KŌDAI MATH. SEM. REP.
19 (1967), 43—52

ON AN ADAPTIVE PROCESS FOR LEARNING FINITE PATTERNS

BY YASUICHI HORIBE

1. Introduction.

It may be possible to state that pattern recognition belongs to a broad concept of classification. When an abstract organ or system is exposed to a sequence of elements from a specified set of stimuli or patterns, one of the important features for the organ's recognition problem is the mechanism of adaptive learning of stimulus classification.

Specifically a certain trainer teaches the organ if it has correctly responded to the current stimulus. Therefore the trainer may be considered in a simple case as a function of input stimulus and its corresponding output response.

It is desirable to find a class of functions of this type which leads the organ to a successful classification of stimuli, whatever the initial state of the organ. Trained step by step under a stimulus sequence by a successfully leading trainer, the organ becomes ultimately to classify the set of stimuli correctly at least to a satisfactory degree. Hence the convergence of organ's state will be an interesting problem.

We shall consider, in this paper, for a class of linear trainers some aspects of the convergence problem, which assure the potentialities of the model formulated here.

2. Formulation and definitions.

Suppose that an *organ* is given a finite set of stimuli: $S=\{1, 2, \cdots, k+l\}$, which is pre-dichotomized into positive class S^+ and negative class S^- such that

$$S^+=\{1, \cdots, k\}, \qquad S^-=\{k+1, \cdots, k+l\}.$$

After the perception of each stimulus, it is *encoded* possibly by a random method into a binary n-sequence, i.e. a vector with n components of 0 or 1. In other words the set S is mapped into the set of all 0 or 1 component vertices of the unit hypercube in the n-dimensional Euclidean space. Therefore we have for stimulus j its code $f'_j=(\sigma_{1j}, \cdots, \sigma_{nj})$, where $\sigma_{ij}=0$ or 1, for $i=1, \cdots, n$ and $j=1, \cdots, k+l$, and f'_j is the transpose of column vector f_j.

Although elements in the code set $F=\{f_1, \cdots, f_k, f_{k+1}, \cdots, f_{k+l}\}$ are not necessarily

Received June 6, 1966.

distinct, it is not inconvenient to assume to regard them as formally different from one another, hence we may call f_j itself *stimulus* f_j, and also sets $F^+=\{f_1, \cdots, f_k\}$ and $F^-=\{f_{k+1}, \cdots, f_{k+l}\}$ may be called positive and negative classes respectively.

DEFINITION 1. *Classification function* ξ is a mapping: $F\to\{1, -1\}$ defined as follows: $\xi(f)=-1$ if $f\in F^+$ and -1 if $f\in F^-$.

DEFINITION 2. *Random stimuli* x_t at times $t=0, 1, \cdots$ are random vectors which are mutually independent, each taking values in F independently with pre-assigned probability distribution for every $t=0, 1, \cdots$.

DEFINITION 3. The *initial state* w_0 *of the organ* is an arbitrarily fixed column vector with n components.

DEFINITION 4. The *state* w_t *at time* t *of the organ* with initial state w_0 is a random column vector defined by the following recurrence relation:

$$w_{t+1}=w_t+\tau(\xi_t, \Delta_t)x_t, \tag{1}$$

where we put $\xi_t=\xi(x_t)$ and $\Delta_t=w_t'x_t-\theta$, $\tau(u, v)$ is a real valued function of two variables u and v which is called *trainer* of the organ, and θ is a fixed real number called *threshold value*.

Since we assume that the organ which is in state w_t at time interval $[t-1, t)$ receives stimulus x_t at time t, we have:

ASSUMPTION 1. w_t and x_t are independent for any $t=0, 1, \cdots$. By this assumption it is not confusing to wright (1) as:

$$w_{t+1}=w_t+\tau(\xi, w_t'x-\theta)x, \tag{1*}$$

t being omitted for ξ_t and x_t.

DEFINITION 5. The *response of the organ* consists of 1, -1, and * which are determined by the current stimulus x_t, the current state w_t, and θ as follows: 1 if $\Delta_t>0$, -1 if $\Delta_t<0$, and * if $\Delta_t=0$, where * means *don't care*.

DEFINITION 6. The *solution space* $S(\theta)$ for θ is the set of all column vectors w with n components such that $w'f>\theta$ if $f\in F^+$, and $w'f<\theta$ if $f\in F^-$. Therefore the solution space forms an open convex set determined by $k+l$ hyperplanes in n-dimensional Euclidean space.

3. A class of linear trainers.

The class of trainers $\{\tau_\alpha; \alpha>0\}$ considered in this paper is essentially owed to B. Widrow introduced in [2], having the following linear form: $\tau_\alpha(u, v)=u-\alpha v$. Hence we have from (1)* that

$$w_{t+1}=w_t+(\xi-\alpha(w_t'x-\theta))x. \tag{2}$$

The behavior of the organ in the case that $\alpha=0$, i.e. the trainer disregards the organ's responses is called *forced learning* [1].

Now let us see how the organ is trained by τ_α.

(a) The case $x \in F^+$.

Then $\xi=1$. If the organ incorrectly responds, i.e.

$$-\Delta=w_t'x-\theta<0, \quad \text{then} \quad w_{t+1}=w_t(1+\alpha\Delta)x.$$

If, however, the organ correctly responds, i.e.

$$\Delta=w_t'x-\theta>0, \quad \text{then} \quad w_{t+1}=w_t+(1-\alpha\Delta)x.$$

(b) The case $x \in F^-$.

Then $\xi=-1$. If the organ incorrectly responds, i.e.

$$\Delta=w_t'x-\theta>0, \quad \text{then} \quad w_{t+1}=w_t-(1+\alpha\Delta)x.$$

If, however, the organ correctly responds, i.e.

$$-\Delta=w_t'x-\theta<0, \quad \text{then} \quad w_{t-1}=w_t-(1-\alpha\Delta)x.$$

In either case a stronger reinforcement is performed in incorrect response rather than in correct response.

4. Martinez's necessary and sufficient condition for the convergence of the expection of w_t.

The first step for considering the validity of the trainer τ_α is to investigate the convergence of the expectation of w_t, when $t\to\infty$.

A proof of the theorem stated at the end of the section, which is due to Martinez [2], will be given here but with more succinctness.

If we define the $n\times n$ random matrix X such that its (i,j)-th element is the product of i-th and j-th components of x, we can easily rewright (2) as:

$$w_{t+1}=(\xi+\alpha\theta)x+(I-\alpha X)w_t, \tag{3}$$

where I is the identity matrix.

By assumption 1, X and w_t are obviously independent for any $t=0,1,\cdots$, so taking expectation E on both sides of (3) results in the following simple vectorial recurrence equation for m_t:

$$m_{t+1}=a+A_\alpha m_t, \qquad t=0,1,\cdots, \tag{4}$$

where we put $m_t=E(w_t)$, $a=E(\xi x+\alpha\theta x)$, $A_\alpha=I-\alpha A$, and $A=E(X)$.

Note that expectation of random matrix (including random vector) is understood as elementwise expectation. Note further than m_0 is equal to initial state w_0.

The solution of the differance equation (4) is

$$m_t=(I+A_\alpha+A_\alpha^2+\cdots+A_\alpha^{t-1})a+A_\alpha^t w_0. \tag{5}$$

We are interested in the convergence of m_t, when $t\to\infty$, regardless of what initial state is a starting point of the training.

Now we know that the necessary and sufficient condition for A_α^t to converge to zero-matrix 0 when $t\to\infty$ is that every eigenvalue of A_α is less than unity in absolute value. Hence we impose that $\max_{1\leqq i\leqq n}|\gamma_i|<1$, where γ_i, $i=1,\cdots,n$, are eigenvalues of A_α. Since $A_\alpha=I-\alpha A$, the set of equalities: $\det(A_\alpha-\gamma_i I)=0$, $i=1,\cdots,n$, may be rewritten as $\det(A-((1-\gamma_i)/\alpha)I)=0$, $i=1,\cdots,n$. Therefore $(1-\gamma_i)/\alpha$, $i=1,\cdots,n$, are eigenvalues of A. If $(1-\gamma_i)/\alpha\leqq 0$ for some i, then $\gamma_i\geqq 1$, since $\alpha>0$. Hence we necessitate the condition that all eigenvalues of A are positive, i.e. A is positive definite. In addition if we choose α such that $\alpha<2/\lambda(A)$ where $\lambda(A)$ is the largest eigenvalue of the positive definite matrix A, then we have that

$$\max_{1\leqq i\leqq n}|1-\alpha\lambda_i|=\max_{1\leqq i\leqq n}|\gamma_i|<1,$$

λ_i, $i=1,\cdots,n$, being eigenvalues of A.

Therefore the necessary and sufficient condition for $A_\alpha^t\to 0$, when $t\to\infty$, is that A is positive definite and $0<\alpha<2/\lambda(A)$.

Note that A is non-negative definite, i.e. either positive definite or positive semi-definite, since for any vector $v'=(v_1,\cdots,v_n)$, we have that $v'Av=E(\sum_{i=1}^n v_i x_i)^2\geqq 0$. A simple example shows that positive definiteness of A can not always be valid.

In conclusion we have the following theorem, since $\max_{1\leqq i\leqq n}|\gamma_i|<1$ ascertains the convergence of the series:

$$I+A_\alpha+A_\alpha^2+\cdots.$$

THEOREM 1. *When $t\to\infty$, m_t converges independently of the initial state w_0 if and only if A is positive definite and the constant α satisfies that* $0<\alpha<2/\lambda(A)$.

Henceforth we shall always be under the following assumption:

ASSUMPTION 2. The matrix A defined above is positive definite.

5. A note on an upper bound for $\lambda(A)$.

The largest eigenvalue of a non-negative matrix, i.e. the matrix whose elements are all non-negative, is the so called Frobenius root of the matrix. By the well known theorem of Frobenius [3], for non-negative matrices A_1 and A_2, it follows that $\lambda(A_1)\geqq\lambda(A_2)$ if $A_1\geqq A_2$, where $\lambda(A_1)$ and $\lambda(A_2)$ are Frobenius roots for A_1 and A_2 respectively, and the order relation for matrices is defined if the same order is preserved componentwise.

Since every element a_{ij} of A is obviously seen to satisfies that $0\leqq a_{ij}\leqq 1$, if we denote by E the $n\times n$ matrix whose elements are all unity, then we have $A\leqq E$, therefore we have that $\lambda(A)\leqq\lambda(E)=n$, since it is readily seen that

$$\det(E-\lambda I)=(-\lambda)^{n-1}(n-\lambda).$$

Note here that $A\neq E$, which will be remarked in section 7.

6. Possibility for the limiting value of m_t to be a solution, for the case of uniform probability distribution.

If we denote by m_∞ the limiting value of m_t when $t\to\infty$, then by (5) we have:

$$m_\infty=(I-A_\alpha)^{-1}a=\frac{1}{\alpha}A^{-1}a. \tag{6}$$

Consider now whether m_∞ itself is a solution, i.e. whether $m_\infty \epsilon S(\theta)$ for some θ and α, where θ is any real number and α satisfies that $0<\alpha<2/\lambda(A)$.

Only for simplicity we consider the special, yet interesting case of uniform distribution, i.e. prob$\{x=f_j\}=1/(k+l)=p$ for every $j=1,\cdots,k+l$.

Since we easily have that $E(\xi x)=pf_1+\cdots+pf_k-pf_{k+1}-\cdots-pf_{k+l}$ and $E(x)=pf_1+\cdots+pf_k+pf_{k+1}+\cdots+pf_{k+l}$, it follows that

$$a=p((1+\alpha\theta)f_1+\cdots+(1+\alpha\theta)f_k+(-1+\alpha\theta)f_{k+1}+\cdots+(-1+\alpha\theta)f_{k+l}). \tag{7}$$

If we denote by Q the $n\times(k+l)$ matrix whose j-th column is f_j, and by g_i the i-th row of Q, therefore

$$g_i=(\sigma_{i1},\cdots,\sigma_{ik},\sigma_{ik+1},\cdots,\sigma_{ik+l}),$$

then it follows that the matrix A may be of the form of

$$A=[p(g_i\cdot g_j)]=pQQ'=pG,$$

where (i,j)-th component of G is the inner product of vectors g_i and g_j. Hence the transpose of (6) results in, since A is symmetric,

$$\begin{aligned} m'_\infty&=\frac{1}{\alpha p}a'G^{-1}\\ &=\frac{1}{\alpha}(f'_1+\cdots+f'_k-f'_{k+1}-\cdots-f'_{k+l})G^{-1}+\theta(f'_1+\cdots+f'_k+f'_{k+1}+\cdots+f'_{k+l})G^{-1}. \end{aligned} \tag{8}$$

Now consider the homogeneous linear equations with $n+2$ unknowns $w_1,\cdots,w_n$, v_1, v_2:

$$\begin{cases} \sigma_{11}w_1+\cdots\cdots+\sigma_{n1}w_n+v_1=0,\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots,\\ \sigma_{1k}w_1+\cdots\cdots+\sigma_{nk}w_n+v_1=0,\\ \sigma_{1k+1}w_1+\cdots\cdots+\sigma_{1k+1}w_n+v_2=0,\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots,\\ \sigma_{1k+l}w_1+\cdots\cdots+\sigma_{1k+l}w_n+v_2=0 \end{cases} \tag{9}$$

where we put $v_1=-\theta-1/\alpha$ and $v_2=-\theta+1/\alpha$. The system (9) may be rewritten as:

$$w_1g_1+\cdots+w_ng_n=\left(\theta+\frac{1}{\alpha},\cdots,\theta+\frac{1}{\alpha},\theta-\frac{1}{\alpha},\cdots,\theta-\frac{1}{\alpha}\right).$$

If both sides of this equation are multiplied by g_i', $i=1,\cdots,n$, we have that

$$w'G=\frac{1}{\alpha}(f_1'+\cdots+f_k'-f_{k+1}'-\cdots-f_{k+l}')+\theta(f_1'+\cdots+f_k'+f_{k+1}'+\cdots+f_{k+l}')$$

hence

$$w'=\frac{1}{\alpha}(f_1'+\cdots+f_k'-f_{k+1}'-\cdots-f_{k+l}')G^{-1}+\theta(f_1'+\cdots+f_k'+f_{k+1}'+\cdots+f_{k+l}')G^{-1}, \tag{10}$$

where $w'=(w_1,\cdots,w_n)$.

LEMMA 1. *If the system* (9) *has a solution*: $\{w^*,\theta^*,1/\alpha^*\}$ *such that* $1/\alpha^*>0$, *then we have that* $w^*\in S(\theta^*)$.

Proof. Obvious.

LEMMA 2. *If the system* (9) *has a solution*: $\{w^*,\theta^*,1/\alpha^*\}$ *such that* $1/\alpha^*>0$, *then for the* θ^* *and* α^* *of this solution,* $m_\infty\in S(\theta^*)$.

Proof. The conclusion is immediate from the comparison of (10) with (8).

Denote by $\bar{Q}$ the $(k+l)\times(n+2)$ matrix formed by coefficients of (9), hence

$$\bar{Q}=\left[\begin{array}{c|cc} & 1 & 0\\ & \vdots & \vdots\\ & 1 & 0\\ Q' & 0 & 1\\ & \vdots & \vdots\\ & 0 & 1\end{array}\right]\begin{array}{l}\left.\begin{array}{c}\\ \\ \\ \end{array}\right\}k\\ \left.\begin{array}{c}\\ \\ \\ \end{array}\right\}l\end{array}$$

Note that rank $\bar{Q}\leqq n+2$.

We now examine sufficient conditions for m_∞ to be a solution for the following possible cases.

(I) The case rank $\bar{Q}=n+2$.

Then the system (9) has at most zero-solution, $w_1=\cdots=w_n=v_1=v_2=0$, therefore it is impossible to have that $1/\alpha>0$.

(II) The case rank $\bar{Q}=n+1$.

If we know the system (9) has a solution such that $v_1=s_1\neq s_2=v_2$, then we can choose α arbitrarily such that $0<\alpha<2/\lambda(A)$, and we have that $m_\infty\in S(\theta)$, where $\theta=((s_1+s_2)/(s_1-s_2))(1/\alpha)$. Otherwise it is impossible to have that $1/\alpha>0$.

(III) The case rank $\bar{Q}=r\leqq n$.

Then components v_1 and v_2 of the general solution of the system (9) may be written as:

$$\begin{cases}v_1=s_1\lambda_1+\cdots+s_{n-r+2}\lambda_{n-r+2},\\ v_2=s_1'\lambda_1+\cdots+s_{n-r+2}'\lambda_{n-r+2}\end{cases} \tag{11}$$

where λ_i, $i=1,\cdots,n-r+2$, are arbitrary, and s_i and s_i', $i=1,\cdots,n-r+2$, are fixed. We call the matrix formed by the coefficients of (11) *s-matrix*, L. Then if rank $L=2$, it is easily seen that we can choose two numbers λ and μ arbitrarily such that $\lambda-\mu>\lambda(A)$, so than we have $\alpha=2/(\lambda-\mu)$ and $\theta=(\lambda+\mu)/2$, for which $m_\infty \in S(\theta)$. But if rank $L=1$, and if $v_1=s_1 \neq s_2=v_2$, then the situation is the same as the case of (II).

In conclusion we have the following theorem:

THEOREM 2. *If the rank of s-matrix is* 2 *in the case that* rank $\bar{Q}\leqq n$, *then for arbitrarily chosen numbers* λ *and* μ *such that* $\lambda-\mu>\lambda(A)$ *we have* $m_\infty \in S(\theta)$, *where* $\alpha=2/(\lambda-\mu)$ *and* $\theta=(\lambda+\mu)/2$.

If the rank of s-matrix is 1 *in the case that* rank $\bar{Q}\leqq n$, *or if* rank $\bar{Q}=n+1$, *then the existence of the solution of the system* (9) *such that* $v_1=s_1\neq s_2=v_2$ *assures that* $m_\infty\in S(\theta)$, *where* α *is arbitrary such that* $0<\alpha<2/\lambda(A)$ *and* $\theta=((s_1+s_2)/(s_1-s_2))(1/\alpha)$.

7. Standard deviation of w_t.

Let us focus our attention on an investigation of the magnitude of the expected deviation of w_t from its mean m_t, since it is necessary to examine the possibility for w_t to become a solution.

First we assume that:

ASSUMPTION 3. prob $\{x=f_j\}>0$ for any $j=1,\cdots,k+l$.

The relation (3) may be written as: $w_t=X_\alpha w_{t-1}+z$, where $X_\alpha=I-\alpha X$, $z=(\xi+\alpha\theta)x$. Hence by Minkowskii's inequality we have that

$$(12)\qquad \sqrt{E\|w_t\|^2}\leqq\sqrt{E(\|X_\alpha w_{t-1}\|+\|z\|)^2}\leqq\sqrt{E\|X_\alpha w_{t-1}\|^2}+\sqrt{E\|z\|^2}.$$

By (12) we shall estimate $E\|w_t\|^2$, where $\|w_t\|$ means the usual norm of w_t in the n-dimensional Euclidean space.

Now the question is whether it is possible to choose a constant ρ such tnat $0<\rho<1$, so that the relation $E\|X_\alpha w_t\|^2\leqq\rho E\|w_t\|^2$ holds independently of t. For this purpose we need the following set of lemmas 3, 4, 5, 6, 7, 8.

LEMMA 3. *Let the n-dimensional random vector w take finite vector values with an assigned probability distribution, and let the* $n\times n$ *random matrix Z which is independent of w take finite matrix values also with certain probability distribution. If* $E(Z)$ *is non-negative definite, then we have that* $E(w'Zw)\geqq 0$.

Proof. Let us put prob $\{w=h_i\}=p_i\geqq 0$ for $i=1,\cdots,N$, $p_1+\cdots+p_N=1$, and also $w'=(w_1,\cdots,w_n)$. Denote by z_{ij} the (i,j)-the element of Z. Then we readily have that

$$\begin{aligned}E(w'Zw)&=E\Big(\sum_{i,j} w_iw_jz_{ij}\Big)=\sum_{i,j}E(w_iw_j)E(z_{ij})\\&=p_1\sum_{i,j}h_{i1}h_{j1}E(z_{ij})+\cdots+p_N\sum_{i,j}h_{iN}h_{jN}E(z_{ij})\\&=p_1h_1'E(Z)h_1+\cdots+p_Nh_N'E(Z)h_N\geqq 0 \qquad \text{q.e.d.}\end{aligned}$$

Denote by U the set of all n-dimensional vectors of unit length, i.e. $U=\{e; ||e||=1\}$.

LEMMA 4. *We have* $M\geqq\lambda(A)\geqq m$, *where* $M=\max_{e\in U} e'\tilde{A}e$, $m=\min_{e\in U} e'Ae$, *and* $\tilde{A}=E(X^2)$.

Proof. We first note that for symmetric matrix B the maximum and the minimum of $e'Be$ under $e\in U$, whose existences are due to the compactness of U and to the continuity of $e'Be$ on U, correspond to the maximum and the minimum eigenvalues of B respectively. Indeed, for example, from $l=\min_{e\in U} e'Be$ it follows that $l=\bar{e}'B\bar{e}$ for some $\bar{e}\in U$ and $l\leqq e'Be$ for all $e\in U$, hence $\bar{e}'(B-lI)\bar{e}=0$ and $e'(B-lI)e\geqq 0$ for all $e\in U$. Therefore $B-lI$ is positive semi-definite. Hence $B-lI$ has at least an eigenvalue equal to 0 and non-zero eigenvalues are all positive. Since for any eigenvalue λ of $B-lI$ we have $\det(B-(l+\lambda)I)=0$, l is the minimum eigenvalue of B.
Therefore it is obvious that $\lambda(A)\geqq m$. To prove that $M\geqq\lambda(A)$ it is only to show that $\tilde{A}\geqq A$, since M is the Frobenius root of $\tilde{A}$ (see section 5).

The (i, j)-th element $\tilde{a}_{ij}$ of $\tilde{A}$ may be written as: $\tilde{a}_{ij}=E(\sum_{k=1}^{n} x_i x_k x_k x_j)=E(x_i x_j r)$, where we put $r=\sum_{k=1}^{n} x_k^2=\sum_{k=1}^{n} x_k$ whose possible values are $0, 1, \cdots, n$, and $x'=(x_1, \cdots, x_n)$. Hence it is readily seen that $\tilde{a}_{ij}\geqq E(x_i x_j)=a_{ij}$, a_{ij} being the (i, j)-th element of A. Therefore we have $\tilde{A}\geqq A$. q.e.d.

LEMMA 5. *We have* $0<m<1$.

Proof. First remark that $0\leqq a_{ij}\leqq 1$, $i, j=1, \cdots, n$.

For some i we have $a_{ii}<1$. (By a similar discussion it may be proved that $0<a_{ii}$ for all $i=1, \cdots, n$.) Indeed, suppose that $a_{ii}=1$ for all $i=1, \cdots, n$, then $a_{ii}=p_i\sigma_{i1}^2+\cdots+p_{k+l}\sigma_{ik+l}^2=1$, hence by assumption 3, $\sigma_{i1}=\cdots=\sigma_{ik+l}=1$, $i=1, \cdots, n$, therefore $a_{ij}=1$, $i, j=1, \cdots, n$. We know from section 5 that A is positive semi-definite, contradicting the assumption 2. We have, therefore, for some i, $a_{ii}<1$. If we choose $(e^i)'=(0, \cdots, 0, \overset{(i)}{1}, 0, \cdots, 0)\in U$, we have $(e^i)'Ae^i=a_{ii}<1$, hence $m<1$. Since A is positive definite it is clear that $0<m$. q.e.d.

LEMMA 6. *Any element in the set of pairs* (α, ρ) *satisfying* $0<\alpha<2/\lambda(A)$ *and* $\rho\geqq M\alpha^2-2m\alpha+1$ *makes* $E(\rho I-X_\alpha^2)$ *non-negative definite.*

Proof. We have $E(\rho I-X_\alpha^2)=E(\rho I-(I-\alpha X)^2)=-(1-\rho)I+2\alpha A-\alpha^2\tilde{A}$. Let denote by ν an eigenvalue of this matrix. Then we have that for non-zero vector b, $(-(1-\rho)I+2\alpha A-\alpha^2\tilde{A})b=\nu b$. Multiplying b' on both sides from the left, we have

$$-(1-\rho)||b||^2+2\alpha(b'Ab)-\alpha^2(b'\tilde{A}b)=\nu||b||^2.$$

If we put $b/||b||=e\in U$, then the above equality reads as:

$$-(1-\rho)+2\alpha(e'Ae)-\alpha^2(e'\tilde{A}e)=\nu.$$

By lemma 4, we obtain $\nu\geqq-(1-\rho)+2m\alpha-M\alpha^2$ which holds for any eigenvalue ν

of the matrix cited above. Hence every pair (α, ρ) satisfying the following inequalities:

$$\rho \geqq M\alpha^2 - 2m\alpha + 1, \qquad 0 < \alpha < \frac{2}{\lambda(A)},$$

makes the matrix $E(\rho I - X_\alpha^2)$ non-negative definite. q.e.d.

The above statement in lemma 6 can be made a little precise by the following lemma.

The discriminant of the quadratic equation: $M\alpha^2 - 2m\alpha + 1 = 0$, is $m^2 - M$ which, by lemmas 4 and 5 is negative, and in addition it follows that $m/M < 1/\lambda(A)$. Hence be lemma 6 we have:

LEMMA 7. *If we denote by R the set of all pairs (α, ρ) satisfying $1 > \rho \geqq M\alpha^2 - 2m\alpha + 1$, then any pair $(\alpha, \rho) \in R$ makes $E(\rho I - X_\alpha^2)$ non-negative definite.*

Note that every pair $(\alpha, \rho) \in R$ satisfies that $0 < \alpha < 2/\lambda(A)$ and $0 < \rho < 1$.

LEMMA 8. *We have that $E\|X_\alpha w_t\|^2 \leqq \rho E\|w_t\|^2$ for any pair (α, ρ) in the set R.*

Proof. By lemma 3 and 7 we obtain $E(w_t'(I\rho - X_\alpha^2)w_t) \geqq 0$, hence the conclusion. q.e.d.

From lemma 8, for every pair $(\alpha, \rho) \in R$, (12) may be put into the form of $\sqrt{E\|w_t\|^2} \leqq \sqrt{\rho}\sqrt{E\|w_{t-1}\|^2} + \pi$ where $\pi = \sqrt{E\|z\|^2}$. Therefore we have

$$\sqrt{E\|w_t\|^2} \leqq (\sqrt{\rho})^t \|w_0\| + \frac{1 - (\sqrt{\rho})^t}{1 - \sqrt{\rho}} \cdot \pi.$$

When $t \to \infty$, we obtain

$$\lim_{t \to \infty} \sqrt{E\|w_t\|^2} \leqq \frac{\pi}{1 - \sqrt{\rho}}.$$

If we put $D_t = \sqrt{E\|w_t - m_t\|^2}$, then we have

$$D_t^2 = E((w_t' - m_t')(w_t - m_t)) = E\|w_t\|^2 - \|m_t\|^2.$$

In conclusion we have the following theorem:

THEOREM 3. *If the mean m_t of w_t converges independently of the initial state w_0, then for any pair (α, ρ) satisfying $1 > \rho \geqq M\alpha^2 - 2m\alpha + 1$ (M and m being defined in lemma 4), the standard deviation D_t of w_t has the following bound:*

$$D_t^2 \leqq \left[(\sqrt{\rho})^t \|w_0\| + \frac{1 - (\sqrt{\rho})^t}{1 - \sqrt{\rho}} \pi \right]^2 - \|m_t\|^2,$$

hence

$$\lim_{t\to\infty} D_t^2 \leqq \left[\frac{\pi}{1-\sqrt{\rho}}\right]^2 - \|m_\infty\|^2.$$

The author is much indebted to Professor K. Kunisawa for his help and encouragement.

References

[1] Block, H. D., The perceptron, a model for brain functioning. Reviews of Modern Physics **34** (1962), 123–135.

[2] Martinez, H. M., A convergence theorem for linear threshold elements. Bulletin of Mathematical Biophysics **27** (1965), 153–159.

[3] Nikaidô, H., Mathematical methods of modern economics. Iwanami Shoten, Tokyo. (in Japanese)

Department of Mathematics,
Tokyo Institute of Technology.