

# Learning of Associative Memory Networks Based Upon Cone-Like Domains of Attraction

Niijima, Koichi  
Department of Informatics Kyushu University

<https://hdl.handle.net/2324/3227>

---

出版情報 : DOI Technical Report. 126, 1996-06-18. Department of Informatics, Kyushu University  
バージョン :  
権利関係 :



# DOI Technical Report

## Learning of Associative Memory Networks Based Upon Cone-Like Domains of Attraction

Koichi NIIJIMA

June 18, 1996

Department of Informatics  
Kyushu University 33  
Fukuoka 812, Japan

E-mail: [niijima@i.kyushu-u.ac.jp](mailto:niijima@i.kyushu-u.ac.jp) Phone: +81-92-583-7631

# Learning of Associative Memory Networks Based Upon Cone-Like Domains of Attraction

KOICHI NIIJIMA

Kyushu University

**Abstract**— *A learning algorithm for single layer perceptrons is proposed. First, cone-like domains, each of which is mapped by the perceptron network into almost an associative pattern, are derived. The learning algorithm is obtained as a process that enlarges the cone-like domains. For autoassociative networks, it is shown that the cone-like domains become domains of attraction for stored patterns in the network. In this case, extended domains of attraction are also obtained by feeding the outputs of the network back to the input layer. In computer simulations, character recognition ability of the autoassociative network is examined.*

**Keywords**—Cone-like domains, Learning algorithm, Autoassociative memory networks, Extended domains of attraction, Pattern recognition.

## 1. INTRODUCTION

Associative memory networks have been studied mainly from two aspects. One of them concerns learning algorithms for the networks, and the other their recalling ability.

There are several learning algorithms of associative memory networks such as the correlation recording, the generalized-inverse recording (Kohonen, 1984), and the Ho-Kashyap algorithm (Hassoun & Song, 1992). These learning techniques have been surveyed in (Hassoun, 1991) together with the capacity and performance of associative memories. Recently, a perceptron learning method on incorporating basins of attraction is proposed (Brouwer, 1995).

Storage capacity of associative memories has been studied by McEliece et al. (1987), Cottrell(1988), Amari(1990), and Kohring(1993), from the viewpoint of domains of attraction in a binary-valued vector space.

Our recent papers Nijima(1994) and Nijima(1995) have treated associative memory networks in a real-valued vector space. We have derived domains of attraction of the networks, and proposed some learning algorithms of the networks based on the domains. In Nijima(1994), domains of attraction for autoassociative memories have been derived under equality associative conditions. Associative conditions, however, are not necessarily of equality type. Equality associative conditions restrict the number of stored patterns and the size of domains of attraction. In Nijima(1995), we relaxed these equality associative conditions into inequality associative ones. Under the relaxed conditions, we derived domains of attraction and gave a learning algorithm based on the domains. However, the domains are much smaller in comparison with domains obtained by a perceptron classifier.

In this paper, we derive cone-like domains, each of which is mapped to almost an associative pattern, under the associative conditions of inequality type. These domains are much larger than those obtained in Nijima(1994) and Nijima(1995), and reasonable from the viewpoint of partitioning the input space by hyperplanes. It is shown that these domains are mutually disjoint. These results are given in Section 2.

The boundaries of the cone-like domain consist of some hyperplanes including undetermined weights and thresholds in the network. It is desirable to determine the weights and thresholds so that the cone-like domain becomes as large as possible. The cone-like domain is a union of a cone and some stripes. To expand the cone and to enlarge the stripes, we derive a functional to be minimized. The functional is minimized under the inequality associative conditions. We adopt a penalty method to solve the minimization problem subject to the inequality constraints. The functional with penalty terms can be minimized using a variety of gradient methods. The minimization process gives a learning algorithm of the network. These are described in Section 3.

Our theory developed in Sections 2 and 3 is applied to analyze autoassociative networks. We put on the network the condition that when a binary valued stored pattern is input in the network, almost the same pattern is output. This condition is expressed in an inequality form and the above theory is applied to obtain domains each of which is mapped by the network function to almost a stored pattern. We show by the contraction mapping theorem that the function has a unique fixed point in the domain, which is extremely near a stored pattern. This implies that the domain is attractive and that the fixed point is its attractor. We further find extended domains of attraction larger than the original domain of attraction. Connection weights

of the network are determined using the learning technique in Section 3. These are given in Section 4.

In computer simulations in Section 5, an autoassociative network is learnt by our algorithm using the English alphabet and its recognition ability is verified by checking which domain some randomly noised patterns are contained in. Concluding remarks are described in Section 6.

## 2. CONE-LIKE DOMAINS OF STORED PATTERNS

Let  $n$  be the number of input nodes, and  $\ell$  the number of output units. We consider a single layer neural network:

$$y_i = f\left(\sum_{k=1}^n w_{ik}x_k - \theta_i\right), \quad i = 1, 2, \dots, \ell, \quad (1)$$

where  $w_{ik}$  denote weights connecting the input nodes and the output layer, and  $\theta_i$  threshold parameters. The function  $f(t)$  indicates the sigmoid function:

$$f(t) = \frac{1}{1 + \exp(-t)}.$$

The equation (1) is expressed using the inner product symbol  $\cdot$  as

$$\begin{aligned} y_i &= f(W_i \cdot x - \theta_i) \\ &\equiv \varphi_i(x), \quad i = 1, 2, \dots, \ell, \end{aligned} \quad (2)$$

where  $W_i = {}^t(w_{i1}, w_{i2}, \dots, w_{in})$ ,  $x = {}^t(x_1, x_2, \dots, x_n)$  with the transpose symbol  ${}^t$ . Putting further  $\varphi(x) = {}^t(\varphi_1(x), \varphi_2(x), \dots, \varphi_\ell(x))$  and  $y = {}^t(y_1, y_2, \dots, y_\ell)$ , we write (2) as

$$y = \varphi(x).$$

Let  $x^\nu$ ,  $\nu = 1, 2, \dots, m$ , denote patterns to be stored in the network (2), and  $\varepsilon$  be a sufficiently small positive number. We impose on the network the associative condition that when  $x^\nu$  is input into (2), a number smaller than  $\varepsilon$  or larger than  $1 - \varepsilon$  is output, namely,

$$\varphi_i(x^\nu) \leq \varepsilon \quad \text{or} \quad \varphi_i(x^\nu) \geq 1 - \varepsilon.$$

The aim of this paper is to clarify a domain whose any element is recognized as the stored pattern  $x^\nu$ , under the imposed associative condition, and to derive a learning algorithm of the network based on the domain.

Let us define two sets of indexes  $i$ , depending on the index  $\nu$  of the stored pattern  $x^\nu$ :

$$\begin{aligned} I_{\nu,-} &= \{i \mid \varphi_i(x^\nu) \leq \varepsilon\}, \\ I_{\nu,+} &= \{i \mid \varphi_i(x^\nu) \geq 1 - \varepsilon\}. \end{aligned}$$

Of course,  $I_{\nu,-} \cup I_{\nu,+} = \{1, 2, \dots, \ell\}$  holds. We shall derive an equivalent condition to  $\varphi_i(x^\nu) \leq \varepsilon$  for  $i \in I_{\nu,-}$ . Since

$s = f(t)$  is monotonically increasing and has the inverse function  $t = \ln(s/(1-s))$ , we have

$$W_i \cdot x^\nu - \theta_i \leq -\ln \frac{1-\varepsilon}{\varepsilon}, \quad i \in I_{\nu,-}, \quad \nu = 1, 2, \dots, m. \quad (3)$$

Similarly, the condition  $\varphi_i(x^\nu) \geq 1 - \varepsilon$  for  $i \in I_{\nu,+}$  is equivalent to

$$W_i \cdot x^\nu - \theta_i \geq \ln \frac{1-\varepsilon}{\varepsilon}, \quad i \in I_{\nu,+}, \quad \nu = 1, 2, \dots, m. \quad (4)$$

We fix  $\nu$  and consider a domain of  $x$  satisfying

$$\varphi_i(x) \leq \varphi_i(x^\nu), \quad i \in I_{\nu,-} \quad (5)$$

and

$$\varphi_i(x^\nu) \leq \varphi_i(x), \quad i \in I_{\nu,+}. \quad (6)$$

The condition (5) means  $f(W_i \cdot x - \theta_i) \leq f(W_i \cdot x^\nu - \theta_i)$  for  $i \in I_{\nu,-}$ , which is equivalent to

$$W_i \cdot (x - x^\nu) \leq 0, \quad i \in I_{\nu,-}.$$

Similarly, (6) is equivalent to

$$W_i \cdot (x - x^\nu) \geq 0, \quad i \in I_{\nu,+}.$$

Therefore, the domain of  $x$  satisfying (5) and (6) can be represented as

$$\begin{aligned} \text{Cone}(x^\nu) &= \{x \mid W_i \cdot (x - x^\nu) \leq 0, \quad i \in I_{\nu,-}, \\ &\quad W_i \cdot (x - x^\nu) \geq 0, \quad i \in I_{\nu,+}\}. \end{aligned}$$

This domain is a cone with the root  $x^\nu$ , and is deeply related to a partitioning of the input space by the hyperplanes of the network (2). Indeed, it holds for  $x \in \text{Cone}(x^\nu)$  that

$$W_i \cdot x - \theta_i \leq W_i \cdot x^\nu - \theta_i \leq -\ln \frac{1-\varepsilon}{\varepsilon}, \quad i \in I_{\nu,-} \quad (7)$$

and

$$W_i \cdot x - \theta_i \geq W_i \cdot x^\nu - \theta_i \geq \ln \frac{1-\varepsilon}{\varepsilon}, \quad i \in I_{\nu,+}, \quad (8)$$

which are similar to the partitioning condition

$$W_i \cdot x - \theta_i < 0 \quad \text{or} \quad > 0, \quad (9)$$

although (7) and (8) are stronger than (9). In  $\text{Cone}(x^\nu)$ , we have a quantitative result.

LEMMA. For any  $x \in \text{Cone}(x^\nu)$ , we have

$$\varphi_i(x) \leq \varepsilon, \quad i \in I_{\nu,-}, \quad (10)$$

$$\varphi_i(x) \geq 1 - \varepsilon, \quad i \in I_{\nu,+}. \quad (11)$$

We further have, for any  $x, \tilde{x} \in \text{Cone}(x^\nu)$ ,

$$|\varphi_i(x) - \varphi_i(\tilde{x})| \leq \varepsilon |W_i \cdot (x - \tilde{x})|, \quad i = 1, 2, \dots, \ell. \quad (12)$$

*Proof.* The inequalities (10) and (11) follow immediately from (5) and the definition of  $I_{\nu,-}$ , and from (6) and the definition of  $I_{\nu,+}$ , respectively.

The inequality (12) is obtained by using the mean value theorem. By the derivative  $f'(t) = f(t)(1 - f(t))$ , we have

$$\begin{aligned}\varphi_i(x) - \varphi_i(\tilde{x}) &= f(W_i \cdot x - \theta_i) - f(W_i \cdot \tilde{x} - \theta_i) \\ &= f(\bar{t})(1 - f(\bar{t})) W_i \cdot (x - \tilde{x}),\end{aligned}$$

where  $\bar{t} = W_i \cdot (\lambda x + (1 - \lambda)\tilde{x}) - \theta_i$  with  $0 < \lambda < 1$ . Putting  $\bar{x} = \lambda x + (1 - \lambda)\tilde{x}$ , we further have

$$\varphi_i(x) - \varphi_i(\tilde{x}) = \varphi_i(\bar{x})(1 - \varphi_i(\bar{x})) W_i \cdot (x - \tilde{x}).$$

Since  $\bar{x}$  belongs to  $\text{Cone}(x^\nu)$  because of its convexity, we obtain (12) from the first results (10) and (11). ■

The inequalities (10) and (11) mean that any pattern in  $\text{Cone}(x^\nu)$  is recognized as the stored pattern  $x^\nu$ .

Next, we shall introduce a domain  $D_\rho(x^\nu)$  larger than  $\text{Cone}(x^\nu)$ , whose any pattern can also be recognized as  $x^\nu$ .

$$D_\rho(x^\nu) = \{x \mid W_i \cdot (x - x^\nu) \leq \rho |W_i \cdot x^\nu - \theta_i|, \quad i \in I_{\nu,-}, \\ W_i \cdot (x - x^\nu) \geq -\rho |W_i \cdot x^\nu - \theta_i|, \quad i \in I_{\nu,+}\},$$

where  $\rho$  satisfies  $0 < \rho < 1$ . It is obvious that  $\text{Cone}(x^\nu) \subset D_\rho(x^\nu)$ . The domains  $\text{Cone}(x^\nu)$  and  $D_\rho(x^\nu)$  are illustrated below.

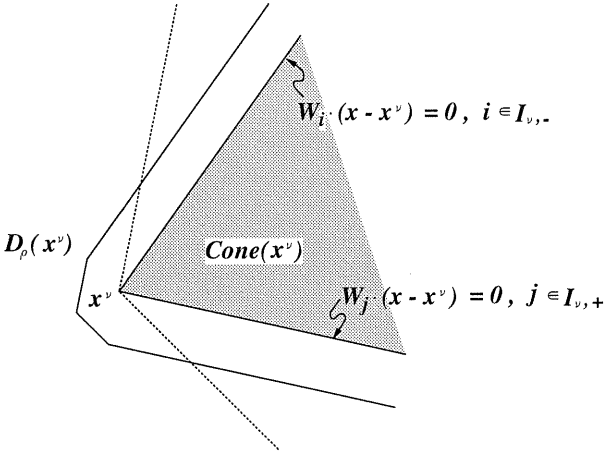


FIGURE 1.  $\text{Cone}(x^\nu)$  and  $D_\rho(x^\nu)$ .

For any  $x \in D_\rho(x^\nu)$ , it holds that

$$\begin{aligned}W_i \cdot x - \theta_i &\leq (1 - \rho) (W_i \cdot x^\nu - \theta_i) \\ &\leq -(1 - \rho) \ln \frac{1 - \varepsilon}{\varepsilon}, \quad i \in I_{\nu,-}\end{aligned}\quad (13)$$

and

$$\begin{aligned}W_i \cdot x - \theta_i &\geq (1 - \rho) (W_i \cdot x^\nu - \theta_i) \\ &\geq (1 - \rho) \ln \frac{1 - \varepsilon}{\varepsilon}, \quad i \in I_{\nu,+}\end{aligned}\quad (14)$$

These conditions approach to the partitioning condition (9) when  $\rho$  tends to 1.

The results in Lemma can be extended to the domain  $D_\rho(x^\nu)$  as follows.

**THEOREM 1.** For any  $x \in D_\rho(x^\nu)$ , we have

$$\varphi_i(x) \leq \varepsilon^{1-\rho}, \quad i \in I_{\nu,-}, \quad (15)$$

$$\varphi_i(x) \geq 1 - \varepsilon^{1-\rho}, \quad i \in I_{\nu,+}. \quad (16)$$

Furthermore, we have for any  $x, \tilde{x} \in D_\rho(x^\nu)$ ,

$$|\varphi_i(x) - \varphi_i(\tilde{x})| \leq \varepsilon^{1-\rho} |W_i \cdot (x - \tilde{x})|, \quad i = 1, 2, \dots, \ell. \quad (17)$$

*Proof.* For  $i \in I_{\nu,-}$ , the inequality (13) and the monotonicity of  $f_i(t)$  leads us to

$$f(W_i \cdot x - \theta_i) \leq f(-(1 - \rho) \ln \frac{1 - \varepsilon}{\varepsilon}). \quad (18)$$

Applying the inequality

$$(1 - \rho) \ln \frac{1 - \varepsilon}{\varepsilon} \geq \ln \frac{1 - \varepsilon^{1-\rho}}{\varepsilon^{1-\rho}}, \quad (19)$$

to (18), we have

$$f(W_i \cdot x - \theta_i) \leq f(\ln \frac{\varepsilon^{1-\rho}}{1 - \varepsilon^{1-\rho}}) = \varepsilon^{1-\rho}$$

which proves (15).

For  $i \in I_{\nu,+}$ , we have from (14) and with the aid of (19),

$$f(W_i \cdot x - \theta_i) \geq f(\ln \frac{1 - \varepsilon^{1-\rho}}{\varepsilon^{1-\rho}}) = 1 - \varepsilon^{1-\rho}$$

which implies (16).

The estimate (17) is derived in the same way as in Lemma. ■

The inequalities (15) and (16) mean that any pattern in  $D_\rho(x^\nu)$  as well as in  $\text{Cone}(x^\nu)$  is recognized as  $x^\nu$ .

The following shows that  $D_\rho(x^\nu)$ ,  $\nu = 1, 2, \dots, m$ , are mutually disjoint.

**COROLLARY.** If  $I_{\nu,-} \neq I_{\mu,-}$  for  $\nu \neq \mu$ , then the domains  $D_\rho(x^\nu)$  and  $D_\rho(x^\mu)$  are disjoint.

*Proof.* This result can be shown using proof by contradiction. By  $I_{\nu,-} \neq I_{\mu,-}$ , there exists an index  $i^*$  such that  $i^* \in I_{\nu,-}$  and  $i^* \notin I_{\mu,-}$ . Since  $I_{\mu,-} \cup I_{\mu,+} = \{1, 2, \dots, \ell\}$ , the index  $i^*$  belongs to  $I_{\mu,+}$ . Suppose that  $D_\rho(x^\nu) \cap D_\rho(x^\mu) \neq \emptyset$ , where  $\emptyset$  denotes the empty set. Then there exists  $x^* \in D_\rho(x^\nu) \cap D_\rho(x^\mu)$ . From  $x^* \in D_\rho(x^\nu)$  and  $i^* \in I_{\nu,-}$ , we have

$$\varphi_{i^*}(x^*) \leq \varepsilon^{1-\rho}. \quad (20)$$

On the other hand, we obtain from  $x^* \in D_\rho(x^\mu)$  and  $i^* \in I_{\mu,+}$ ,

$$\varphi_{i^*}(x^*) \geq 1 - \varepsilon^{1-\rho}$$

which contradicts (20). ■

This corollary means that the domains  $D_\rho(x^\nu)$ ,  $\nu = 1, 2, \dots, m$ , can be classified if target patterns for  $x^\nu$ ,  $\nu = 1, 2, \dots, m$ , are different from each other.

It is desirable for the domain  $D_\rho(x^\nu)$  to be as large as possible. In Section 3, we propose a method for determining weights and thresholds of the network so that  $D_\rho(x^\nu)$  becomes large.

### 3. LEARNING ALGORITHM

Classification ability of the network (2) can be measured by the size of the domain  $D_\rho(x^\nu)$  in Theorem 1. It is desirable for the domain to be as large as possible. The domain  $D_\rho(x^\nu)$  is a union of the cone  $Cone(x^\nu)$  and the stripe  $Str(x^\nu)$ :

$$Str(x^\nu) = \{x \mid 0 < W_i \cdot (x - x^\nu) \leq \rho |W_i \cdot x^\nu - \theta_i|, i \in I_{\nu,-}, \\ 0 > W_i \cdot (x - x^\nu) \geq -\rho |W_i \cdot x^\nu - \theta_i|, i \in I_{\nu,+}\}.$$

Therefore, the domain  $D_\rho(x^\nu)$  can be enlarged by expanding both  $Str(x^\nu)$  and  $Cone(x^\nu)$ .

The first task is to expand the stripe  $Str(x^\nu)$ . The width of stripe between the hyperplanes  $W_i \cdot (x - x^\nu) = 0$  and  $W_i \cdot (x - x^\nu) = \rho |W_i \cdot x^\nu - \theta_i|$  is given by

$$\frac{\rho |W_i \cdot x^\nu - \theta_i|}{\|W_i\|}. \quad (21)$$

which was derived in Nijima (1995). We want to determine  $W_i$  and  $\theta_i$  so as to maximize (21). However, the quantity (21) depends on the index  $\nu$  and so  $\sum_{\nu=1}^m (W_i \cdot x^\nu - \theta_i)^2 / \|W_i\|^2$  is maximized, that is,

$$\frac{\|W_i\|^2}{\sum_{\nu=1}^m (W_i \cdot x^\nu - \theta_i)^2} \quad (22)$$

is minimized. The next task is to expand the cone  $Cone(x^\nu)$ . The relation between  $Cone(x^\nu)$  and its boundaries is given below.

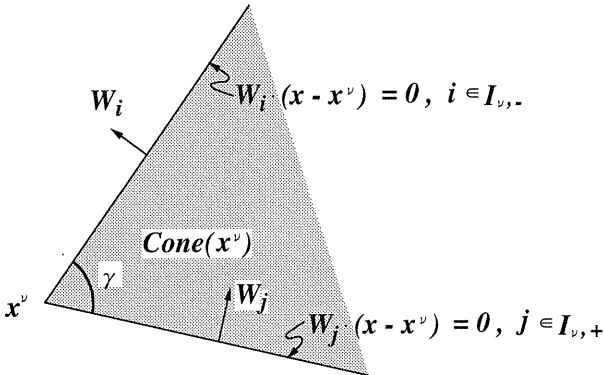


FIGURE 2.  $Cone(x^\nu)$  and its boundaries.

The domain  $Cone(x^\nu)$  is an intersection of wedge-shaped domains each of which lies between the hyperplanes  $H_i : W_i \cdot (x - x^\nu) = 0$  for  $i \in I_{\nu,-}$  and  $H_j : W_j \cdot (x - x^\nu) = 0$  for  $j \in I_{\nu,+}$ . To enlarge the wedge-shaped domain between  $H_i$  and  $H_j$ , it suffices to maximize an angle  $\gamma$  at which  $H_i$  and  $H_j$  cross. This maximization implies the minimization of

$$\cos \gamma = \frac{W_i \cdot W_j}{\|W_i\| \|W_j\|}, \quad (23)$$

where  $\|\cdot\|$  denotes the Euclidean norm. This minimization is done only for a pair  $(i, j)$  satisfying  $i \in I_{\nu,-}$  and  $j \in I_{\nu,+}$  for at least one  $\nu$ . For latter convenience, we denote by  $S$  the set of  $(i, j)$  belonging to  $I_{\nu,-} \times I_{\nu,+}$  for at least one  $\nu$ .

We next focus the inequality associative conditions (3) and (4). Introducing two index sets

$$N_{i,-} = \{\nu \mid \varphi_i(x^\nu) \leq \varepsilon\}, \quad N_{i,+} = \{\nu \mid \varphi_i(x^\nu) \geq 1 - \varepsilon\},$$

we rewrite (3) and (4) as

$$W_i \cdot x^\nu - \theta_i \leq -\ln \frac{1 - \varepsilon}{\varepsilon}, \quad \nu \in N_{i,-}, \quad i = 1, 2, \dots, \ell, \quad (24)$$

$$W_i \cdot x^\nu - \theta_i \geq \ln \frac{1 - \varepsilon}{\varepsilon}, \quad \nu \in N_{i,+}, \quad i = 1, 2, \dots, \ell, \quad (25)$$

respectively.

Learning of connection weights is done by minimizing a weighted sum of (22) and (23) under the conditions (24) and (25). The formulation of this minimization problem by the penalty method is

$$\sum_{i=1}^{\ell} \frac{\|W_i\|^2}{\sum_{\nu=1}^m (W_i \cdot x^\nu - \theta_i)^2} + C_1 \sum_{\substack{i < j \\ (i,j) \in S}} \frac{W_i \cdot W_j}{\|W_i\| \|W_j\|} \\ + C_3 \sum_{i=1}^{\ell} \left[ \sum_{\nu \in N_{i,-}} \left( \ln \frac{1 - \varepsilon}{\varepsilon} + W_i \cdot x^\nu - \theta_i \right)_+^2 \right. \\ \left. + \sum_{\nu \in N_{i,+}} \left( \ln \frac{1 - \varepsilon}{\varepsilon} - W_i \cdot x^\nu + \theta_i \right)_+^2 \right] \\ \longrightarrow \min,$$

where  $z_+^2 = z^2$  for  $z > 0$ , 0 otherwise, and  $C_1$  and  $C_3$  denote penalty constants. In actual computation, however, we minimize the following functional to avoid numerical instability:

$$J = \sum_{i=1}^{\ell} \frac{1}{\sum_{\nu=1}^m (V_i \cdot x^\nu - \eta_i)^2} + C_1 \sum_{\substack{i < j \\ (i,j) \in S}} V_i \cdot V_j \\ + C_2 \sum_{i=1}^{\ell} (\|V_i\|^2 - 1)^2 \\ + C_3 \sum_{i=1}^{\ell} \left[ \sum_{\nu \in N_{i,-}} (\alpha + \beta_i (V_i \cdot x^\nu - \eta_i))^2 \right.$$

$$+ \sum_{\nu \in N_{i,+}} (\alpha - \beta_i(V_i \cdot x^\nu - \eta_i))^2_+ ] \longrightarrow \min, \quad (26)$$

where  $C_2$  denotes a penalty constant and we have put  $V_i = W_i/\|W_i\|$ ,  $\eta_i = \theta_i/\|W_i\|$ ,  $\alpha = \ln((1-\varepsilon)/\varepsilon)$  and  $\beta_i = \|W_i\|$ . We solve (26), for example, by the conjugate gradient method to obtain  $V_i$ ,  $\eta_i$  and  $\beta_i$ , and compute the connection weights  $W_i = \beta_i V_i$  and  $\theta_i = \beta_i \eta_i$ .

#### 4. AUTOASSOCIATIVE MEMORY NETWORK

It is well known that a single layer perceptron has limitations in its classification ability (Minsky & Papert, 1969). However, if the network discussed previously is of autoassociative type, we can improve the recalling ability by feeding the output back to the input. In this section, we deal with an autoassociative memory network. Let us choose  $\ell = n$  in the network (2). The stored patterns  $x^\nu, \nu = 1, 2, \dots, m$ , are assumed to be  $(0, 1)$ -valued vectors. We impose on the network the following condition for  $x^\nu = {}^t(x_1^\nu, x_2^\nu, \dots, x_n^\nu)$ :

$$\varphi_i(x^\nu) \begin{cases} \geq 1 - \varepsilon, & x_i^\nu = 1, \\ \leq \varepsilon, & x_i^\nu = 0. \end{cases} \quad (27)$$

The condition (27) means that for the input pattern  $x^\nu$ , a pattern extremely near  $x^\nu$  is output, because  $\varepsilon$  is small enough. Namely, the network satisfying (27) is almost of autoassociative type. However, the stored pattern  $x^\nu$  itself is not a fixed point of  $\varphi$ , namely,  $x^\nu \neq \varphi(x^\nu)$ . In Theorem 2 below, we show that there exists a pattern  $x^{\nu,*}$  extremely near  $x^\nu$  such that  $x^{\nu,*} = \varphi(x^{\nu,*})$ .

The condition (27) can be written as (3) and (4), but now  $I_{\nu,-} = \{i \mid x_i^\nu = 0\}$  and  $I_{\nu,+} = \{i \mid x_i^\nu = 1\}$ . From now on, we employ these index sets. Theorem 1 holds for such a network and the domains  $D_\rho(x^\nu)$ ,  $\nu = 1, 2, \dots, m$ , are mutually disjoint. Hence, our learning algorithm in Section 3 can be applied to compute the connection weights  $W_i$  and the threshold parameters  $\theta_i$  of the present network. For this network, we have the following results.

**THEOREM 2.** For any  $x, \tilde{x} \in D_\rho(x^\nu)$ , we have

$$\|\varphi(x) - \varphi(\tilde{x})\| \leq \kappa \|x - \tilde{x}\|,$$

where

$$\kappa = \sqrt{\sum_{i=1}^n \|W_i\|^2 \varepsilon^{1-\rho}}.$$

If  $\varepsilon$  is sufficiently small so as to satisfy  $\kappa < 1$ , the function  $\varphi$  becomes a contractive mapping in  $D_\rho(x^\nu)$ .

Moreover, the equation  $x = \varphi(x)$  has a unique solution  $x^{\nu,*}$  in  $D_\rho(x^\nu)$  satisfying

$$\|x^{\nu,*} - x^\nu\| \leq \frac{\sqrt{n} \varepsilon}{1 - \kappa}. \quad (28)$$

*Proof.* Estimating (17) using Schwarz' inequality leads to

$$|\varphi_i(x) - \varphi_i(\tilde{x})| \leq \varepsilon^{1-\rho} \|W_i\| \|x - \tilde{x}\|.$$

Summing up both sides after squared, we obtain the first assertion.

The latter can be proved by the contraction mapping theorem (Rall, 1969). Before applying this theorem, we describe a simple geometrical notion. We find the distance from  $x^\nu$  to the boundary of  $D_\rho(x^\nu)$ , which consists of the hyperplanes

$$G_i: W_i \cdot (x - x^\nu) = \begin{cases} \rho |W_i \cdot x^\nu - \theta_i|, & i \in I_{\nu,-}, \\ -\rho |W_i \cdot x^\nu - \theta_i|, & i \in I_{\nu,+}. \end{cases}$$

Since  $W_i$  is a normal vector for  $G_i$ , a vector  $x$  from  $x^\nu$  with the direction  $W_i$  is represented as  $x = x^\nu + a W_i$ . From  $x \in G_i$ , we obtain  $a = \pm \rho |W_i \cdot x^\nu - \theta_i|$  and hence, the distance is given by  $\rho |W_i \cdot x^\nu - \theta_i|$  which is larger than  $\rho \ln((1-\varepsilon)/\varepsilon)$  because of (3) and (4). Therefore, the ball

$$B_\delta(x^\nu) = \{x \mid \|x - x^\nu\| \leq \delta, \quad \delta = \rho \ln \frac{1-\varepsilon}{\varepsilon}\}$$

is contained in  $D_\rho(x^\nu)$ . We put  $b = \|\varphi(x^\nu) - x^\nu\|/(1-\kappa)$ . By the contraction mapping theorem,  $\varphi$  has a fixed point  $x^{\nu,*}$  in  $B_\delta(x^\nu)$ , namely,  $x^{\nu,*} = \varphi(x^{\nu,*})$  holds. Furthermore, we have

$$\|x^{\nu,*} - x^\nu\| \leq b.$$

Since  $b \leq \sqrt{n} \varepsilon/(1-\kappa)$  by (27), we obtain (28). ■

In the present network, we can feed the output back to the input. Namely, we can generate  $\{x^{(r)}\}_{r=0,1,\dots}$  by the iteration

$$x^{(r+1)} = \varphi(x^{(r)}), \quad r = 0, 1, \dots,$$

where  $x^{(0)}$  is in  $D_\rho(x^\nu)$ . This sequence converges to the fixed point  $x^{\nu,*}$ . Therefore, we may say  $D_\rho(x^\nu)$  a domain of attraction.

For latter analysis, it is needed to introduce the domain  $D_\rho(x^{\nu,*})$  instead of  $D_\rho(x^\nu)$ . Since  $x^{\nu,*}$  is extremely near  $x^\nu$ , the domain  $D_\rho(x^{\nu,*})$  is almost equal to  $D_\rho(x^\nu)$ . If  $x^{\nu,*} \in \text{Cone}(x^\nu)$ , then the inclusion  $D_\rho(x^{\nu,*}) \subset D_\rho(x^\nu)$  holds and hence, Theorem 1 is valid in  $D_\rho(x^{\nu,*})$ . If not so, Theorem 1 does not hold in  $D_\rho(x^{\nu,*})$ . We have therein a weakened result of Theorem 1.

**THEOREM 3.** Suppose that  $\varepsilon$  is small enough so as to satisfy  $\kappa < 1$  and

$$\frac{\sqrt{n} \varepsilon}{1 - \kappa} \max_{i=1,2,\dots,n} \|W_i\| \leq \frac{1}{2} \ln \frac{1-\varepsilon}{\varepsilon}. \quad (29)$$

Then, we have for any  $x \in D_\rho(x^{\nu,*})$ ,

$$\begin{aligned} \varphi_i(x) &\leq \varepsilon^{(1-\rho)/2}, & i \in I_{\nu,-}, \\ \varphi_i(x) &\geq 1 - \varepsilon^{(1-\rho)/2}, & i \in I_{\nu,+}. \end{aligned}$$

Furthermore, we have for any  $x, \tilde{x} \in D_\rho(x^{\nu,*})$ ,

$$|\varphi_i(x) - \varphi_i(\tilde{x})| \leq \varepsilon^{(1-\rho)/2} |W_i \cdot (x - \tilde{x})|, \quad i = 1, 2, \dots, n. \quad (30)$$

*Proof.* From  $x \in D_\rho(x^{\nu,*})$ , we have for  $i \in I_{\nu,-}$ ,

$$W_i \cdot x - \theta_i \leq W_i \cdot x^{\nu,*} - \theta_i + \rho |W_i \cdot x^{\nu,*} - \theta_i|. \quad (31)$$

Using (3), the term  $W_i \cdot x^{\nu,*} - \theta_i$  is estimated as

$$\begin{aligned} W_i \cdot x^{\nu,*} - \theta_i &= W_i \cdot x^\nu - \theta_i + W_i \cdot (x^{\nu,*} - x^\nu) \\ &\leq -\ln \frac{1-\varepsilon}{\varepsilon} + \|W_i\| \|x^{\nu,*} - x^\nu\|. \end{aligned}$$

A further estimation by (28) and (29) yields  $W_i \cdot x^{\nu,*} - \theta_i \leq -1/2 \ln((1-\varepsilon)/\varepsilon)$ . Combining this estimation with (31) gives

$$\begin{aligned} W_i \cdot x - \theta_i &\leq (1-\rho)(W_i \cdot x^{\nu,*} - \theta_i) \\ &\leq -\frac{1-\rho}{2} \ln \frac{1-\varepsilon}{\varepsilon}. \end{aligned}$$

Similarly, we obtain for  $i \in I_{\nu,+}$ ,

$$W_i \cdot x - \theta_i \geq \frac{1-\rho}{2} \ln \frac{1-\varepsilon}{\varepsilon}.$$

The remaining proof can be done along exactly the same line as in Theorem 1. ■

The first part of Theorem 3 asserts that any pattern in  $D_\rho(x^{\nu,*})$  is recognized as the pattern  $x^{\nu,*}$ . From the second assertion (30), it follows that the network function  $\varphi$  is contractive in  $D_\rho(x^{\nu,*})$  provided  $\varepsilon$  is sufficiently small. Thus  $D_\rho(x^{\nu,*})$  also becomes a basin of attraction.

Let  $\varphi^0$  be the identity mapping in  $R^n$  and define  $\varphi^q$ ,  $q \geq 1$ , by  $\varphi^q(x) = \varphi(\varphi^{q-1}(x))$  recursively. We define a domain of  $x$  whose  $q$ -step recall  $\varphi^q(x)$  belongs to  $D_\rho(x^{\nu,*})$ .

$$\begin{aligned} D_\rho^q(x^{\nu,*}) &= \{ x \mid W_i \cdot (\varphi^q(x) - x^{\nu,*}) \leq \rho |W_i \cdot x^{\nu,*} - \theta_i|, \\ &\quad i \in I_{\nu,-}, \\ &\quad W_i \cdot (\varphi^q(x) - x^{\nu,*}) \geq -\rho |W_i \cdot x^{\nu,*} - \theta_i|, \\ &\quad i \in I_{\nu,+} \}, \end{aligned}$$

where  $D_\rho^0(x^{\nu,*}) = D_\rho(x^{\nu,*})$ . Let  $x$  be any pattern in  $D_\rho^q(x^{\nu,*})$ . Then  $\varphi^q(x)$  belongs to a domain of attraction  $D_\rho(x^{\nu,*})$ . Therefore, we may say  $D_\rho^q(x^{\nu,*})$  also a domain of attraction. It will be shown below that  $D_\rho^q(x^{\nu,*})$  expand together with the recalling step  $q$ .

**THEOREM 4.** Suppose that  $\varepsilon$  satisfies the assumption of Theorem 3 and

$$\sqrt{n} (\varepsilon + \varepsilon^{(1-\rho)/2}) \max_{i=1,2,\dots,n} \|W_i\| \leq \frac{\rho}{2} \ln \frac{1-\varepsilon}{\varepsilon}. \quad (32)$$

Then we have the inclusions

$$D_\rho(x^{\nu,*}) = D_\rho^0(x^{\nu,*}) \subset D_\rho^1(x^{\nu,*}) \subset D_\rho^2(x^{\nu,*}) \subset \dots$$

*Proof.* It suffices to prove  $D_\rho^q(x^{\nu,*}) \subset D_\rho^{q+1}(x^{\nu,*})$  for any  $q \geq 0$ . Let  $x$  be any pattern in  $D_\rho^q(x^{\nu,*})$ . From the definition of the domain,  $\varphi^q(x)$  belongs to  $D_\rho(x^{\nu,*})$ . Put

$u = \varphi^q(x)$ . By the first assertion of Theorem 3 and the property of  $x^{\nu,*}$ , we have  $|\varphi_i(u) - x_i^{\nu,*}| \leq \varepsilon + \varepsilon^{(1-\rho)/2}$  for all  $i$ . Therefore,

$$\|\varphi(u) - x^{\nu,*}\| \leq \sqrt{n} (\varepsilon + \varepsilon^{(1-\rho)/2}). \quad (33)$$

To prove  $x \in D_\rho^{q+1}(x^{\nu,*})$ , it suffices to show that  $\varphi(u)$  is in  $D_\rho(x^{\nu,*})$ . For the purpose, we estimate  $W_i \cdot (\varphi(u) - x^{\nu,*})$ . For  $i \in I_{\nu,-}$ , we have

$$\begin{aligned} W_i \cdot (\varphi(u) - x^{\nu,*}) &\leq \|W_i\| \|\varphi(u) - x^{\nu,*}\| \\ &\leq \frac{\rho}{2} \ln \frac{1-\varepsilon}{\varepsilon}, \end{aligned}$$

where we have used (32) and (33) in the second line. Remembering here the inequality  $|W_i \cdot x^{\nu,*} - \theta_i| \geq 1/2 \ln((1-\varepsilon)/\varepsilon)$  which was used in the proof of Theorem 3, we obtain

$$W_i \cdot (\varphi(u) - x^{\nu,*}) \leq \rho |W_i \cdot x^{\nu,*} - \theta_i|.$$

Similarly, we have for  $i \in I_{\nu,+}$ ,

$$W_i \cdot (\varphi(u) - x^{\nu,*}) \geq -\rho |W_i \cdot x^{\nu,*} - \theta_i|.$$

These show that  $\varphi(u)$  is in  $D_\rho(x^{\nu,*})$ . ■

In the same way as the proof in Corollary, it is shown that the domains  $D_\rho^p(x^{\nu,*})$  and  $D_\rho^q(x^{\mu,*})$  are disjoint for any  $p, q \geq 0$  and  $\nu \neq \mu$ .

## 5. SIMULATIONS

For computer simulations, we store the alphabet of Figure 3 in the autoassociative network in Section 4.



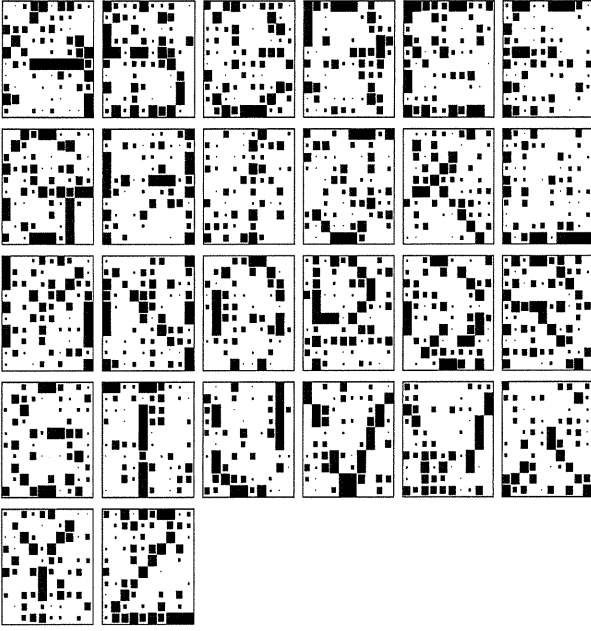
**FIGURE 3.** 26 stored alphabet. The characters  $x^\nu$ ,  $\nu = 1, 2, \dots, 26$  are listed from top to down and from left



to right. Each character is represented by a  $10 \times 10$  grid matrix. Black and white squares indicate 1 and 0, respectively.

We now have  $m = 26$  and  $n = 100$ . The network has 100 input nodes, the same number of output units, 10000 connection weights  $W_i = (w_{ik})$  and 100 thresholds  $\theta_i$ . To determine these parameters, we chose  $\varepsilon = \exp(-9)$  and minimized the functional in (26) using the conjugate gradient method. We check the assumptions of Theorems 2, 3 and 4 based on the computed parameters. We chose  $\rho = 0.4$  for  $\varepsilon = \exp(-9)$ . The quantity  $\sqrt{\sum_{i=1}^{100} \|W_i\|^2}$  is 16.183 and the maximum of  $\|W_i\|$ ,  $i = 1, 2, \dots, 100$  is 2.173. Hence, the contractive factor  $\kappa$  in Theorem 2 is 0.658 which is smaller than 1. Since the left and right hand sides of (29) in Theorem 3 are 0.0078 and 4.5, respectively, the condition (29) is satisfied. The inequality (32) in Theorem 4 is also fulfilled, because the left and right hand sides of (32) are 1.4629 and 1.8, respectively.

First, we shall try the recognition of 80% random noised patterns:



**FIGURE 4.** 80% random noised patterns. Each pattern is represented by a  $10 \times 10$  grid matrix. Each square denotes a gray level.

To make noised patterns, we chose a position at random from 100 pixels in each characters, and put there a random number from 0 to 9. The 80% noised patterns above were made by repeating this procedure 80 times.

To check which domain of attraction each noised pattern is contained in, we computed two kinds of ratio

$$r_{-}^{q,\nu} = \max_{i \in I_{\nu,-}} \frac{W_i \cdot (\varphi^q(x) - x^{\nu,*})}{|W_i \cdot x^{\nu,*} - \eta_i|}$$

and

$$r_{+}^{q,\nu} = \min_{i \in I_{\nu,+}} \frac{W_i \cdot (\varphi^q(x) - x^{\nu,*})}{|W_i \cdot x^{\nu,*} - \eta_i|}$$

for  $0 \leq q \leq 10$  and  $\nu = 1, 2, \dots, 26$ , where we have used  $x^\nu$  in place of  $x^{\nu,*}$ . If we have  $r_{-}^{q,\nu} \leq 0.4$  and  $r_{+}^{q,\nu} \geq -0.4$  for some  $q$  and  $\nu$ , the noised pattern  $x$  can be recognized as  $x^\nu$ .

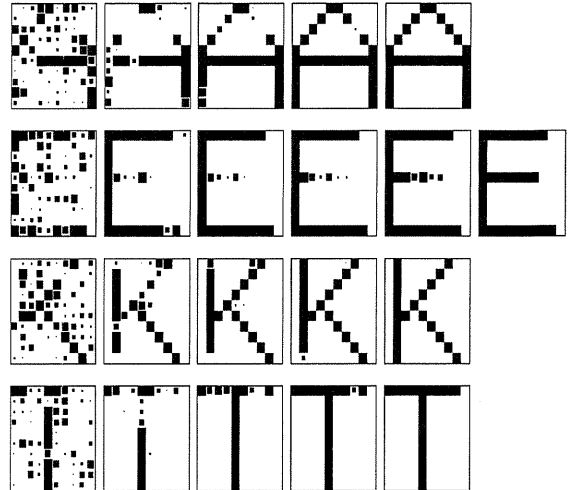
In the table below, we list, for each noised pattern in the Figure 4, the recalling number  $q$ , two values  $r_{-}^{q,\nu}$  and  $r_{+}^{q,\nu}$  such that  $r_{-}^{q,\nu} \leq 0.4$  and  $r_{+}^{q,\nu} \geq -0.4$  are satisfied. The bottom in each box represents the recognized pattern  $x^\nu$ .

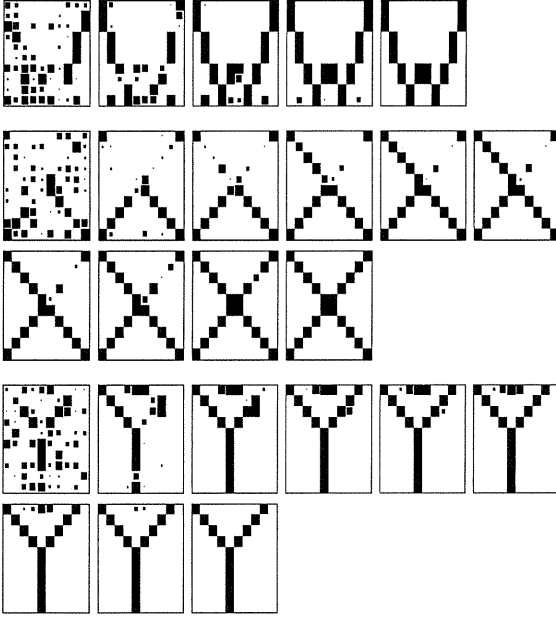
**TABLE 1**  
Recognition results for 80% random noised patterns

4 0.010 -0.051 A	3 0.083 -0.326 B	2 0.287 -0.292 C	3 0.060 -0.230 D	5 0.002 -0.012 E	2 0.033 -0.121 F
2 0.329 -0.157 G	2 0.250 -0.066 H	2 0.064 -0.173 I	2 0.045 -0.142 J	4 0.002 -0.006 K	2 0.153 -0.261 L
3 0.004 -0.009 M	3 0.162 -0.047 N	2 0.082 -0.176 O	2 0.104 -0.313 P	3 0.049 -0.026 Q	3 0.002 -0.005 R
2 0.132 -0.272 S	4 0.001 -0.005 T	2 0.013 -0.040 U	2 0.018 -0.031 V	4 0.024 -0.006 W	8 0.064 -0.301 X
8 0.005 -0.002 Y	3 0.113 -0.370 Z				

This table shows that all noised patterns in Figure 4 are recognized correctly.

We illustrate here the recalling processes for some noised pattern in Figure 4.

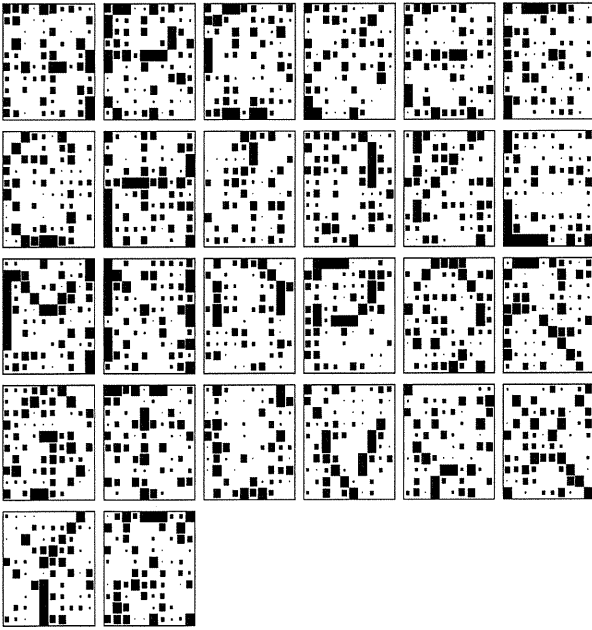




**FIGURE 5.** The recalling processes for some noisy patterns in Figure 4.

Figure 5 presents the situation that the target patterns are recalled by feeding output patterns back to the input layer successively.

The next simulation is done for the following 90% random noisy patterns.



**FIGURE 6.** 90% random noisy patterns. Each pattern is represented by a  $10 \times 10$  grid matrix. Each square denotes a gray level.

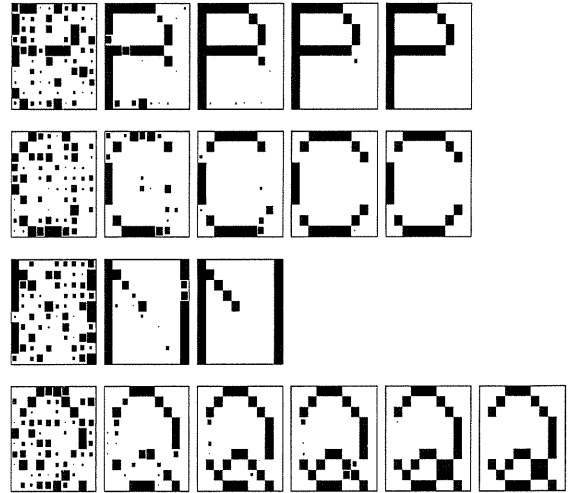
The results of recognition are shown below:

**TABLE 2**  
Recognition results for 90% random noisy patterns

6		3	2	4	4
0.051		0.125	0.115	0.034	0.014
-0.263		-0.134	-0.396	-0.101	-0.053
A	×	C	D	E	F
3	4	4	8	4	3
0.129	0.002	0.033	0.003	0.013	0.230
-0.065	-0.005	-0.010	-0.007	-0.069	-0.035
C	H	I	J	K	L
2		4	4		6
0.099		0.037	0.006		0.002
-0.070		-0.101	-0.002		-0.005
M	×	O	P	×	R
3	7	3	2	2	1
0.015	0.006	0.003	0.005	0.101	0.284
-0.029	-0.023	-0.011	-0.008	-0.258	-0.382
S	T	U	V	W	X
3	8				
0.156	0.008				
-0.017	-0.026				
Y	Z				

The table 2 shows that three patterns failed to be recognized. Also the 7-th pattern was recognized as “C”. However, this failure is permitted because of the similarity of the characters “C” and “G”.

We illustrate below some examples failed to recall.



**FIGURE 7.** Examples failed to recall.

These results show that some noisy patterns in Figure 6 converge to spurious memory patterns which exist outside the extended domains of attractions.

From the results of Tables 1 and 2, we see that our autoassociative memory has a powerful recognition ability.

## 6. CONCLUDING REMARKS

We derived a cone-like domain of attraction for associative memories. The domain is much larger than the domain of attraction given in Nijijima (1994) and Nijijima (1995). The cone-like domain is similar to that obtained by partitioning the input space using the conventional hyperplanes. The domain includes only one associative pattern and any pattern in the domain is mapped by the network function into almost the associative pattern. This enables us to recognize noised patterns.

We also proposed a learning algorithm based upon our cone-like domain. The algorithm is a minimizing process for a functional which was derived along the principle that enlarges the cone-like domain.

Furthermore, we applied our theory to autoassociative memory networks, and derived larger domains of attraction than the original one.

In the simulation, an autoassociative memory network was constructed by using our learning algorithm based on the English alphabet. Its recognition ability is superior to that by the associative memory networks in Nijijima (1994) and Nijijima (1995).

Since our network contains only one nonlinear layer, a linear theory can be applied. Even if a network is of multilayer type, domains of attraction themselves can be obtained under associative conditions imposed on the final layer neurons of the network. However, these domains contain nonlinear functions and their shape is complicated. It is a future work to clarify such domains and to find out learning algorithms for multilayer networks.

## REFERENCES

- Amari, S. I. (1990). Mathematical foundations of neuro computing. *Proceedings of the IEEE*, **78**, 1443-1463.
- Brouwer, R. K. (1995). A method for training recurrent neural networks for classification by building basins of attraction. *Neural Networks*, **8**, 597-603.
- Cottrell, M. (1988). Stability and attractivity in associative memory networks. *Biological Cybernetics*, **58**, 129-139.
- Hassoun, M. H. (1991). Dynamic associative memories. In I. K. Sethi & A. K. Jain, (Eds.), *Artificial neural networks and statistical pattern recognition*, (pp. 195-218). Amsterdam: North-Holland.
- Hassoun, M. H. & Song, J. (1992). Adaptive Ho-Kashyap rules for perceptron training. *IEEE Trans. Neural Networks*, **3**, 51-61.
- Kohonen, T. (1984). *Self-organization and associative memory*. New York: Springer.
- Kohring, G. A. (1993). On the Q-state neuron problem in attractor neural networks. *Neural Networks*, **6**, 573-581.
- McEliece, R. J., Posner, E. C., Rodemich, E. R. & Venkatesh, S. S. (1987). The capacity of the Hopfield associative memory. *IEEE Trans. Inform. Theory*, **33**, 461-482.
- Minsky, M. L. & Papert, S. A. (1969). *Perceptrons*. Cambridge: MIT Press.
- Nijijima, K. (1994). Domains of attraction in autoassociative memory networks. *New Generation Computing*, **12**, 395-407.
- Nijijima, K. (1995). Learning of associative memory networks by penalty methods. *RIFIS Technical Reports*, **TR-CS-113**, 1-9.
- Rall, L. B. (1969). *Computational solution of nonlinear operator equations*. New York: Wiley.