

Generalized perceptron learning rule and its implications for photorefractive neural networks

Chau-Jern Cheng, Pochi Yeh,* and Ken Yuh Hsu

Institute of Electro-Optical Engineering, National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu, Taiwan 30050, China

Received September 23, 1993; revised manuscript received April 5, 1994

We consider the properties of a generalized perceptron learning network, taking into account the decay or the gain of the weight vector during the training stages. A mathematical proof is given that shows the conditional convergence of the learning algorithm. The analytical result indicates that the upper bound of the training steps is dependent on the gain (or decay) factor. A sufficient condition of exposure time for convergence of a photorefractive perceptron network is derived. We also describe a modified learning algorithm that provides a solution to the problem of weight vector decay in an optical perceptron caused by hologram erasure. Both analytical and simulation results are presented and discussed.

INTRODUCTION

Learning is one of the most intriguing properties of the neural network. In the process of learning, both memory and information processing are involved. In a primitive network such as the perceptron¹ the network can be trained to classify a number of inputs by supervised learning with examples. The capability of information processing in a neural network is mainly determined by its structure, i.e., the arrangement of neurons, and its interconnection. Once the structure of a neural network is given, the interconnection can be obtained by learning algorithms. The perceptron with a single layer structure is a basic neural network that can be trained to sort a set of input patterns into category 1 (C1) and category 2 (C2). A trained perceptron will perform the following pattern classification when the network is interrogated with an input vector \mathbf{X} :

$$\begin{aligned} \mathbf{W} \cdot \mathbf{X} &> 0 && \text{for } \mathbf{X} \in \text{C1}, \\ \mathbf{W} \cdot \mathbf{X} &< 0 && \text{for } \mathbf{X} \in \text{C2}. \end{aligned}$$

In the learning process the interconnection weight \mathbf{W} is modified according to a simple learning algorithm given by

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \epsilon_k \eta \mathbf{X}_k, \quad (1)$$

where \mathbf{X}_k is the input pattern vector, \mathbf{W}_k is the interconnection weight vector, η is a constant representing the learning rate, and ϵ_k is an error signal denoted by

$$\epsilon_k = \begin{cases} 0 & \text{if } \mathbf{X}_k \text{ is correctly classified} \\ 1 & \text{if } \mathbf{X}_k \in \text{C1 but is misclassified} \\ -1 & \text{if } \mathbf{X}_k \in \text{C2 but is misclassified} \end{cases}, \quad (2)$$

where the subscript k is an integer ($k = 1, 2, 3, \dots$) registering the number of tests (or interrogations). We note that the interconnection weight is changed whenever there is a misclassification of an input vector. The process of weight vector update continues until all the

input patterns are correctly classified. The convergence proof of the simple perceptron can be found in Refs. 2 and 3.

A recent development in photoinduced holograms in photorefractive materials (e.g., LiNbO_3 , $\text{Sr}_x\text{Ba}_{1-x}\text{Nb}_2\text{O}_6$, and BaTiO_3)⁴ and spatial light modulators (SLM's) offers unique possibilities for the optical implementation of neural networks with a learning capability. The holograms can be recorded and erased in these media by an appropriate optical interferometric technique. This is ideal for implementing the interconnection weight, which must be modified in the process of learning. Several optical architectures for implementing the perceptronlike learning networks have been proposed and demonstrated.⁵⁻⁸ Learning is achieved by use of real-time holographic techniques in photorefractive crystals to record the modifiable interconnection weight, which is proportional to the amplitude of the photoinduced hologram. By virtue of the dynamic response of the photorefractive crystal, the hologram may decay during the learning process. The hologram decay leads to a decrease of the interconnection weight. We have shown that the holographic decay may affect the convergence property of the perceptron.⁹ Significant decay of the interconnection weight may lead to a divergence of the learning process. To overcome the hologram decay, the interconnection weight must be strengthened by optical techniques. Mathematically this is equivalent to adding a gain factor in the update equation. Although some special cases of optical perceptron with weight decay⁹⁻¹² have been studied, a general perceptron theory that addresses the issue of gain or decay of the interconnection weight is not available. To illustrate the need, we point out that the gain in the interconnection weight may also lead to a divergence of the learning process. In this paper we consider the convergence properties of a generalized perceptron algorithm with an arbitrary gain factor ξ for the weight vector. The decay is represented by $0 < \xi < 1$. An important special case is the photorefractive perceptron learning network, in which the gain factor $\xi = \exp(-t/\tau_e)$ is less than unity.

GENERALIZED PERCEPTRON LEARNING ALGORITHM

The weight update equation of a generalized perceptron algorithm can be written as

$$\mathbf{W}_{k+1} = \xi \mathbf{W}_k + \epsilon_k \eta \mathbf{X}_k, \tag{3}$$

where \mathbf{X}_k is the input vector, \mathbf{W}_k is the weight vector, and ϵ_k is an error signal at the k th interrogation, respectively. η is the learning rate, and ξ , depending on its value relative to 1, is the gain or decay factor of the perceptron network.

Following the perceptron learning algorithm and carrying out a similar analysis leading to the proof of convergence, we obtain a solution for the interconnection weight vector \mathbf{W}_{p+1} that must satisfy the following relation (see Appendix A):

$$f_1(p) \leq |\mathbf{W}_{p+1}|^2 \leq f_2(p), \tag{4}$$

where p is the total number of weight changes and

$$f_1(p) \equiv \frac{\beta^2 \eta^2}{|\mathbf{W}|^2} \left(\frac{1 - \xi^p}{1 - \xi} \right)^2, \tag{5}$$

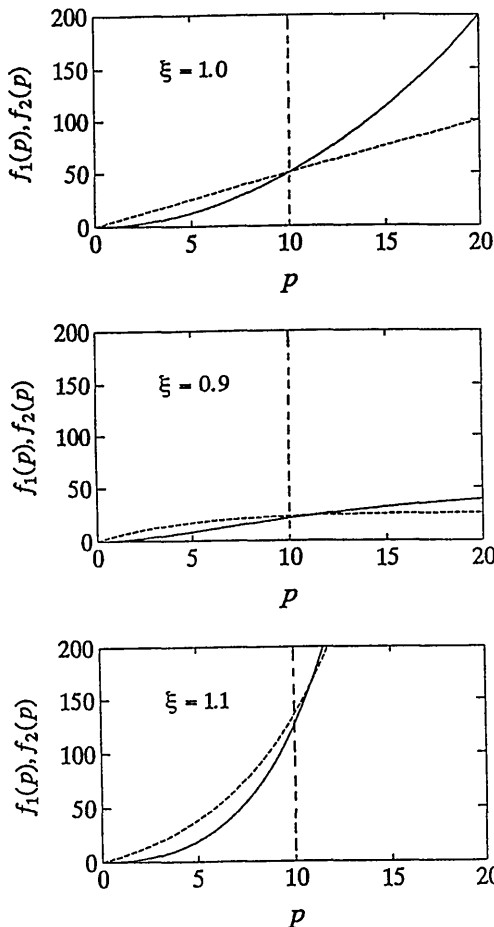


Fig. 1. Upper and lower bounds of the squared magnitude of the weight vector $|\mathbf{W}_{p+1}|^2$ versus updating steps p of the generalized perceptron learning network for (a) $\xi = 1.0$, (b) $\xi = 0.9$, and (c) $\xi = 1.1$. Solid curves, $f_1(p)$; dashed curves, $f_2(p)$. The intersection p_o provides an upper bound for p . In the example $f_1(1) = 0.5$ and $f_2(1) = 5$ (arbitrary units) were used.

$$f_2(p) \equiv \alpha \eta^2 \left(\frac{1 - \xi^{2p}}{1 - \xi^2} \right)^2, \tag{6}$$

with \mathbf{W} being an arbitrary solution and

$$\min(\mathbf{W} \cdot \mathbf{Y}_k) \equiv \beta \quad \text{for all the pattern vectors } \mathbf{Y}_k \text{ in } \mathbf{C}, \tag{7}$$

$$\max(|\mathbf{Y}_k|^2) \equiv \alpha \quad \text{for all the pattern vectors } \mathbf{Y}_k \text{ in } \mathbf{C}, \tag{8}$$

where $\mathbf{C} \equiv \mathbf{C1} \cup \mathbf{C2}$. In Eqs. (7) and (8) \mathbf{Y} is the input pattern redefined in Eq. (A3) of Appendix A, and we note that \mathbf{W}_{p+1} is, in general, different from \mathbf{W} . According to inequality (4) the squared magnitude of the weight vector \mathbf{W}_{p+1} is confined in the region bounded by $f_1(p)$ and $f_2(p)$. We recall that p is the total number of weight changes for training. The convergence of the learning process requires that an upper bound exist for the total number of updating steps p . The existence of an upper bound for p requires that a solution p_o exist that satisfies the following equation:

$$f_1(p_o) = f_2(p_o). \tag{9}$$

In other words, these two curves must intersect. The point of intersection provides an upper bound p_o for the total number of updating steps, which depends on the values of the gain factor ξ of the generalized perceptron algorithm. To understand the convergence property, we plot $f_1(p)$ and $f_2(p)$ as functions of p for various values of ξ in Fig. 1. In the example we used $f_1(1) = 0.5$ and $f_2(1) = 5$ with arbitrary units for heuristic explanation. We note that the intersection occurs at $p_o = 10$ for the case $\xi = 1$. For cases in which the gain factors $\xi = 0.9$ and $\xi = 1.1$, the point of intersection shifts toward $p > 10$. In other words, either gain or decay of the weight vector will lead to a higher upper bound for the number of weight changes.

Using Eqs. (5) and (6) and solving Eq. (9), we obtain the following expression for the upper bound p_o :

$$p_o = \log_{\xi} \left[\frac{(1 - \xi^2) - p_m(1 - \xi)^2}{(1 - \xi^2) + p_m(1 - \xi)^2} \right], \tag{10}$$

where p_m is an upper bound of the total number of weight changes of the normal perceptron algorithm² defined by

$$p_m = \frac{\alpha |\mathbf{W}|^2}{\beta^2}. \tag{11}$$

Using Eqs. (5) and (6), we also find that

$$\frac{f_2(1)}{f_1(1)} = p_m. \tag{12}$$

We note that p_m is greater than 1 and covers a wide range of numbers because of the large number of solution weights \mathbf{W} . Equation (10) shows that the upper bound p_o is a function of the parameters ξ and p_m . We also note that the maximum number of updating steps p_o of the generalized perceptron algorithm is dependent

on the gain factor (or decay rate) ξ and is independent of the learning rate η . To illustrate this, Fig. 2 shows the computer simulation results of the learning behaviors of the generalized learning algorithm described in Eq. (3) for various values of ξ and η . In the simulation we used 10 training patterns (C1: {B, 3, 5, E, 6} and C2: {A, 2, C, 4, D}); each pattern consists of 32×32 pixels) as shown in Fig. 2(g). In the learning process each iteration represents a complete cycle of interrogation of all 10 patterns. We note that, in our example, either too much gain [see Fig. 2(e), $\xi = 1.3$] or too much decay [see Fig. 2(a), $\xi = 0.85$] leads to a divergence in the learning process. Comparing Figs. 2(c) and 2(f), we also see that the convergence behaviors are identical for different values of the learning rate η .

In Fig. 3 p_o is plotted as function of the gain factor ξ for various values of p_m . We note that for a given gain factor ξ the upper bound p_o is an increasing function of p_m and the upper bound p_o of the generalized perceptron learning algorithm is always larger than that of the origi-

nal one ($\xi = 1$). That is to say, the minimum p_o occurs at $\xi = 1$, where

$$\lim_{\xi \rightarrow 1} p_o = \lim_{\xi \rightarrow 1} \log_{\xi} \left[\frac{(1 - \xi^2) - p_m(1 - \xi)^2}{(1 - \xi^2) + p_m(1 - \xi)^2} \right] = p_m. \quad (13)$$

This is the case of the normal perceptron ($\xi = 1$), which has the smallest upper bound.

According to the properties of the logarithm and analyzing Eq. (10), we find that a finite upper bound p_o for the total number of updating steps of the generalized perceptron learning algorithm exists only when ξ is bounded in a region given by

$$\xi_{\min} < \xi < \xi_{\max}, \quad (14)$$

where

$$\xi_{\max} = \frac{p_m + 1}{p_m - 1}, \quad (15)$$

$$\xi_{\min} = \frac{p_m - 1}{p_m + 1}. \quad (16)$$

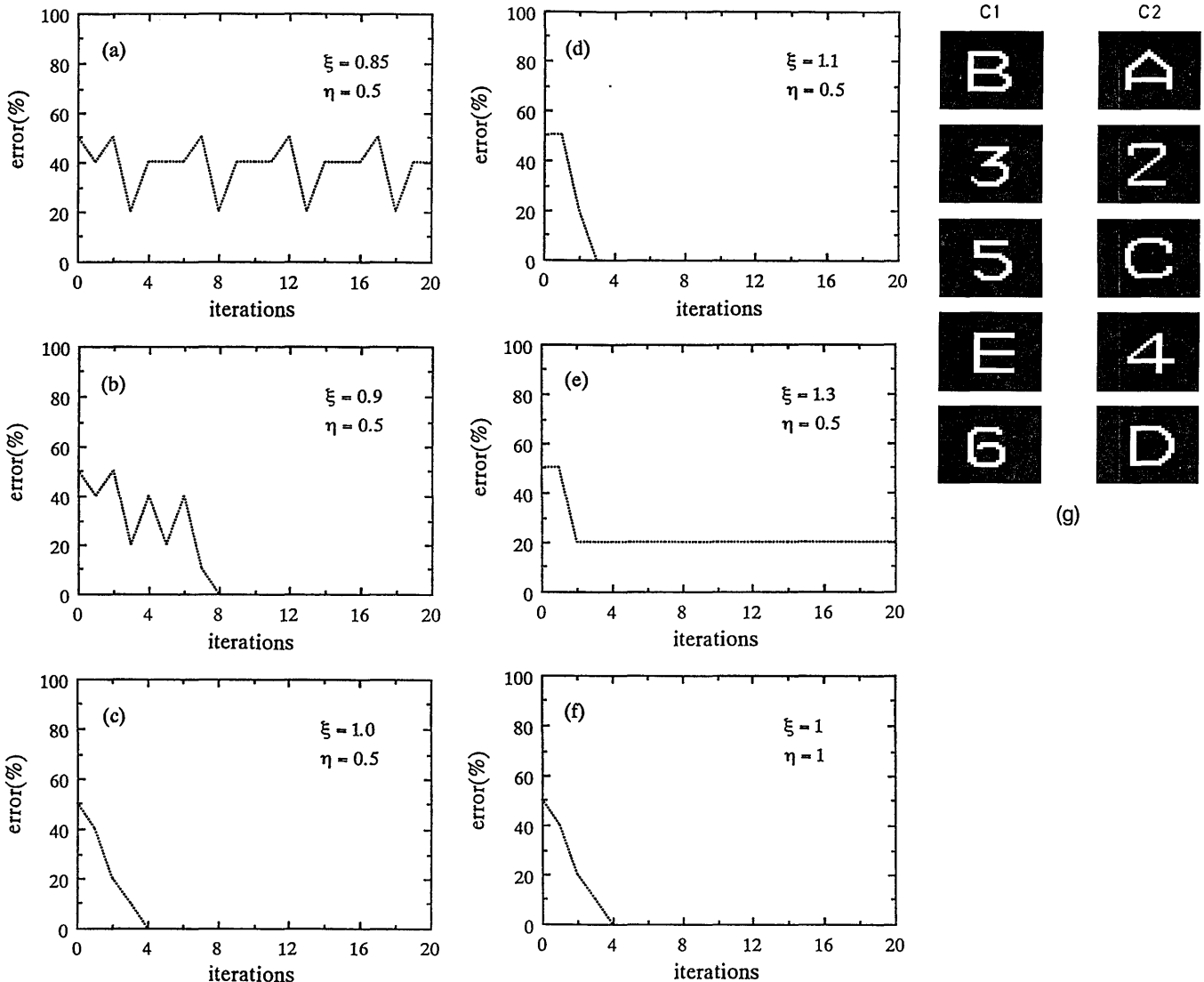


Fig. 2. Learning behaviors of a generalized perceptron with 10 input patterns (C1: {B, 3, 5, E, 6} and C2: {A, 2, C, 4, D}); pattern consists of 32×32 pixels) and for various gain factors: (a) $\xi = 0.85$; (b) $\xi = 0.9$; (c) $\xi = 1.0$; (d) $\xi = 1.1$; (e) $\xi = 1.3$, $\eta = 0.5$; (f) normal perceptron, $\xi = 1$, $\eta = 1$; (g) training patterns. Each iteration represents a complete cycle of interrogation of all 10 patterns.

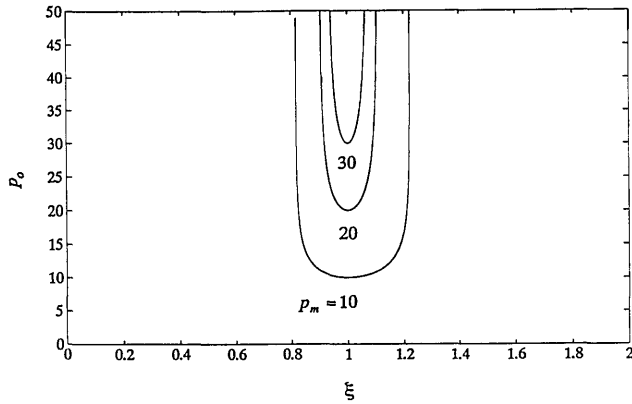


Fig. 3. Upper bound of updating steps p_o of the generalized perceptron algorithm as a function of gain factor ξ for $p_m = 10, 20, 30$ as indicated.

In other words, the generalized perceptron learning algorithm may not converge if the gain factor ξ is out of the range specified by inequality (14). The range for the gain factor is dependent on the magnitude of p_m . Table 1 lists the limits for various values of p_m . Figure 4 plots the limits as functions of p_m . The results in Fig. 4 indicate that, for a given set of input patterns with a given classification, p_m covers a wide range of values, and the minimum of all these p_m 's is useful in determining the range of the gain factor ξ . It should be pointed out that the upper bound p_m depends on knowledge of solution weight \mathbf{W} . Some values of p_m can be derived from Eq. (11), in which the weight vector \mathbf{W} is obtained by the perceptron algorithm described in Eq. (1). By definition these values of p_m are all greater than $\min(p_m)$. In addition, the results of the simulation also include the number of weight changes p for convergence. These values of p_m as well as p obtained by perceptron simulation provide information about $\min(p_m)$. Specifically,

$$p < \min(p_m) \leq p_m.$$

In our case, Fig. 2(f), we obtained $p = 7$, $p_m = \alpha|\mathbf{W}_{p+1}|^2/\beta^2 = 368$. This leads to $7 < \min(p_m) \leq 368$. Further simulation results are likely to decrease the range of possible values of $\min(p_m)$. As mentioned earlier, information about $\min(p_m)$ is useful in determining the range of ξ for convergence. For Fig. 2(f) the range of ξ for convergence is at least $0.995 < \xi < 1.005$, which we obtain by taking $p_m = 368$ as the worst case. The range of ξ for perceptron convergence is useful in the implementation in which the interconnection weight is likely to suffer gain or loss because of practical issues such as noise, electrical resistance, holographic decay, and even system imperfection. The range of ξ for convergence would allow a finite range of tolerance in the system design.

PHOTOREFRACTIVE PERCEPTRON NETWORK

A special case of interest is the photorefractive perceptron, where the gain factor (or decay factor) is given by $\xi = \exp(-t/\tau_e)$, and the learning rate is given by $\eta = 1 -$

$\exp(-t/\tau_r)$. The weight update equation is thus written as

$$\mathbf{W}_{p+1} = \exp(-t/\tau_e)\mathbf{W}_p + [1 - \exp(-t/\tau_r)]\mathbf{Y}_p, \quad (17)$$

where \mathbf{Y}_p is the input pattern, \mathbf{W}_p is the interconnection weight, and τ_r and τ_e are the rise and decay time of the photorefractive crystal, respectively. t is the exposure time for each of the weight updates. Equation (17) indicates that the learning rate η and decay rate ξ of the optical perceptron network are determined by the exposure time during the holographic recording and the response times of the photorefractive crystal. Using $\xi = \exp(-t/\tau_e)$ in Eq. (10), we obtain the following expression for the upper bound of the total number of weight changes of the photorefractive perceptron:

$$p_o = \frac{\tau_e}{t} \ln \left[\frac{1 + p_m \tanh(t/2\tau_e)}{1 - p_m \tanh(t/2\tau_e)} \right]. \quad (18)$$

According to Eq. (18), a finite upper bound requires that the exposure time of the optical perceptron network be written as

$$0 < t < \tau_e \ln \left(\frac{p_m + 1}{p_m - 1} \right). \quad (19)$$

For $p_m \gg 1$, inequality (19) can be written as

$$t p_m < 2\tau_e. \quad (20)$$

Inequality (20) may provide a guideline for the exposure time t provided that $\min(p_m)$ is known. The photorefractive perceptron will converge in a finite number of steps provided that the exposure time is limited by inequality (20). We note that inequality (20) is a sufficient condition for the convergence of a photorefractive perceptron. Although there are infinite numbers of p_m , the minimum of all these p_m 's provides the guideline for the exposure time.

Table 1. Limits of Gain Factor ξ

p_m	ξ_{\min}	ξ_{\max}
5	0.667	1.500
10	0.818	1.222
15	0.875	1.143
20	0.905	1.105

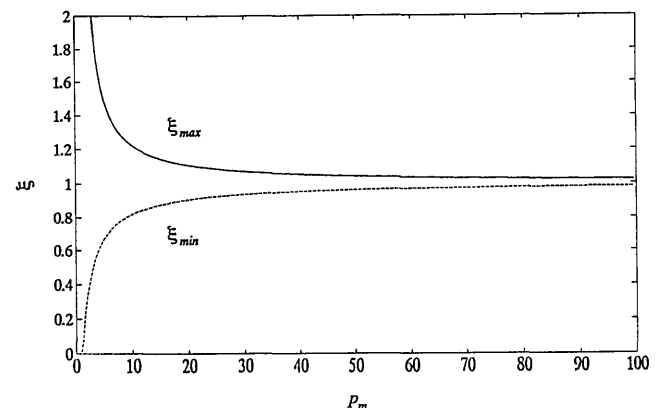


Fig. 4. Limits of gain factor ξ as functions of p_m . Solid curve, ξ_{\max} ; dashed curve, ξ_{\min} .

MODIFIED LEARNING ALGORITHM

As a result of hologram decay during the training procedure, the photorefractive perceptron network may not converge or convergence may require more iterations. This can be seen from the simulation results illustrated in Figs. 2(a)–2(f). Physically the hologram decay is consistent with the decay of the weight vector. When there is a large number of input patterns, the weight decay may become too severe for the perceptron network to converge. To overcome the weight decay in a photorefractive perceptron in the learning process, a method for weight restoration^{13,14} is required. The system can realize this by reading out the weight vector and then rerecording it optically in the same medium by using the technique of phase conjugation during each exposure for update by previous records in holographic associative memory.^{15,16} Mathematically, the equation for the weight update becomes

$$\mathbf{W}_{p+1} = \exp(-t/\tau_e)\mathbf{W}_p + [1 - \exp(-t/\tau_r)](\mathbf{Y}_p + A\mathbf{W}_p), \quad (21)$$

where A is a constant. In other words, a previous weight vector is added to the input vector to compensate for the weight decay. The equation can be rewritten as

$$\mathbf{W}_{p+1} = \{\exp(-t/\tau_e) + A[1 - \exp(-t/\tau_r)]\}\mathbf{W}_p + [1 - \exp(-t/\tau_r)]\mathbf{Y}_p. \quad (22)$$

According to the general results, inequality (14), the sufficient condition for the convergence of the perceptron becomes

$$\xi_{\min} < \exp(-t/\tau_e) + A[1 - \exp(-t/\tau_r)] < \xi_{\max}. \quad (23)$$

The constant A can be properly chosen to tailor the gain factor. For the case of a unity gain factor, i.e., $\exp(-t/\tau_e) + A[1 - \exp(-t/\tau_r)] = 1$, we obtain

$$A = \frac{1 - \exp(-t/\tau_e)}{1 - \exp(-t/\tau_r)}. \quad (24)$$

For short exposure time we obtain $A \approx \tau_r/\tau_e$. Since each constant A corresponds to a unique value of ξ , the results of simulation are in exact agreement with those obtained in Fig. 2.

CONCLUSIONS

In conclusion, we have considered the properties of a generalized perceptron learning network, taking into account a gain factor in the update of the interconnection weight. We have shown that the learning rate will not affect the convergence of the perceptron. It merely affects the scale of the input vectors. The gain factor (or decay factor) may affect the convergence of the perceptron learning process. We have derived conditions for the convergence of the learning process. The range of the gain factor (ξ) for perceptron convergence is useful in hardware implementation, where the interconnection weight may suffer gain or loss during the training stage because of practical issues, which would allow a finite range of tolerance in the system design. We have also considered the spe-

cial case of photorefractive perceptrons in which the gain factor is an exponential term accounting for holographic erasure during illumination. A feasible guideline for the exposure time of a photorefractive perceptron network is presented. A modified photorefractive perceptron with a compensation technique for overcoming the weight decay is proposed and analyzed.

APPENDIX A. CONVERGENCE PROOF OF THE GENERALIZED PERCEPTRON

We assume that the sets $C1$ and $C2$ are linearly separable and that the union of these two subsets is the complete training set C . Here, without loss of generality, we also assume that the perceptron network has a zero threshold value. That is, at least one solution weight vector \mathbf{W} exists such that

$$\mathbf{W} \cdot \mathbf{X} > 0 \quad \text{for } \mathbf{X} \in C1, \quad (A1)$$

$$\mathbf{W} \cdot \mathbf{X} < 0 \quad \text{for } \mathbf{X} \in C2. \quad (A2)$$

For the purpose of proving the convergence, we define a set of new vectors such that

$$\mathbf{Y} = \begin{cases} \mathbf{X} & \text{for } \mathbf{X} \in C1 \\ -\mathbf{X} & \text{for } \mathbf{X} \in C2 \end{cases}. \quad (A3)$$

Thus inequalities (A1) and (A2) can be written as

$$\mathbf{W} \cdot \mathbf{Y} > 0. \quad (A4)$$

The weight update equation (3) now becomes

$$\mathbf{W}_{p+1} = \xi\mathbf{W}_p + \eta\mathbf{Y}_p, \quad (A5)$$

where the pattern \mathbf{Y}_p is misclassified by the weight \mathbf{W}_p and p is now an integer registering the number of weight changes. We assume that there is no gain or decay of the weight vector during all the interrogations whose classifications are correct. In other words, $\xi = 1$, where $\epsilon_k = 0$. If we start with the initial weight $\mathbf{W}_1 = 0$, we obtain the weight vector after p updating steps:

$$\begin{aligned} \mathbf{W}_{p+1} &= \xi^{p-1}\eta\mathbf{Y}_1 + \xi^{p-2}\eta\mathbf{Y}_2 + \dots + \eta\mathbf{Y}_p \\ &= \eta \sum_{k=1}^p \xi^{p-k}\mathbf{Y}_k. \end{aligned} \quad (A6)$$

Taking the dot product of the above equation with solution weight vector \mathbf{W} , we obtain

$$\mathbf{W} \cdot \mathbf{W}_{p+1} = \eta \sum_{k=1}^p \xi^{p-k}(\mathbf{W} \cdot \mathbf{Y}_k). \quad (A7)$$

For a given solution weight vector \mathbf{W} , let β be the minimum for all the inner products, i.e.,

$$\min(\mathbf{W} \cdot \mathbf{Y}_k) \equiv \beta \quad \text{for all the pattern vectors } \mathbf{Y}_k, \quad (A8)$$

where $\beta > 0$. From Eqs. (A7) and (A8) we obtain

$$\mathbf{W} \cdot \mathbf{W}_{p+1} \geq \beta\eta \left(\frac{1 - \xi^p}{1 - \xi} \right). \quad (A9)$$

Using the Cauchy–Schwarz inequality,

$$|\mathbf{W}_{p+1}|^2 \geq \frac{(\mathbf{W} \cdot \mathbf{W}_{p+1})^2}{|\mathbf{W}|^2}, \quad (\text{A10})$$

and inequalities (A9) and (A10), we obtain

$$|\mathbf{W}_{p+1}|^2 \geq \frac{\beta^2 \eta^2}{|\mathbf{W}|^2} \left(\frac{1 - \xi^p}{1 - \xi} \right)^2. \quad (\text{A11})$$

We define a function of p as

$$f_1(p) \equiv \frac{\beta^2 \eta^2}{|\mathbf{W}|^2} \left(\frac{1 - \xi^p}{1 - \xi} \right)^2. \quad (\text{A12})$$

From Eq. (A5) we have

$$|\mathbf{W}_{k+1}|^2 = \xi^2 |\mathbf{W}_k|^2 + \eta^2 |\mathbf{Y}_k|^2 + 2\xi\eta(\mathbf{W}_k \cdot \mathbf{Y}_k). \quad (\text{A13})$$

Using $\mathbf{W}_k \cdot \mathbf{Y}_k < 0$ during training and $2\xi\eta > 0$, we obtain

$$|\mathbf{W}_{k+1}|^2 \leq \xi^2 |\mathbf{W}_k|^2 + \eta^2 |\mathbf{Y}_k|^2. \quad (\text{A14})$$

By adding these inequalities for $\mathbf{k} = 1, 2, \dots, p$, we obtain

$$|\mathbf{W}_{p+1}|^2 \leq \eta^2 \sum_{k=1}^p \xi^{2(p-k)} |\mathbf{Y}_k|^2. \quad (\text{A15})$$

Let the maximum magnitude of the pattern vectors be written as

$$\max(|\mathbf{Y}_k|^2) \equiv \alpha. \quad (\text{A16})$$

This leads to

$$|\mathbf{W}_{p+1}|^2 \leq \alpha \eta^2 \left(\frac{1 - \xi^{2p}}{1 - \xi^2} \right). \quad (\text{A17})$$

We define another function of p as

$$f_2(p) \equiv \alpha \eta^2 \left(\frac{1 - \xi^{2p}}{1 - \xi^2} \right). \quad (\text{A18})$$

Combining relations (A11), (A12), (A17), and (A18), we obtain the following relation:

$$f_1(p) \leq |\mathbf{W}_{p+1}|^2 \leq f_2(p). \quad (\text{A19})$$

Inequality (A19) indicates that the weight vector on completion of the training \mathbf{W}_{p+1} will be localized in the region bounded by $f_1(p)$ and $f_2(p)$. We recall that p is the total number of updating steps leading to a trained perceptron. The total number of steps p must satisfy the above equation. An upper bound for p exists provided that the following equation yields a finite solution p_o :

$$f_1(p_o) = f_2(p_o). \quad (\text{A20})$$

Using Eqs. (A12) and (A18) and solving Eq. (A20), we obtain the upper bound for p :

$$p_o = \log_{\xi} \left[\frac{(1 - \xi^2) - p_m(1 - \xi)^2}{(1 - \xi^2) + p_m(1 - \xi)^2} \right], \quad (\text{A21})$$

where

$$p_m = \frac{\alpha |\mathbf{W}|^2}{\beta^2}, \quad (\text{A22})$$

with p_m being the maximum number of finite updating steps of the normal perceptron algorithm.²

ACKNOWLEDGMENTS

The research is supported by the National Science Council, Taiwan, under contract NSC 83-0416-E-009-012. Pochi Yeh acknowledges the support of the K. T. Lee/K. Y. Chin Foundation.

*Permanent address, Department of Electrical and Computer Engineering, University of California, Santa Barbara, California 93106.

REFERENCES

1. F. Rosenblatt, *Principle of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms* (Spartan, Washington, D.C., 1962).
2. N. J. Nilsson, *Learning Machines: Foundation of Trainable Pattern-Classifying Systems* (McGraw-Hill, New York, 1965).
3. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis* (Wiley, New York, 1973).
4. See, for example, P. Yeh, *Introduction to Photorefractive Nonlinear Optics* (Wiley, New York, 1993).
5. D. Psaltis, D. Brady, and K. Wagner, "Adaptive optical networks using photorefractive crystals," *Appl. Opt.* **27**, 1752–1759 (1988).
6. E. G. Paek, J. Wullert, and J. S. Patel, "Holographic implementation of a learning machine based on a multicategory perceptron algorithm," *Opt. Lett.* **14**, 1303–1305 (1989).
7. J. Hong, S. Campbell, and P. Yeh, "Optical pattern classifier with perceptron learning," *Appl. Opt.* **29**, 3019–3025 (1990).
8. K. Y. Hsu, S. H. Lin, C. J. Chen, T. C. Hsieh, and P. Yeh, "An optical neural network for pattern recognition," *Int. J. Opt. Comput.* **2**, 409–423 (1991).
9. K. Y. Hsu, S. H. Lin, and P. Yeh, "Conditional convergence of photorefractive perceptron learning," *Opt. Lett.* **18**, 2135–2137 (1993).
10. J. Hertz, A. Krogh, and R. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Reading, Mass., 1991), Chap. 6, p. 157.
11. G. Parisi, "Asymmetric neural networks and the process of learning," *J. Phys. A* **19**, L675–L680 (1986).
12. J. P. Nadal, G. Toulouse, J. P. Changeux, and S. Dehaene, "Networks of formal neurons and memory palimpsests," *Europhys. Lett.* **1**, 535–542 (1986).
13. D. Brady, K. Hsu, and D. Psaltis, "Periodically refreshed multiply exposed photorefractive holograms," *Opt. Lett.* **15**, 817–819 (1990).
14. Y. Qiao and D. Psaltis, "Sampled dynamic holographic memory," *Opt. Lett.* **17**, 1376–1378 (1992).
15. B. H. Soffer, G. J. Dunning, Y. Owechko, and E. Marom, "Associative holographic memory with feedback using phase-conjugation mirrors," *Opt. Lett.* **11**, 118–120 (1986).
16. Y. Qiao, D. Psaltis, C. Gu, J. Hong, P. Yeh, and R. R. Neurgaonkar, "Phase-locked sustainment of photorefractive holograms using phase conjugation," *J. Appl. Phys.* **70**, 4648–4648 (1991).