# Convergence properties of symmetric learning algorithm for pattern classification

# Convergence Properties of Symmetric Learning Algorithm for Pattern Classification

S.Miyoshi, Kobe City College of Technology, Japan
K.Ikeda, K.Nakayama, Kanazawa University, Japan
E-mail : miyoshi@kobe-kosen.ac.jp

*Abstract*—**The geometric learning algorithm (GLA) was proposed as an application of the affine projection algorithm (APA) for an adaptive filter to perceptron. In the GLA, the connection weight vector $w(n)$ is updated vertically towards the orthogonal complement of $k$ patterns. The GLA demonstrates some typical behavior when the learning rate $\lambda$ is 2, which means that $w(n)$ and $w(n+1)$ are symmetric with respect to the complement. Therefore, in this paper, the GLA with $\lambda = 2$ is discriminated as "symmetric learning algorithm (SLA)" and the convergence properties of the SLA are analyzed. The convergence condition among the order $k$ of the SLA, the number $P$ of patterns and the dimension $N$ of patterns is analyzed theoretically. It is proved that $k < N$ is the necessary condition for convergence when $P \geq 2N$. The relation between $k$ and the learning speed is analyzed theoretically. It becomes clear that the maximum learning speed on average can be obtained when $k = N/2$. These properties are supported by computer simulations. Furthermore, the goodness of the solution by the SLA is investigated through computer simulation. That is, there exists little difference in the goodness of solution by changing the order $k$.**

*Keyword-* **perceptron, pattern classification, affine projection algorithm, geometric learning algorithm, symmetric learning algorithm.**

## I. INTRODUCTION

In the field of adaptive filters, the affine projection algorithm (APA)[1] is well known as a generalized algorithm of the normalized LMS algorithm[2],[3] into the block signal processing.

We proposed the geometric learning algorithm (GLA) as an application of the APA to perceptron[4],[5]. The connection weight vector is updated vertically towards the orthogonal complement of $k$ pattern vectors by the $k$th order GLA. In[4],[5], the necessary condition for the convergence of the 1st order GLA for 2 patterns' classification within a finite number of steps regardless of the initial weights was derived theoretically. That is, the upper and the lower bounds of the learning rate $\lambda$ for convergence are determined by the angle $\theta$ of the solution area. When the number of the patterns is larger than 2, a new concept "an angle $\psi_{\min}$ of the solution area" was introduced, in which the weight vector solutions exist. It was numerically confirmed that the $\theta - \lambda$ relation is good approximation of $\psi_{\min} - \lambda$. Furthermore, it was theoretically proved that the Ist-order GLA with $\lambda = 2$ always converges for any number of patterns.

Considering the practical pattern classification problem, the information about the angle of the solution area isn't given generally. Therefore, the condition that the learning rate is 2 has a special significance on the GLA. From the geometrical point of view, the GLA in which the learning rate is 2 can be expressed as the algorithm in which the connection weight vector is updated to the symmetric vector with respect to the orthogonal complement of the space spanned by $k$ pattern vectors used for the update. In this way, the GLA in which the learning rate is 2 has a special significance in the convergence properties and the geometrical sense. Therefore, in this paper, the GLA with $\lambda = 2$ is discriminated as "symmetric learning algorithm (SLA)" and the convergence properties of the SLA are analyzed[6].

In [4] , [5], [7], the convergence properties of the 1st order GLA have been analyzed theoretically and numerically. In this paper, the convergence properties of the higher order cases are analyzed.

In this paper, the patterns to be classified are random vectors of whichelements are real numbers. Therefore, the patterns are on general position with probability 1. The word "convergence" means that the learning process finishes by reaching the solution area. The words "global convergence" means that the learning converges within a finite number of steps regardless of initial weights.

## II. PERCEPTRON

Figure 1 shows a perceptron proposed by Rosenblatt[8]. The operation can be described by Eqs.(1) and (2).

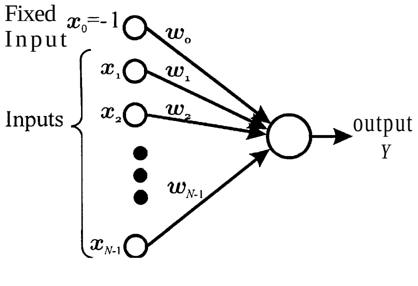

Fig. 1. perceptron.

$$U = \sum_{i=0}^{N-l} w_i x_i \tag{1}$$

$$Y = \begin{cases} +1, & u \geq 0 \\ -1, & u < 0 \end{cases} \tag{2}$$

When 0 is substituted in u of Eq. (l), this means a hyperplane of which gradient and position are determined by the connection weights $w_i$. Therefore, perceptron has the ability to discriminate the two classes which are divided by this hyperplane in the pattern space. In Fig.1, the threshold is fixed to 0 by equipping with the input $x_0$ which always takes -1 and the weight $w_0$. This type is convenient because the hyperplane includes the origin point in exchange for only 1 dimensional enlargement of input vectors. Therefore, this type of perceptron is considered in this paper. The number $N$ of inputs including this fixed input is the dimension of the patterns.

## III. Geometric Learning Algorithm (GLA) and Symmetric Learning Algorithm (SLA)

### A. Geometric Learning Algorithm

In the field of adaptive filters, the affine projection algorithm (APA)[1] is well known as a generalized algorithm of the normalized LMS algorithm[2],[3] into the block signal processing. The geometric learning algorithm (GLA) was proposed as an application of APA to perceptron for pattern classification[4],[5],[9]. In the $k$th order GLA, a weights update is done by using $k$ patterns which need to be learned, that is, the patterns which are misclassified by the current weight vector. The $k$th order GLA is described as follows:

```
[stepl] initialization :
  w (0) is randomly set;
[step2] $k_0$ is set to the number of patterns which are misclassified by the current weight vector.
[step3] judgement of convergence :
  if $k_0$ = 0 then stop
[step4]
  if $k_0 \geq k$
      $\cdot\ (\boldsymbol{x}^1, \boldsymbol{x}^2, \cdots, \boldsymbol{x}^k)^T$        (3)
  else
      $\mathbf{x} = (\boldsymbol{x}^1, \boldsymbol{x}^2, \cdots, \boldsymbol{x}^{k_0})^T$        (4)
  end if;
[step5] weight update :
      $w(n+1) = w(n) - \lambda \boldsymbol{X}^+ \boldsymbol{X} \boldsymbol{w}(n)$        (5)
[step6] return to step2 :
```

where, $\lambda$ is the learning rate. $\boldsymbol{X}^+$ means the Moore–Penrose generalized inverse of X. In Eqs. (3) and (4), $\boldsymbol{x}^1 - \boldsymbol{x}^k$ or $\boldsymbol{x}^1 - \boldsymbol{x}^{k_0}$ are selected from the patterns which are misclassified by the current weight vector. It isn't defined here how to select these patterns.

### B. Convergence Condition of 1st order GLA

Figure 2 shows the weight update process in the 1st order GLA for 2 patterns' classification. Though the learning converges in Fig.2, the GLA learning doesn't always converge within a finite number of steps for linearly separable patterns. In [4],[5],[7], the condition for the global convergence of the 1st order GLA was derived theoretically through the analysis about the condition that the learning oscillates for 2 patterns. That is, the necessary and sufficient condition that the 1st order GLA converges globally for 2 patterns is given as follows:

- if $0 < \theta \leq \frac{\pi}{2}$,

$$\frac{\tan(\frac{\pi}{4} + \frac{\theta}{2})}{\tan(\frac{\pi}{4} - \frac{\theta}{2})} + 1 > \lambda > \frac{\tan(\frac{\pi}{4} - \frac{\theta}{2})}{\tan(\frac{\pi}{4} + \frac{\theta}{2})} + 1 \tag{6}$$

- if $\frac{\pi}{2} < \theta \leq \pi$,

$$\lambda > 1 \tag{7}$$

where, $\theta$ is defined by the 2 patterns as shown in Fig.2.

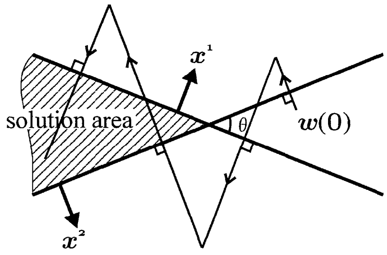

Fig. 2. Weights update process in l-GLA.

In the case of more than 2 patterns (expressed as "many patterns" in this paper), the learning process until convergence or oscillation becomes more complicated. One of the reasons is that the shape of the solution area is more complicated and another is that the order of patterns to be presented is added to the degrees of freedom. However, it was shown that the global convergence condition for many patterns almost agrees with that for 2 patterns by considering the angle $\psi_{\min}$ of the solution area defined as the minimal angle of the solution area viewed from the origin point[4],[5].

### C. Symmetric Learning Algorithm

Eqs.(6) and (7) show that the 1st order GLA converges globally for 2 patterns if the learning rate $\lambda$ is 2. Furthermore, even in the case of many patterns, it was proved that the 1st order GLA converges globally if $\lambda$ is 2[5], [7].

Considering the practical pattern classification problem, the information about the angle of the solution area isn't given generally. Therefore, the condition that the learning rate is 2 has a special significance on the GLA.

In the GLA, the weight vector is updated vertically towards the orth.ogonal complement of the space spanned by $k$ pattern vectors selected at each step. Therefore, from a geometrical point of view, the weight vector is updated to the "symmetric"vector with respect to the orthogonal complement when $\lambda$ is 2.

In this way, the GLA in which the learning rate $\lambda$ is 2 has the special significance from the viewpoint of the convergence properties and the geometrical sense. Therefore, the GLA with $\lambda = 2$ is discriminated as "symmetric learning algorithm (SLA)" .

## IV. Convergence Condition of $k$TH order SLA

In this section, the relation among the order $k$, the number $P$ of pat terns and the dimension N of patterns for the global convergence of $k$th order SLA is analyzed theoretically.

Figure 3 shows an example of pattern classification when N = 3 , $P$ = 6. We consider the classification of patterns A – F to black and white classes. The hyperplane S is perpendicular to one of the weights which divide all patterns correctly. The hyperplane S(i) is perpendicular to w(i) which is the connection weights after the ith update. w(i) has not reached the solution, area yet.

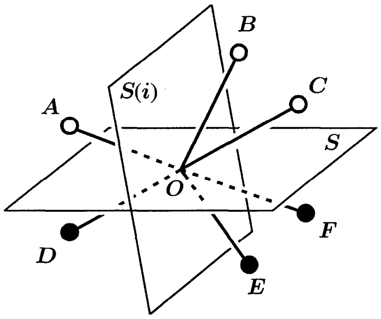

Fig. 3. Example of pattern classification ($N = 3, P = 6$).

After the ith update, there are 3 misclassified patterns, A, E, F. Considering the 3rd order SLA, these 3 patterns are used for the next update. As the dimension N is 3, the dimension of the orthogonal complement of these 3 pattern vectors is 0, that is, the weight vector is updated towards the origin point. This means that the weight vector is multiplied by -1. That is, the direction of the weight vector is changed by $\pi$ radian. As the position of S(i) doesn't change, after $(i+1)$th update, there are 3 patterns of which classes don't agree with Eqs.(1) and (2), that is, B, C, D. The next update is done in the same manner. Therefore, the learning oscillates at this position, that is, the learning doesn't converge.

In the above discussion, just the one case is considered, that is, $k$ = 3, N = 3, $P$ = 6. However, it can be shown easily that this oscillation happens when $k \geq$ N and $P \geq$ 2N. That is, $k$ < N is the necessary condition for the global convergence when $P \geq 2N$.

## V. Learning Speed of $k$TH order SLA

In this section, it is analyzed theoretically that how $k$ should be selected to maximize the learning speed of the SLA. For simplicity, the patterns used in each update are considered to be selected at random from the misclassified patterns at each step.

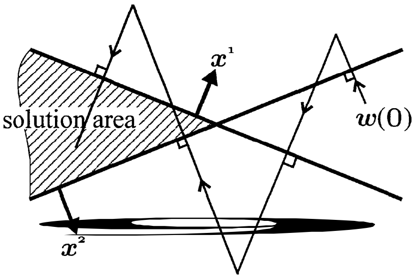

Fig. 4. Weights update process in 1–SLA.

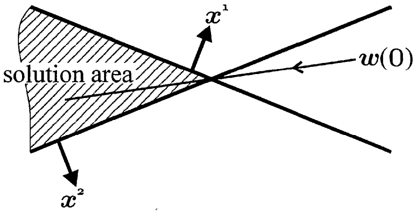

Fig. 5. Weights update process in 2-SLA.

Figures 4 and 5 are the examples of weight updates by the 1st order SLA and the 2nd order SLA, respectively. In these examples, the number $P$ of patterns is 2. The learnings from the initial weight w (0) converge in 4 updates and 1 update, respectively. Of course, the number of updates until convergence by the SLA depends on the initial weights. Therefore, Figs 4 and 5 are only certain examples for the comparison of the learning speed of the 1st and 2nd order SLAs. However, in the 1st order SLA, each update is done for the solution of the contradiction between Eqs.(1),(2) and class about only one pattern. In the 2nd order SLA, each update is done for the solution about two patterns. Accordingly, it is considered that the learning of the 2nd order SLA is faster than that of the 1st order SLA on average.

Does the learning speed grow monotonously as the order increases ? The answer is "NO". In the following, we show that the $k$th order SLA and the $(N-k)$th order SLA are equivalent for the learning speed if the number of patterns $P$ is large enough and the patterns can be considered to be distributed uniformly. This leads to the conclusion that the maximum learning speed is obtained when $k$ = N/2.

As described before, the weight vector is updated to the symmetric vector with respect to the orthogonal complement by the SLA. Using another expression, only the component in the space spanned by the $k$ patterns are reversed in the $k$th order SLA. Therefore, the following equations are valid.

$$\boldsymbol{w}(0) = \boldsymbol{w}_k^c + \boldsymbol{w}_k^p \tag{8}$$

$$\boldsymbol{w}_k(1) = \boldsymbol{w}_k^c - \boldsymbol{w}_k^p \tag{9}$$

where, $\boldsymbol{w}_k^c$ is the projection of the initial weight $\boldsymbol{w}(0)$ to the orthogonal complement of $k$ pattern vectors, $\boldsymbol{w}_k^p$ is the projection to the space spanned by $k$ pattern vectors, and $\boldsymbol{w}_k(1)$ is the weight after the 1st update.

In the same manner, only the component in the space spanned by the (N $-$ $k$) patterns are reversed in the ($N-k$)th order SLA. Therefore, the following equations are valid.

$$\boldsymbol{w}(0) = \boldsymbol{w}_{N-k}^c + \boldsymbol{w}_{N-k}^p \tag{10}$$

$$\boldsymbol{w}_{N-k}(1) = \boldsymbol{w}_{N-k}^c - \boldsymbol{w}_{N-k}^p \tag{11}$$

It is assumed that the learnings by the $k$th order SLA and the (N $-k$)th order SLA proceed in parallel from $\boldsymbol{w}(0)$ and the spaces spanned by patterns selected each step are orthogonal to each other. In this case, the following equations are valid.

$$\boldsymbol{w}_k^c = \boldsymbol{w}_{N-k}^p \tag{12}$$

$$\boldsymbol{w}_k^p = \boldsymbol{w}_{N-k}^c \tag{13}$$

Accordingly, the following is valid from Eqs. (9), (11) and Eqs.(12),(13)

$$\begin{aligned} \boldsymbol{w}_{N-k}(1) &= \boldsymbol{w}_{N-k}^c - \boldsymbol{w}_{N-k}^p \\ &= \boldsymbol{w}_k^p - \boldsymbol{w}_k^c \\ &= -\boldsymbol{w}_k(1) \end{aligned} \tag{14}$$

This equation shows that the weight vector after the 1st update by the $k$th order SLA and that by the (N $-$ $k$)th order SLA are -1 times each other, that is, the relation of reverse. Assuming that the next update is done in the same manner, the weight vectors by the $k$th order SLA and by the (N $-k$)th order SLA exactly coincide after the 2nd update. In this way, the two parallel learnings can be considered that their weight vectors repeat the agreements and the reverses.

In order to explain the theoretical discussion above, the 1st order SLA and the 2nd order SLA for 3 dimensional patterns are considered as an example. Figures 6 and 7 show the 1st and 2nd updates by those SLAs. In these figures, the pattern vectors are omitted. The relation between the complement towards which the weight is updated and the weight vectors are shown.

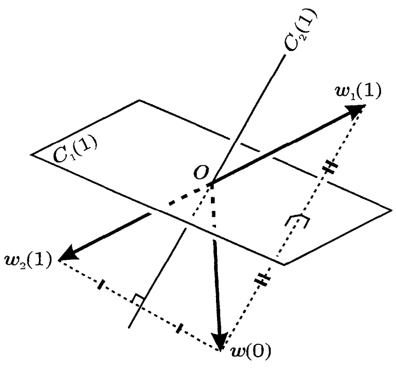

Fig. 6. l-SLA and 2-SLA (1st update).

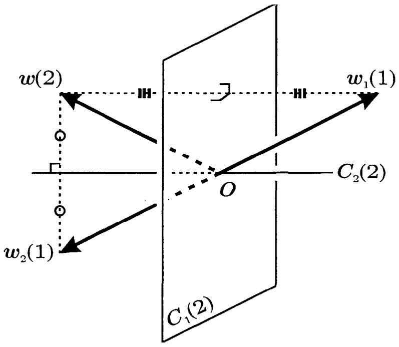

Fig. 7. l-SLA and 2-SLA (2nd update).

In Fig.6, $C_1$ (1) is the complement towards which the 1st update by the 1st order SLA is done. $C_2$( 1) is the complement towards which the 1st update by the 2nd order SLA is done. w (0) denotes the initial weight. $\boldsymbol{w}_1(1)$ and $\boldsymbol{w}_2(1)$ denote the weights after the 1st updates by the 1st order SLA and by the 2nd order SLA, respectively. Assuming that $C_1$ (1) and $C_2$ (1) are selected to be orthogonal to each other, $\boldsymbol{w}_1(1) = -\boldsymbol{w}_2(1)$ is true as shown in Fig.6.

Assuming that the complements are selected to be orthogonal at the 2nd update, the weights after the 2nd update by the 1st and 2nd order SLAs exactly agrees with each other, that is, w(2) as shown in Fig.7.

After that, assuming that the complements are selected in the same manner, the update processes by the 1st and the 2nd order SLAs repeat the agreements and the reverses.

Of course, the consideration above shows only that the $k$th order SLA and the (N $-k$)th order SLA of which update processes repeat the agreements and the reverses can exist when the number $P$ of patterns is large enough and the patterns can be considered to be distributed uniformly. The situation is different in the practical pattern classification. First, the patterns are not distributed uniformly. Second, as the patterns used at each step are selected at random, the processes of the $k$th order SLA and the (N $-k$) th order SLA don't always repeat the agreements and the reverses even if the learnings begin from the same initial weight.

However, the $k$th order SLA and the (N $-k$)th order SLA can be considered to be equivalent for the learning speed on average about many initial weights, many patterns and many orders of pattern presentations.

The above consideration can be summarized as follows.

- The higher the order is, the faster the learning is, if the order is relatively low, for example the 1st order SLA and the 2nd order SLA.
- The $k$th order SLA and the (N $-k$)th order SLA are equivalent from the viewpoint of the learning speed on average.

Consequently, it can be said that the maximal learning speed of the SLA is obtained when $k$ = N/2 on average. That is, the order $k$ of the SLA should be selected to be 1/2 of the dimension N from the viewpoint of learning speed. This is a very interesting and remarkable property of the SLA comparing with that the learning speed of the APA increases monotonously as the order increases [ 1].

## VI. Computer Simulation

### A. Order $k$ of SLA and Convergence Property

The relation between the order $k$ and the convergence properties is investigated through computer simulation. First, the linearly separable pattern sets of which the dimension N = 3, 4, 7, 10 and the number of patterns $P$ = 2 – 20 are generated. 20 sets are generated for each combination of N and $P$. Therefore, the total number of generated sets is 1520. Next, the convergence properties of the 1st$\sim P$th order SLAs are investigated. The object is to investigate the condition for the global convergence, that is, the condition for convergence regardless of the initial weights. Therefore, the convergence properties are judged by 100 trials using different random initial weights. When all the 100 trials have converged for a certain set, it is judged to be "global convergence"for the set. When the 100 trials include any that has not converged, it is judged to be "not global convergence"for the set. Each trial is judged by whether the learning converges in 1000 updates or not.

Table I gives the upper bounds of $k$ for global convergence. In this table, for example, "6" at $P$ = 18, N = 7 means that all the 20 pattern sets composed of 7 dimensional 18 patterns have converged globally by the 1st $\sim$ 6th order SLA and have not converged globally by more than the 6th order SLA. The plural numbers as the upper bound means that the upper bounds for global convergence have varied with pattern sets. For example, "8, 9, $P$" at $P$ = 15, N = 10 means that there have been 3 kinds of pattern sets in the 20 pattern sets composed of 10 dimensional 15 patterns, that is, which have converged globally by only the 1st – 8th order SLA, which have converged globally by only the 1st – 9th order SLA and which have converged globally by all the 1st – 15th order SLA.

TABLE I

Upper bound of order $k$ for SLA convergence.

| $N$<br>$P$ | 3 | 4 | 7 | 10 |
|---|---|---|---|---|
| 2 | $P$ | $P$ | $P$ | $P$ |
| 3 | $P$ | $P$ | $P$ | $P$ |
| 4 | 1, $P$ | $P$ | $P$ | $P$ |
| 5 | 2, $P$ | 2, 3, $P$ | $P$ | $P$ |
| 6 | 2 | 2, 3, $P$ | $P$ | $P$ |
|  | 2 | 3 | $P$ | $P$ |
| 8 | 2 | 3 | $P$ | $P$ |
| 9 | 2 | 3 | $P$ | $P$ |
| 10 | 2 | 3 | 5, 6, $P$ | $P$ |
| 11 | 2 | 3 | 5, 6, $P$ | $P$ |
| 12 | 2 | 3 | 5, 6, $P$ | $P$ |
| 13 | 2 | 3 | 6 | $P$ |
| 14 | 2 | 3 | 6 | 9, $P$ |
| 15 | 2 | 3 | 6 | 8, 9, $P$ |
| 16 | 2 | 3 | 6 | 7, 8, 9, $P$ |
| 17 | 2 | 3 | 6 | 8, 9, $P$ |
| 18 | 2 | 3 | 6 | 8, 9 |
| 19 | 2 | 3 | 6 | 9 |
| 20 | 2 | 3 | 6 | 9 |

At all combinations of N and $P$, though the global convergence can be obtained until a certain $k$, it can not be obtained about over a certain $k$. The global convergences have been obtained in the case of $k$ = 1 about all pattern sets. This supports the proof in [5],[7]. That is, the 1st order SLA always converges within a finite number of steps for linearly separable patterns regardless of the initial weights and the number of patterns.

This table shows the interesting relation among $k$, $P$, N and convergence. The relation includes the result of Sec.IV. That is:

- If $P \gtrsim$ 2N and $k >$_N, the learning doesn't converge globally. This result supports the analysis in Sec.IV.
- If $P \gtrsim$ 2N and $k <$ N, the learning converges globally.
- If $2 < P \lesssim N$, the learning converges globally regardless of the order $k$.
- If N $\lesssim P \lesssim 2N$, the upper bounds of $k$ for global convergence vary with pattern sets. Furthermore, there are the case in which the learning doesn't converge globally even at the small $k$ at which the learning always converges globally when 2_$< P \lesssim N$ or $P \geq 2N$.

### B. Order $k$ of SLA and Learning Speed

Figure 8 shows the average numbers of updates until convergence for $(N, P) = (7,16),(10,20)$ in order to confirm the analysis in Sec.V. In this figure, each polygonal line means the average number of updates until convergence about 100 trials with various initial weights for a certain pattern set.

This figure shows that there exist some scatter on the average number of updates until convergence. Furthermore, these figures proved the following relation between the order $k$ and the learning speed.

- The learning is slow at small $k$.
- The learning is slow at large $k$.
- The maximal learning speed is obtained at around $k = N/2$ on average.

These results support the theoretical analysis in Sec.V.

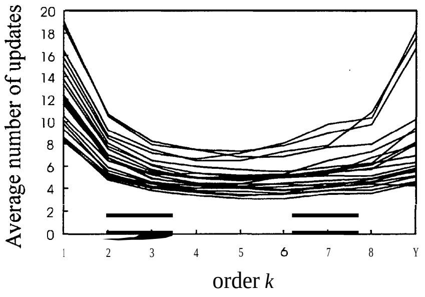

Fig. 8. . Average update number for convergence (N = 10, $P = 20$).

### C. Order $k$ of SLA and Noise Performance

In the former section, it has been proved that the learning speed varies with the order $k$. In this section, the goodness of the solution by various $k$ is investigated through computer simulations. First, random noises distributed from -0.7 to +0.7 are added to the components of the pattern vectors used in the case of $N = 10$, $P = 20$ in the former section. The components of the pattern vectors are random values distributed from -1.0 to +l.O. Next, the association rates that the noisy patterns are classified correctly are estimated by 10000 trials (100 trials using different noises for each of 100 solutions by different initial weigths). Figure 9 shows the results.

From Fig.9, it is proved that there is little difference on the association rates with $k$ variation. Analyzing this figure in detail, there is the slight tendency that the larger the order $k$ becomes, the lower the association rate becomes. Considering this tendency is slight, it can be said that there is little difference on the goodness of the solution with various $k$. From this result and the learning speed, $k = N/2$ is the best point of the SLA.

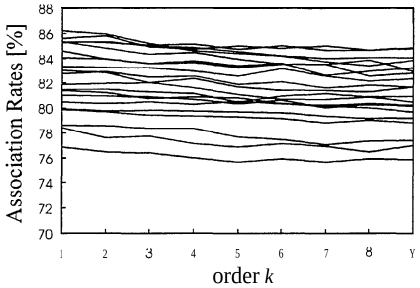

Fig. 9. Noise performance of solution (N = 10, $P = 20$).

## VII. Conclusion

The symmetric learning algorithm (SLA) is proposed as a special case of the geometric learning algorithm in which the learning rate is 2. The relation among the order $k$ of the SLA, the number $P$ of patterns, the dimension N of patterns for convergence has been analyzed theoretically. It has been proved that $k < N$ is the necessary condition for convergence when $P \geq 2N$. The relation between $k$ and the learning speed has been analyzed theoretically. It has become clear that the maximum learning speed on average can be obtained when $k = N/2$. These properties have been supported by computer simulation. Furthermore, the goodness of the solution by the SLA has, been investigated through comput er simulations. That is, there is little difference in the goodness of solution by changing the order $k$.

## References

[1] Ozeki,K. and Umeda,T., *An adaptive filtering algorithm using an orthogonal projection to an Affine subspace and its properties* (in Japanese), IECE Trans., vol.J67–A, no.2, pp.126–132, 1984.

[2] Widrow,B. and Hoff,Jr.M.E., *Adaptive switching circuits,* IRE WESCON Conv. Rec., pt.4, pp.96–104, 1960.

[3] Nagumo,J.I. and Noda,A., *A learning method for system identification,* IEEE Trans. on Automatic Control, vol.AC–12, pp.282–287, 1967.

[4] Miyoshi,S. and Nakayama,K., *A geometric learning algorithm for elementary perceptron and its convergence analysis,* Proc. 1997 IEEE Int. Conf. on Neural Networks, Houston, Texas, USA, pp.1913–1918, June 1997.

[5] Miyoshi,S., Ikeda,K. and Nakayama,K., *A geometric learning algorithm for elementary perceptron and its convergence condition* (in Japanese), IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences, in press.

[6] Miyoshi,S., Ikeda,K. and Nakayama,K., *Convergence Properties **of** Symmetric Learning Algorithm **for** Pattern Classification* (in Japanese), IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences, in press.

[7] Ikeda,K., Miyoshi,S. and Nakayama,K., *Conditions **for** convergence **of** the normalized LMS algorithm in neural learning,* Proc. Int. Symp. on Nonlinear Theory and its Applications, Honolulu, USA, pp.743–746, Nov.-Dec. 1997.

[8] Rosenblatt ,F., *The perceptron: A probabilistic model **for** information storage and organization in the brain,* Psychological Review, vol.65, pp.386–408, 1958.

[9] Hattori,M. and Hagiwara,M., *Intersection learning **for** bidirectional associative memory* (in Japanese), Trans. of the Institute of Electrical Engineers of Japan, vol.116–C, no.7, pp.755–761, 1996.