ON THE PATTERN CLASSIFICATION PROBLEMS BY LEARNING II

Tanaka, Kensuke
Department of Mathematics, Niigata University | Department of Mathematics, Kyushu University

https://doi.org/10.5109/13045

出版情報：統計数理研究. 14 (1/2), pp.61-73, 1970-03. Research Association of Statistical Sciences
バージョン：

ON THE PATTERN CLASSIFICATION PROBLEMS BY LEARNING (II)†

By

Kensuke TANAKA*

(Received February 7, 1970)

§ 1. Introduction and Summary.

In recent years, many authors have found various methods of constructing the adaptive system for the pattern classification problems. In such methods, they assume that the characteristic of the object can be represented by a single time-invariant real valued function defined on an input space, whereas the true type of the function is unknown to them. They give several methods by which the above function can be constructed on the basis of a pair of sequences of independent observed inputs, whose distributions are unknown, and of the corresponding observed outputs.

These learning theories were developed mainly by two types of approach. One is the method of potential functions by M. A. Aizerman, E. M. Braverman and L. I. Rozonoer under certain conditions ([1], [2], [3], [5]), another an application of stochastic approximation theory by Ya. Z. Tsypkin. ([3]) He tried to approximate the unknown real valued function by a linear system of linearly independent functions.

However, neither the assumption that the inputs are observed from the independent random variables, nor the one that the characteristic of the object is time-invariant does seem to us to cover the whole real situation in pattern recognition. On this reason, we make both of the above assumptions weaker in the following: (i) the inputs are observed from dependent random variables with certain conditions, (ii) the characteristic of the object representable by an unknown real valued function defined on an input space may not be time-invariant but tends to a limit function as time become infinite. In the situation, it seems natural to us that a sequence of the functions, which represent the characteristic of the object, converges to a certain function in a course of infinite learning processes.

Thus, our problem is to construct an approximation to the limit function on the basis of a pair of sequences of the inputs and the outputs. To solve this problem, on the basis of a pair of sequences of the inputs and the outputs we approxi-

* Department of Mathematics, Kyushu University, Fukuoka and Department of Mathematics, Niigata University, Niigata.

† This paper was prepared while the author was at Kyushu University on leave from Niigata University.

mate sequentially the limit function by a sequence of linear systems of linearly independent continuous functions.

Our algorithm is an application of the method introduced by T. Kitagawa [11] in the successive process of statistical control. This method, which may be called modified stochastic approximation, was investigated by V. Dupač [5] in detail.

This paper consists of six sections. In Section 2, we shall state several lemmas [12] necessary for the proofs of main results in this paper. In Section 3, we shall give the formulation of the problem. In Section 4, we shall investigate a method of approximation to the limit function on the basis of a pair of sequences of the inputs and the outputs. In Section 5, we shall be concerned with a method of approximation in the case when the outputs of the object have noise.

§ 2. Preliminaries.

In this section, three lemmas are stated without proof in order to prove main results of this paper. Let us consider an N-dimensional stochastic process $\{y^n\}_{n=1}^{\infty}$ and three sequences of non-negative real valued measurable functions $\{U_n\}_{n=1}^{\infty}$, $\{V_n\}_{n=1}^{\infty}$ and $\{\zeta_n\}_{n=1}^{\infty}$, where each U_n, V_n and ζ_n are defined on R^{nN}. Then accordingly $\{U_n(y^1, \cdots, y^n)\}_{n=1}^{\infty}$, $\{V_n(y^1, \cdots, y^n)\}_{n=1}^{\infty}$ and $\{\zeta_n(y^1, \cdots, y^n)\}_{n=1}^{\infty}$ become again three stochastic processes respectively. Let us write $U_n = U_n(y^1, \cdots, y^n)$, $V_n = V_n(y^1, \cdots, y^n)$ and $\zeta_n = \zeta_n(y^1, y^2, \cdots, y^n)$ for the sake of simplicity. We denote the expected values of three stochastic variables U_n, V_n and ζ_n by $E[U_n]$, $E[V_n]$ and $E[\zeta_n]$. Furthermore, we denote the conditional expectations of three stochastic variables U_{n+1}, V_{n+1} and ζ_{n+1} given the random variables $y^1, y^2, \cdots, y^n$ by $E[U_{n+1}|y^1, \cdots, y^n]$, $E[V_{n+1}|y^1, \cdots, y^n]$ and $E[\zeta_n|y^1, \cdots, y^n]$, respectively.

In what follows, let $\{\gamma_n\}_{n=1}^{\infty}$ and $\{\mu_n\}_{n=1}^{\infty}$ be two sequences of real numbers. Now, we introduce the fundamental conditions for three stochastic processes $\{U_n\}_{n=1}^{\infty}$, $\{V_n\}_{n=1}^{\infty}$ and $\{\zeta_n\}_{n=1}^{\infty}$:

(A1) $E[U_1]$ and $E[V_1]$ exist,

(A2) $E[U_{n+1}|y^1, \cdots, y^n] \leqq (1+\mu_n)U_n - \gamma_n V_n + \zeta_n$ hold for all n,

(A3) $\sum_{n=1}^{\infty} \gamma_n = \infty$ $(\gamma_n \geqq 0,\ n=1, 2, \cdots)$,

(A4) $\sum_{n=1}^{\infty} |\mu_n| < \infty$,

(A5) There exists a sequence of positive numbers $\{M_n\}_{n=1}^{\infty}$ such that $P[\zeta_n \leqq M_n] = 1$ for all n, and such that $\sum_{n=1}^{\infty} M_n < \infty$.

The following Lemma 1 and Lemma 2 were essentially proved in [12].

LEMMA 1. *Let the hypotheses for three stochastic processes* $\{U_n\}_{n=1}^{\infty}$, $\{V_n\}_{n=1}^{\infty}$ *and* $\{\zeta_n\}_{n=1}^{\infty}$ *be satisfied:* (i) *conditions* (A1)~(A5) *hold,* (ii) $\lim_{n\to\infty} \gamma_n = 0$, (iii) *if there exists a subsequence* $\{n_k\}_{k=1}^{\infty}$ *of a sequence* $\{n\}_{n=1}^{\infty}$ *such that* $P[\lim_{k\to\infty} V_{n_k} = 0] = 1$, *then* $P[\lim_{k\to\infty} U_{n_k} = 0] = 1$. *Then, it holds that*

$$P[\lim_{n\to\infty} U_n = 0] = 1 \quad \textit{and} \quad \lim_{n\to\infty} E[U_n^{\beta}] = 0 \qquad \textit{for all } 0 < \beta < 1 .$$

LEMMA 2. *Suppose that a sequence of non-negative real numbers* $\{a_n\}_{n=1}^{\infty}$ *satisfies* *the condition: there exist a positive integer* n_0, *two sequences of non-negative real numbers* $\{\gamma_n\}_{n=1}^{\infty}$ *and* $\{A_n\}_{n=1}^{\infty}$ *such that*

(2.1) $$a_{n+1} \leqq (1-\gamma_{n+1})a_n + A_{n+1} \quad \textit{for all } n \geqq n_0\,,$$

(2.2) $$\sum_{n=1}^{\infty} \gamma_n = \infty\,,$$

(2.3) $$\lim_{n\to\infty} \gamma_n = 0\,,$$

(2.4) $$\sum_{n=1}^{\infty} A_n < \infty\,.$$

Then, it holds that $\lim_{n\to\infty} a_n = 0$.

Next, we state the lemma which was modified by V. Dupač [9] without proof.

LEMMA 3. *Suppose that a sequence of non-negative real numbers* $\{a_n\}_{n=1}^{\infty}$ *satisfies the condition: there exist a positive integer* n_0, *two positive constants* A *and* B *such that*

(2.5) $$a_{n+1} \leqq (1-A/n^s)a_n + B/n^t \quad \textit{for all } n \geqq n_0$$

(2.6) $$t \textit{ real number and } 0 < s < 1$$

Then, it holds that

$$\limsup_{n\to\infty} n^{t-s} a_n \leqq B/A\,.$$

§ 3. The formulation of the problem.

In our pattern classification problems, the characteristic of the object is unknown, while we can observe an input and the corresponding output of the object. Now we assume that there exists a time-dependent real valued function defined on an input space, which represent the characteristic of the object at each instant, and that this function tends to a limit function as time become infinite.

Thus, our problem is to construct an approximation to the limit function on the basis of a pair of sequences of the inputs and the outputs. An outcome in this problem is denoted by a pair (x, y). The element x is an observable input and y is an observable output corresponding to x. Let X and Y denote respectively an input space and an output space. For the inputs $x^1, x^2, \cdots$, we consider a sequence

$$(x^1, y^1),\ (x^2, y^2),\ \cdots\ (x^n, y^n),\ \cdots,$$

where $x^n \in X$ and $y^n \in Y$.

Here, we assume that y^n is expressed by a function of n and $x^n: y^n = f^{(n)}(x^n)$, while the true type of this function is unknown to us. We let $\xi^n = (x^1, x^2, \cdots, x^n)$. In what follows, we shall assume that, for each n, the transition probability distribution of an observed sample value x^{n+1} at instant $n+1$ given ξ^n at instant n has the density function with respect to Lebesgue measure. We denote this transition probability density function by $p(x^{n+1}|\xi^n)$. Through this paper, we treat the case when each transition probability density is unknown to the observer.

§ 4. The construction of an unknown limit function.

In this section, we shall investigate the formulated problem in Section 3. In advance, let us take a set of systems of linearly independent continuous functions $\{\varphi_i^{(n)}(x)\}_{i=1}^N$ ($n \geqq 1$) defined on an input space X. At each instant $n+1$, we would like to approximate an unknown function $f^{(n+1)}(x^{n+1})$ by a finite series

$$\hat{f}_*^{(n+1)}(x^{n+1}\mid\xi^n)=\sum_{i=1}^N c_{i*}^{(n+1)}(\xi^n)\varphi_i^{(n+1)}(x^{n+1}) \tag{4.1}$$

which minimizes a quantity I_{n+1} defined by

$$\begin{aligned} I_{n+1} &= E[(y^{n+1}-\hat{f}^{(n+1)}(x^{n+1}\mid\xi^n))^2\mid\xi^n] \\ &= \int_X (y^{n+1}-\hat{f}^{(n+1)}(x^{n+1}\mid\xi^n))^2 p(x^{n+1}\mid\xi^n)dx^{n+1}; \end{aligned} \tag{4.2}$$

where

$$\hat{f}^{(n+1)}(x^{n+1}\mid\xi^n)=\sum_{i=1}^N c_i^{(n+1)}(\xi^n)\varphi_i^{(n+1)}(x^{n+1}) \tag{4.3}$$

and $\{c_i^{(n+1)}(\xi^n)\}_{i=1}^N$ are unknown coefficients for ξ^n. But we cannot find $\{c_{i*}^{(n+1)}(\xi^n)\}_{i=1}^N$. Hence, we shall reduce this problem to the problem of finding an algorithm by which we can construct, on the basis of a sequence of outcomes (x^1, y^1), (x^2, y^2), $\cdots$, $(x^n y^n)$, random variable $c_j^{(n)}(\xi^n)$ $(j=1, 2, \cdots, N)$ converging to $c_{j*}^{(n)}(\xi^{n-1})$ as $n\to\infty$ in some sense.

Now, in order to simplify the description we use the following notations:

$$\boldsymbol{C}^{(n)}(\xi^n)=(c_1^{(n)}(\xi^n), c_2^{(n)}(\xi^n), \cdots, c_N^{(n)}(\xi^n)), \tag{4.4}$$

$$\boldsymbol{C}_*^{(n)}(\xi^{n-1})=(c_{1*}^{(n)}(\xi^{n-1}), c_{2*}^{(n)}(\xi^{n-1}), \cdots, c_{N*}^{(n)}(\xi^{n-1})), \tag{4.5}$$

$$u_j^{(n)}=c_j^{(n)}(\xi^n)-c_{j*}^{(n)}(\xi^{n-1}), \qquad j=1, 2, \cdots, N, \tag{4.6}$$

$$U^{(n)}=(u_1^{(n)}, u_2^{(n)}, \cdots, u_N^{(n)}), \tag{4.7}$$

$$\theta_j^{(n)}=c_{j*}^{(n)}(\xi^{n-1})-c_{j*}^{(n+1)}(\xi^n), \qquad j=1, 2, \cdots, N, \tag{4.8}$$

$$\Theta^{(n)}=(\theta_1^{(n)}, \theta_2^{(n)}, \cdots, \theta_N^{(n)}), \tag{4.9}$$

$$\boldsymbol{\varphi}^{(n)}(x^n)=(\varphi_1^{(n)}(x^n), \varphi_2^{(n)}(x^n), \cdots, \varphi_N^{(n)}(x^n))', \tag{4.10}$$

where the prime of a matrix denotes its transpose. The norm $\|A\|$ of a matrix A and the inner product $\langle A, B\rangle$ of two matrices A and B are given as follows:

$$\langle A, B\rangle=\sum_{j=1}^N\sum_{k=1}^N a_{jk}b_{jk} \tag{4.11}$$

$$\|A\|^2=\langle A, A\rangle \tag{4.12}$$

where a_{jk} is the (j, k) element of matrix A and b_{jk} is the (j, k) element of matrix B.

By differentiating I_{n+1} with respect to $\boldsymbol{C}^{(n+1)}(\xi^n)$, equating the derivative to zero, we would like to obtain, under some condition, a solution vector $\boldsymbol{C}_*^{(n+1)}(\xi^n)$ such that

$$\nabla_{\boldsymbol{C}^{(n+1)}}I_{n+1}=-2E[(y^{n+1}-\boldsymbol{C}_*^{(n+1)}(\xi^n)\boldsymbol{\varphi}^{(n+1)}(x^{n+1}))\boldsymbol{\varphi}^{(n+1)}(x^{n+1})'\mid\xi^n]=0, \tag{4.13}$$

where $\nabla_{C^{(n+1)}}$ is the gradient operator with respect to $C^{(n+1)}$.

In view of the above argument, we shall construct the following algorithm with a sequence of non-negative real numbers $\{\gamma_n\}_{n=1}^{\infty}$ such that

$$\sum_{n=1}^{\infty} \gamma_n = \infty , \tag{4.14}$$

$$\sum_{n=1}^{\infty} \gamma_n^2 < \infty . \tag{4.15}$$

Firstly, using an observed input x^1 and the corresponding output y^1 to x^1 at instant 1, we construct

$$C^{(1)}(\xi^1) = C^{(0)} + \gamma_1 (y^1 - C^{(0)} \varphi^{(1)}(x^1)) \varphi^{(1)}(x^1)' ,$$

where $C^{(0)} = 0$.

Secondly, using an observed input x^2 and the corresponding output y^2 to x^2 at instant 2, we construct

$$C^{(2)}(\xi^2) = C^{(1)}(\xi^1) + \gamma_2 (y^2 - C^{(1)}(\xi^1) \varphi^{(2)}(x^2)) \varphi^{(2)}(x^2)' .$$

In general, using an observed input x^{n+1} and the corresponding output y^{n+1} to x^{n+1} at instant $n+1$, we constructe

$$C^{(n+1)}(\xi^{n+1}) = C^{(n)}(\xi^n) + \gamma_{n+1} (y^{n+1} - C^{(n)}(\xi^n) \varphi^{(n+1)}(x^{n+1})) \varphi^{(n+1)}(x^{n+1})' .$$

We state the following lemma concerning the existence of a solution vector $C_*^{(n+1)}(\xi^n)$ such that

$$\nabla_{C^{(n+1)}} I_{n+1} = 0 .$$

LEMMA 4. *Let the hypotheses be satisfied:*

(i) $p(x^{n+1} | \xi^n) > 0$ *for all* ξ^{n-1} *and all* $x^{n+1} \in X$.

(ii) $\{\varphi_i^{(n+1)}(x)\}_{i=1}^{N}$ *is a system of linearly independent continuous functions defined on* X,

(iii)

$$E[\,| f^{(n+1)}(x^{n+1}) \varphi_i^{(n+1)}(x^{n+1})| \,|\, \xi^n] < \infty \qquad \text{for all } i \text{ and } n \tag{4.16}$$

(iv)

$$E[\,| \varphi_i^{(n+1)}(x^{n+1}) \varphi_j^{(n+1)}(x^{n+1})| \,|\, \xi^n] < \infty \qquad \text{for all } i, j \text{ and } n . \tag{4.17}$$

Then, there exist a unique solution vector $C_*^{(n-1)}(\xi^n)$ *and* $k_0(\xi^n) > 0$ *such that, for any vector* $C = (c_1, \cdots, c_N)$, $\sum_{i=1}^{N} c_i^2 \neq 0$,

$$\langle C A^{(n+1)}(\xi^n), C \rangle \geqq k_0(\xi^n) \| C \|^2 > 0 , \tag{4.18}$$

where $A^{(n+1)}(\xi^n) = E[\varphi^{(n+1)}(x^{n+1}) \varphi^{(n+1)}(x^{n+1})' | \xi^n]$.

PROOF. It follows from (i) and (ii) that for any vector $C = (c_1, \cdots, c_N)$, $\sum_{i=1}^{N} c_i^2 \neq 0$,

$$\begin{aligned} \langle C A^{(n+1)}(\xi^n), C \rangle &= \sum_{i=1}^{N} \sum_{j=1}^{N} c_i \Big(\int_X \varphi_i^{(n+1)}(x^{n+1}) \varphi_j^{(n+1)}(x^{n+1}) p(x^{n+1} | \xi^n) dx^{n+1} \Big) c_j \\ &= \int_X \Big(\sum_{i=1}^{N} \sum_{j=1}^{N} \varphi_i^{(n+1)}(x^{n+1}) \varphi_j^{(n+1)}(x^{n+1}) c_i c_j \Big) p(x^{n+1} | \xi^n) dx^{n+1} \end{aligned} \tag{4.19}$$

$$= \int_X \Big(\sum_{i=1}^{N} c_i \varphi_i^{(n+1)}(x^{n+1}) \Big)^2 p(x^{n+1} | \xi^n) dx^{n+1} > 0 .$$

Since $A^{(n+1)}(\xi^n)$ is a positive definite matrix, there exist $\gamma(\xi^n) > 0$ and $R(\xi^n) > 0$ such that

$$\gamma(\xi^n) \|C\|^2 \leqq \langle C A^{(n+1)}(\xi^n), C \rangle \leqq R(\xi^n) \|C\|^2 , \tag{4.20}$$

where $\gamma(\xi^n)$ and $R(\xi^n)$ are respectively the minimum and the maximum eigenvalue of the matrix $A^{(n+1)}(\xi^n)$. Noting that $A^{(n+1)}(\xi^n)$ is a positive definite matrix, we can obtain that $A^{(n+1)}(\xi^n)$ is a non-singular matrix. According to the above fact, we have from (4.13)

$$C_*^{(n+1)}(\xi^n) = B^{(n+1)}(\xi^n)(A^{(n+1)}(\xi^n))^{-1} , \tag{4.21}$$

where $B^{(n+1)}(\xi^n) = E[f^{(n+1)}(x^{n+1}) \boldsymbol{\varphi}^{(n+1)}(x^{n+1})' | \xi^n]$. Furthermore, by letting $k_0(\xi^n) = \gamma(\xi^n)$, we have (4.18). Thus, the proof of the theorem is completed.

Concerning these two random variables $C^{(n+1)}(\xi^{n+1})$ and $C_*^{(n+1)}(\xi^n)$, we have the following theorems.

THEOREM 4.1. *Let the hypotheses be satisfied:*

(i) $p(x^{n+1} | \xi^n) > 0$ *for all* ξ^n *and all* $x^{n+1} \in X$,

(ii)

$$E[| f^{(n+1)}(x^{n+1}) \varphi_i^{(n+1)}(x^{n+1}) | \, | \xi^n] < \infty \qquad \text{for all } i \tag{4.22}$$

(iii) *for all* ξ^n, *there exist two positive constants* M *and* k_0 *such that*

$$E[| \varphi_i^{(n+1)}(x^{n+1}) \varphi_j^{(n+1)}(x^{n+1}) | \, | \xi^n] \leqq M \qquad \text{for all } i \text{ and } j , \tag{4.23}$$

$$0 < k_0 \leqq k_0(\xi^n) , \tag{4.24}$$

where $k_0(\xi^n)$ *is the minimum eigenvalue of a matrix* $A^{(n+1)}(\xi^n)$,

(iv) *there exist three positive constants* K_1, K_2 *and* K_3 *such that, for all* n,

$$\mathrm{Var}[Y^{(n+1)} | \xi^n] \leqq K_1 \|U^{(n)}\|^2 + K_2 \|\Theta^{(n)}\|^2 + K_3 , \tag{4.25}$$

where $Y^{(n+1)} = (y^{n+1} - C^{(n)}(\xi^n) \varphi^{(n+1)}(x^{n+1})) \varphi^{(n+1)}(x^{n+1})'$ *and*

$$\mathrm{Var}[Y^{(n+1)} | \xi^n] = E[\|Y^{(n+1)} - E[Y^{(n+1)} | \xi^n]\|^2 | \xi^n]$$

(v) *there exists a sequence of non-negative real numbers* $\{M_n\}_{n=1}^{\infty}$ *such that*

$$P[\|\Theta^{(n)}\|^2 \leqq \gamma_{n+1} M_n] = 1 \tag{4.26}$$

$$\sum_{n=1}^{\infty} M_n < \infty . \tag{4.27}$$

Then, it holds that

$$P[\lim_{n \to \infty} U^{(n)} = 0] = 1 , \qquad \lim_{n \to \infty} E[\|U^{(n)}\|^{2\beta}] = 0 \qquad \text{for all } 0 < \beta \leqq 1 .$$

PROOF. By the construction of $C^{(n+1)}(\xi^{n+1})$, we have

$$\begin{aligned} &C^{(n+1)}(\xi^{n+1}) - C_*^{(n+1)}(\xi^n) \\ &= C^{(n)}(\xi^n) + \gamma_{n+1}(y^{n+1} - C^{(n)}(\xi^n) \varphi^{(n+1)}(x^{n+1})) \varphi^{(n+1)}(x^{n+1})' - C_*^{(n+1)}(\xi^n) \\ &= C^{(n)}(\xi^n) - C_*^{(n)}(\xi^{n-1}) + \gamma_{n+1} E[Y^{(n+1)} | \xi^n] + \gamma_{n+1}(Y^{(n+1)} - E[Y^{(n+1)} | \xi^n]) \end{aligned} \tag{4.28}$$

$$+C_*^{(n)}(\xi^{n-1})-C_*^{(n+1)}(\xi^n)$$
$$=C^{(n)}(\xi^n)-C_*^{(n)}(\xi^{n-1})+\gamma_{n+1}(B^{(n+1)}(\xi^n)-C^{(n)}(\xi^n)A^{(n+1)}(\xi^n))$$
$$+\gamma_{n+1}(Y^{(n+1)}-E[Y^{(n+1)}|\xi^n])+C_*^{(n)}(\xi^{n-1})-C_*^{(n+1)}(\xi^n)$$
$$=C^{(n)}(\xi^n)-C_*^{(n)}(\xi^{n-1})+\gamma_{n+1}(B^{(n+1)}(\xi^n)-C_*^{(n+1)}(\xi^n)A^{(n+1)}(\xi^n)$$
$$+C_*^{(n+1)}(\xi^n)A^{(n+1)}(\xi^n)-C_*^{(n)}(\xi^{n-1})A^{(n+1)}(\xi^n)+C_*^{(n)}(\xi^{n-1})A^{(n+1)}(\xi^n)$$
$$-C^{(n)}(\xi^n)A^{(n+1)}(\xi^n))+C_*^{(n)}(\xi^{n-1})-C_*^{(n+1)}(\xi^n)+\gamma_{n+1}(Y^{(n+1)}-E[Y^{(n+1)}|\xi^n])$$
$$=(C^{(n)}(\xi^n)-C_*^{(n)}(\xi^{n-1}))(I-\gamma_{n+1}A^{(n+1)}(\xi^n))+(C_*^{(n)}(\xi^{n-1})$$
$$-C_*^{(n+1)}(\xi^n))(I-\gamma_{n+1}A^{(n+1)}(\xi^n))+\gamma_{n+1}(Y^{(n+1)}-E[Y^{(n+1)}|\xi^n]),$$

where I is an identity matrix.

The equality (4.28) can be written in terms of $U^{(n+1)}$, $U^{(n)}$, $\Theta^{(n)}$ as

(4.29)
$$U^{(n+1)}=U^n(I-\gamma_{n+1}A^{(n+1)}(\xi^n))+\Theta^{(n)}(I-\gamma_{n+1}A^{(n+1)}(\xi^n))$$
$$+\gamma_{n+1}(Y^{(n+1)}-E[Y^{(n+1)}|\xi^n]).$$

Squareing norm of both sides of (4.29) and taking conditional expectation, we have

(4.30)
$$E[\|U^{(n+1)}\|^2|\xi^n]=\|U^{(n)}(I-\gamma_{n+1}A^{(n+1)}(\xi^n))\|^2$$
$$+\|\Theta^{(n)}(I-\gamma_{n+1}A^{(n+1)}(\xi^n))\|^2+2\langle U^n(I-\gamma_{n+1}A^{(n+1)}(\xi^n)),$$
$$\Theta^{(n)}(I-\gamma_{n+1}A^{(n+1)}(\xi^n))\rangle+\gamma_{n+1}^2\,\mathrm{Var}\,[Y^{(n+1)}|\xi^n].$$

By (4.23), (4.24) and (4.25), (4.30) gives us

(4.31)
$$E[\|U^{(n+1)}\|^2|\xi^n]=\langle U^{(n)}(I-\gamma_{n+1}A^{(n+1)}(\xi^n),\ U^{(n)}(I-\gamma_{n+1}A^{(n+1)}(\xi^n))\rangle$$
$$+\langle\Theta^{(n)}(I-\gamma_{n+1}A^{(n+1)}(\xi^n),\ \Theta^{(n)}(I-\gamma_{n+1}A^{(n+1)}(\xi^n))$$
$$+2\langle U^{(n)}(I-\gamma_{n+1}A^{(n+1)}(\xi^n),\ \Theta^{(n)}(I-\gamma_{n+1}A^{(n+1)}(\xi^n))\rangle+\gamma_{n+1}^2\mathrm{Var}[Y^{(n+1)}|\xi^n]$$
$$=\|U^{(n)}\|^2-2\gamma_{n+1}\langle U^{(1)},\ A^{(n+1)}(\xi^n)U^{(n)}\rangle+\gamma_{n+1}^2\|U^{(n)}A^{(n+1)}(\xi^n)\|^2$$
$$+\|\Theta^{(n)}\|^2-2\gamma_{n+1}\langle\Theta^{(n)},\ A^{(n+1)}(\xi^n)\Theta^{(n)}\rangle+\gamma_{n+1}^2\|\Theta^{(n)}A^{(n+1)}(\xi^n)\|^2$$
$$+2\langle U^{(n)}-\gamma_{n+1}U^{(n)}A^{(n+1)}(\xi^n),\ \Theta^{(n)}-\gamma_{n+1}\Theta^{(n)}A^{(n+1)}(\xi^n)\rangle$$
$$+\gamma_{n+1}^2\,\mathrm{Var}\,[Y^{(n+1)}|\xi^n)]$$
$$\leqq\|U^{(n)}\|^2-2\gamma_{n+1}k_0\|U^{(n)}\|^2+\gamma_{n-1}^2c_1\|U^{(n)}\|^2+\|\Theta^{(n)}\|^2-2\gamma_{n+1}k_0\|\Theta^{(n)}\|^2$$
$$+\gamma_{n-1}^2c_1\|\Theta^{(n)}\|^2+2c_2\|U^{(n)}\|\|\Theta^{(n)}\|+\gamma_{n+1}^2(K_1\|U^{(n)}\|^2+K_2\|\Theta^{(n)}\|^2+K_3)$$
$$\leqq\{1-2\gamma_{n+1}k_0+\gamma_{n+1}^2(c_1+K_1)\}\|U^{(n)}\|^2+\{1+\gamma_{n+1}^2(c_1+K_2)\}\|\Theta^{(n)}\|^2$$
$$+2c_2\|U^{(n)}\|\|\Theta^{(n)}\|+\gamma_{n+1}^2K_3,$$

where c_1 and c_2 are some constants. Noting that

$$2c_2\|U^{(n)}\|\|\Theta^{(n)}\|\leqq\gamma_{n+1}k_0\|U^{(n)}\|^2+(c_2^2/k_0)(\|\Theta^{(n)}\|^2/\gamma_{n+1})$$

and using (4.31), we have

(4.32) $$E[\|U^{(n+1)}\|^2|\xi^n] \leqq \{1+\gamma_{n+1}^2(c_1+k_1)\}\|U^{(n)}\|^2-\gamma_{n+1}k_0\|U^{(n)}\|^2$$
$$+\{1+\gamma_{n-1}^2(c_1+K_2)\}\|\Theta^{(n)}\|^2+(c_2^2/k_0)(\|\Theta^{(n)}\|^2/\gamma_{n+1})+\gamma_{n-1}^2K_3 .$$

$U_{n+1}=\|U^{(n+1)}\|^2$, $V_n=\|U^{(n)}\|^2$, $\mu_n=\gamma_{n+1}^2(c_1+K_1)$ and $\zeta_n=\{1+\gamma_{n-1}^2(c_1+K_2)\}\|\Theta^{(n)}\|^2+(c_2^2/k_0)(\|\Theta^{(n)}\|^2/\gamma_{n+1})+\gamma_{n+1}^2K_3$ satisfy (A1)~(A5) and (i), (ii), (iii) of Lemma 1 by (4.14), (4.15), (4.26), (4.27) and (4.32). Therefore, by Lemma 1, it follows that

$$P[\lim_{n\to\infty} U^{(n)}=0]=1 \quad \text{and} \quad \lim_{n\to\infty} E[\|U^{(n)}\|^{2\beta}]=0 , \qquad \text{for all } 0<\beta<1 .$$

Furthermore, taking the unconditional expectation on both sides of (4.32), we have

(4.33) $$E[\|U^{(n+1)}\|^2] \leqq \{1-\gamma_{n+1}k_0+\gamma_{n+1}^2(c_1+K_1)\}E[\|U^{(n)}\|^2]$$
$$+\{1+\gamma_{n+1}^2(c_1+K_2)\}E[\|\Theta^{(n)}\|^2]+(c_2^2/k_0)(E[\|\Theta^{(n)}\|^2]/\gamma_{n+1})+\gamma_{n+1}^2K_3 .$$

$a_n=E[\|U^{(n)}\|^2]$ and $A_n=\{1+\gamma_{n+1}^2(c_1+K_2)\}E[\|\Theta^{(n)}\|^2]+(c_2^2/k_0)(E[\|\Theta^{(n)}\|^2]/\gamma_{n+1})+\gamma_{n+1}^2K_3$ satisfy (2.1), (2.2), (2.3) and (2.4) by (4.15), (4.26) and (4.27). Therefore, by Lemma 2, it follows that

$$\lim_{n\to\infty} E[\|U^{(n)}\|^2]=0$$

Thus, the proof of the theorem is completed.

We have the following theorem concerning the order of mean convergence.

THEOREM 4.2. *Suppose that the conditions* (i), (ii), (iii), (iv) *of Theorem* 4.1 *and*

(4.34) $$\gamma_n=a/n^\alpha , \qquad a>0 , \qquad \frac{1}{2}<\alpha<1 ,$$

(4.35) $$E[\|\Theta^{(n)}\|^2]=O(n^{-2\omega}) , \qquad \omega>\alpha ,$$

where the notation $f(n)=O(g(n))$ *means* $\limsup_{n\to\infty}|f(n)/g(n)|<\infty$. *Then,*

$$E[\|U^{(n)}\|^2]=\begin{cases} O(n^{-2(\omega-\alpha)}) & \text{if } \omega<\frac{3}{2}\alpha \\ O(n^{-\alpha}) & \text{if } \omega\geqq\frac{3}{2}\alpha . \end{cases}$$

PROOF. By (4.33), (4.34) and (4.35), there exist a positive integer N and three positive constants C_1, C_2, C_3 such that for all $n\geqq N$

(4.36) $$E[\|U^{(n+1)}\|^2]\leqq(1-C_1/n^\alpha)E[\|U^{(n)}\|^2]+C_2/n^{2\alpha}+C_3/n^{2\omega-\alpha} .$$

Consequently, we have for $\omega<\frac{3}{2}\alpha$

(4.37) $$E[\|U^{(n+1)}\|^2]\leqq(1-C_1/n^\alpha)E[\|U^{(n)}\|^2]+C_4/n^{2\omega-\alpha}$$

and for $\omega\geqq\frac{3}{2}\alpha$

(4.38) $$E[\|U^{(n+1)}\|^2]\leqq(1-C_1/n^\alpha)E[\|U^{(n)}\|^2]+C_5/n^{2\alpha} ,$$

where C_4 and C_5 are some constants.

Thus an application of Lemma 3 for $a_n=E[\|U^{(n)}\|^2]$ gives us the result of the theorem.

EXAMPLE 4.1. We consider an algorithm with a sequence $\{1/n\}_{n=1}^\infty$ and a model which satisfies the following conditions:

(i) an input state space X is a bounded closed subset of R^m

(ii) $f^{(1)}=f^{(2)}=\cdots=f$ is defined on X and a bounded function

(iii) $p(x^{n+1}|\xi^n)=p(x^{n+1}|x^n)>0$ for all x^{n+1} and $x^n\in X$

(iv) for some probability density $p(x)$ ($p(x)>0$ for all $x\in X$) defined on X, there are two positive numbers C and α such that for all $x^n\in X$

(4.39) $$|p(x^{n+1}|x^n)-p(x^{n+1})|\leqq C/(n+1)^{1+\alpha}.$$

Suppose that, at each instant, a system of linearly independent continuous functions $\{\varphi_i(x)\}_{i=1}^N$ defined on X is given.

In this case, by (iv), $A=E[\varphi(x)\varphi(x)']$ is a positive definite matrix and for all i and j

(4.40) $$|E[\varphi_i(x^{n+1})\varphi_j(x^{n+1})|x^n]-E[\varphi_i(x)\varphi_j(x)]|\leqq\frac{C_1}{(n+1)^{1+\alpha}}$$

(4.41) $$|E[f^{(n+1)}(x^{n+1})\varphi_i(x^{n+1})|x^n]-E[f^{(n+1)}(x)\varphi_i(x)]|\leqq\frac{C_2}{(n+1)^{1+\alpha}}$$

where C_1 and C_2 are some constants and $\varphi(x)=(\varphi_1(x),\cdots,\varphi_N(x))'$. From (4.40) and (4.41), there exist two positive constants C_3 and C_4 such that

(4.42) $$\|A^{(n+1)}(x^n)-A\|^2\leqq C_3/(n+1)^{2(1+\alpha)}$$

(4.43) $$\|B^{(n+1)}(x^n)-B\|^2\leqq C_4/(n+1)^{2(1+\alpha)},$$

where $A^{(n+1)}(x^n)=E[\varphi(x^{n+1})\varphi(x^{n+1})'|x^n]$, $B^{(n+1)}(x^n)=E[f^{(n+1)}(x^{n+1})\varphi(x^{n+1})'|x^n]$ and $B=E[f(x)\varphi(x)']$.

In view of the above argument, there exist a positive integer N_1, two positive constants k_0 and C such that for all $n\geqq N_1$

(4.44) $$\|\Theta^{(n)}\|^2=\|C_*^{(n)}(x^{n-1})-C_*^{(n+1)}(x^n)\|^2\leqq C/n^2(1+\alpha)$$

(4.45) $$0<k_0<k_0(x^n),$$

where $k_0(x^n)$ is the minimum eigenvalue of the matrix $A^{(n+1)}(x^n)=E[\varphi(x^{n+1})\varphi(x^{n+1})'|x^n]$.

Furthermore, from (4.42) and (4.43), we can easily obtain for all $n\geqq N_2$

(4.46) $$\mathrm{Var}[Y^{(n+1)}|x^n]\leqq K_1\|U^{(n)}\|^2+K_2,$$

where K_1 and K_2 are some positive constants and N_2 is some positive integer. Thus, the results of Theorem 4.1 hold.

EXAMPLE 4.2. We consider an algorithm with a sequence $\{1/n^\alpha\}_{n=1}^\infty$ $\left(\frac{1}{2}<\alpha<1\right)$ and a model which satisfies (i), (ii), (iii) in Example 4.1 and the following condition:

(v) for some probability density $p(x)$ ($p(x)>0$ for all $x\in X$) defined on X,

(4.47) $$|p(x^{n+1}|x^n)-p(x^{n+1})|^2=O(n^{-2\omega}),\qquad \omega>\alpha$$

is true for all n.

Suppose that a system of linearly independent continuous functions $\{\varphi_i(x)\}_{i=1}^N$ defined on X is given.

In this case, we can easily obtain, from the conditions of the example,

$$E[\|\Theta^{(n)}\|^2]=O(n^{-2\omega}) \tag{4.48}$$

$$0<k_0\leqq k_0(x^n) \qquad \text{for all } n\geqq N \tag{4.49}$$

and

$$\mathrm{Var}[Y^{(n+1)}\,|\,x^n]\leqq K_1\|U^{(n)}\|^2+K_2 \tag{4.50}$$

where k_0, K_1 and K_2 are some positive constants and N is a positive integer. Thus, the result of Theorem 4.2 holds.

§ 5. The construction of an unknown limit function in the presence of noise.

In this section, we shall consider the same problem as Section 4. An algorithm used by us is the same form as the algorithm in Section 4.

However, the corresponding observed output to an input x^n at each instant n is

$$\bar{y}^{(n)}=y^n+\varepsilon^n\,, \tag{5.1}$$

where a random variable ε^n depends on only x^n.

This random variable ε^n is called a noise random variable. Now, according to the argument of Section 4, we use the following algorithm with a sequence of non-negative numbers $\{\gamma_n\}_{n=1}^{\infty}$ such that

$$\sum_{n=1}^{\infty}\gamma_n=\infty \tag{5.2}$$

$$\sum_{n=1}^{\infty}\gamma_n^2<\infty\,. \tag{5.3}$$

Firstly, using an observed input x^1 and the corresponding an observed output $\bar{y}^{(1)}=y^1+\varepsilon^1$ at instant 1, we construct

$$\boldsymbol{C}^{(1)}(\xi^1,\alpha^1)=\boldsymbol{C}^{(0)}+\gamma_1(\bar{y}^{(1)}-\boldsymbol{C}^{(0)}\varphi^{(1)}(x^1))\varphi^{(1)}(x^1)'\,,$$

where $\boldsymbol{C}^{(0)}\equiv 0$, $\xi^1=x^1$ and $\alpha^1=\varepsilon^1$.

Secondly, using an observed input x^2 and the corresponding an observed output $\bar{y}^{(2)}=y^2+\varepsilon^2$ at instant 2, we construct

$$\boldsymbol{C}^{(2)}(\xi^2,\alpha^2)=\boldsymbol{C}^{(1)}(\xi^1,\alpha^1)+\gamma_2(\bar{y}^{(2)}-\boldsymbol{C}^{(1)}(\xi^1,\alpha^1)\varphi^{(2)}(x^2))\varphi^{(2)}(x^2)'\,,$$

where $\xi^2=(x^1,x^2)$ and $\alpha^2=(\varepsilon^1,\varepsilon^2)$.

In general, using an observed input x^{n+1} and the corresponding an observed output $\bar{y}^{(n+1)}$ at instant $n+1$, we construct

$$\boldsymbol{C}^{(n+1)}(\xi^{n+1},\alpha^{n+1})=\boldsymbol{C}^{(n)}(\xi^n,\alpha^n)+\gamma_{n+1}(\bar{y}^{(n+1)}-\boldsymbol{C}^{(n)}(\xi^n,\alpha^n)$$
$$\cdot\varphi^{(n+1)}(x^{n+1}))\varphi^{(n+1)}(x^{n+1})'$$

where $\xi^{n+1}=(x^1,x^2,\cdots,x^{n+1})$ and $\alpha^{n+1}=(\varepsilon^1,\varepsilon^2,\cdots,\varepsilon^{n+1})$.

Concerning these two random variables $\boldsymbol{C}_*^{(n)}(\xi^{n-1})$ and $\boldsymbol{C}^{(n)}(\xi^n,\alpha^n)$ we have the following theorem.

THEOREM 5.1. *Let the hypotheses be satisfied:*

(i) $p(x^{n+1}|\xi^n)>0$ *for each* ξ^n *and all* $x^{n+1}\in X$

(ii)

(5.4) $$E[|f^{(n+1)}(x^{n+1})\varphi_i^{(n+1)}(x^{n+1})||\xi^n]<\infty \qquad \text{for all } i \text{ and } n\,,$$

(iii) *for all* ξ^n, *there exist two positive numbers* k_0 *and* M *such that*

(5.5) $$E[|\varphi_i^{(n+1)}(x^{n+1})\varphi_j^{(n+1)}(x^{n+1})||\xi^n]\leqq M \qquad \text{for all } i \text{ and } j$$

(5.6) $$0<k_0\leqq k_0(\xi^n) \qquad \text{for all } \xi^n$$

where $k_0(\xi^n)$ *is the minimum eigenvalue of a matrix* $A^{(n+1)}(\xi^n)$, (iv) *there exist three non-negative constants* K_1, K_2 *and* K_3 *such that*

(5.6) $$\mathrm{Var}[Y^{(n+1)}|\xi^n,\alpha^n]\leqq K_1\|U^{(n)}\|^2+K_2\|\Theta^{(n)}\|^2+K_3\,,$$

where $U^{(n)}=\boldsymbol{C}^{(n)}(\xi^n,\alpha^n)-\boldsymbol{C}_*^{(n)}(\xi^{n-1})$, $\Theta^{(n)}=\boldsymbol{C}_*^{(n)}(\xi^{n-1})-\boldsymbol{C}_*^{(n+1)}(\xi^n)$, $Y^{(n+1)}=(y^{n+1}-\boldsymbol{C}^{(n)}(\xi^n,\alpha^n)\varphi^{(n+1)}(x^{n+1}))\varphi^{(n+1)}(x^{n+1})'$ *and* $\mathrm{Var}[Y^{(n+1)}|\xi^n,\alpha^n]=E[\|Y^{(n+1)}-E[Y^{(n+1)}|\xi^n,\alpha^n]\|^2|\xi^n,\alpha^n]$.

(v) *for each noise random variable* ε^n

(5.8) $$E[\varepsilon^n|x^n]=0$$

and there is a positive constant K *such that*

(5.9) $$E[(\varepsilon^n)^2|x^n]\leqq K\,.$$

(vi) *there is a sequence of non-negative real numbers* $\{M_n\}_{n=1}^{\infty}$ *such that*

(5.10) $$P[\|\Theta^{(n)}\|^2\leqq\gamma_{n+1}M_n]=1$$

(5.11) $$\sum_{n=1}^{\infty}M_n<\infty\,.$$

Then,

$$P[\lim_{n\to\infty}U^{(n)}=0]=1\,,\qquad \lim_{n\to\infty}E[\|U^{(n)}\|^{2\beta}]=0 \qquad \text{for all } 0<\beta\leqq 1\,.$$

PROOF. By the construction of $\boldsymbol{C}^{(n+1)}(\xi^{n+1},\alpha^{n+1})$, we have

(5.12)
$$\begin{aligned}
&\boldsymbol{C}^{(n+1)}(\xi^{n+1},\alpha^{n+1})-\boldsymbol{C}_*^{(n+1)}(\xi^n)\\
&=\boldsymbol{C}^{(n)}(\xi^n,\alpha^n)-\boldsymbol{C}_*^{(n+1)}(\xi^n)+\gamma_{n+1}(\bar{y}^{(n+1)}-\boldsymbol{C}^{(n)}(\xi^n,\alpha^n)\varphi^{(n+1)}(x^{n+1}))\varphi^{(n+1)}(x^{n+1})'\\
&=\boldsymbol{C}^{(n)}(\xi^n,\alpha^n)-\boldsymbol{C}_*^{(n)}(\xi^{n-1})+\gamma_{n+1}(y^{(n+1)}-\boldsymbol{C}^{(n)}(\xi^n,\alpha^n)\varphi^{(n+1)}(x^{n+1})\\
&\quad+\varepsilon^{n+1})\varphi^{(n+1)}(x^{n+1})'+\boldsymbol{C}_*^{(n)}(\xi^{n-1})-\boldsymbol{C}_*^{(n+1)}(\xi^n)\\
&=\boldsymbol{C}^{(n)}(\xi^n,\alpha^n)-\boldsymbol{C}_*^{(n)}(\xi^{n-1})+\gamma_{n+1}(B^{(n+1)}(\xi^n)-\boldsymbol{C}^{(n)}(\xi^n,\alpha^n)A^{(n+1)}(\xi^n))\\
&\quad+\boldsymbol{C}_*^{(n)}(\xi^{n-1})-\boldsymbol{C}_*^{(n+1)}(\xi^n)+\gamma_{n+1}(Y^{(n+1)}-E[Y^{(n+1)}|\xi^n,\alpha^n]\\
&\quad+\varepsilon^{n+1}\varphi^{(n+1)}(x^{n+1})')\\
&=\boldsymbol{C}^{(n)}(\xi^n,\alpha^n)-\boldsymbol{C}_*^{(n)}(\xi^{n-1})+\gamma_{n+1}(B^{(n+1)}(\xi^n)-\boldsymbol{C}_*^{(n+1)}(\xi^n)A^{(n+1)}(\xi^n)\\
&\quad+\boldsymbol{C}_*^{(n+1)}(\xi^n)A^{(n+1)}(\xi^n)-\boldsymbol{C}_*^{(n)}(\xi^{n-1})A^{(n+1)}(\xi^n)+\boldsymbol{C}_*^{(n)}(\xi^{n-1})A^{(n+1)}(\xi^n)\\
&\quad-\boldsymbol{C}^{(n)}(\xi^n,\alpha^n)A^{(n+1)}(\xi^n))+\boldsymbol{C}_*^{(n)}(\xi^{n-1})-\boldsymbol{C}_*^{(n+1)}(\xi^n)\\
&\quad+\gamma_{n+1}(Y^{(n+1)}-E[Y^{(n+1)}|\xi^n,\alpha^n]+\varepsilon^{n+1}\varphi^{(n+1)}(x^{n+1})')
\end{aligned}$$

$$=(C^{(n)}(\xi^n, \alpha^n)-C_*^{(n)}(\xi^{n-1}))(I-\gamma_{n+1}A^{(n+1)}(\xi^n))+(C_*^{(n)}(\xi^{n-1})$$
$$-C_*^{(n-1)}(\xi^n))(I-\gamma_{n+1}A^{(n+1)}(\xi^n))+\gamma_{n+1}(Y^{(n+1)}-E[Y^{(n+1)}|\xi^n, \alpha^n]$$
$$+\varepsilon^{n+1}\varphi^{(n+1)}(x^{n+1})')$$

where I is an identity matrix.

The equality (5.12) can be written in terms of $U^{(n+1)}$, $U^{(n)}$ $\Theta^{(n)}$ as

$$U^{(n+1)}=U^{(n)}(I-\gamma_{n+1}A^{(n+1)}(\xi^n))+\Theta^{(n)}(I-\gamma_{n+1}A^{(n+1)}(\xi^n))$$
$$+\gamma_{n+1}(Y^{(n+1)}-E[Y^{(n+1)}|\xi^n, \alpha^n]+\varepsilon^{n+1}\varphi^{(n+1)}(x^{n+1})'). \tag{5.13}$$

Squareing norm of both sides of (5.13) and taking conditional expectation, we have

$$E[\|U^{(n+1)}\|^2|\xi^n, \alpha^n]=\|U^{(n)}(I-\gamma_{n+1}A^{(n+1)}(\xi^n))\|^2$$
$$+\|\Theta^{(n)}(I-\gamma_{n+1}A^{(n+1)}(\xi^n)\|^2+2\langle U^{(n)}(I-\gamma_{n+1}A^{(n+1)}(\xi^n)),$$
$$\Theta^{(n)}(I-\gamma_{n+1}A^{(n+1)}(\xi^n))\rangle+2\gamma_{n+1}^2(\mathrm{Var}[Y^{(n+1)}|\xi^n, \alpha^n]$$
$$+E[\|\varepsilon^{(n+1)}\varphi^{(n+1)}(x^{n+1})'\|^2|\xi^n, \alpha^n]), \tag{5.14}$$

where the expectation is taken with respect to the probability distribution of $(x^{n+1}, \varepsilon^{n+1})$. By (5.5), (5.6), (5.7), (5.8) and (5.9), we can easily obtain

$$E[\|U^{(n+1)}\|^2|\xi^n, \alpha^n]\leqq(1-2\gamma_{n+1}k_0+\gamma_{n+1}^2c_1)\|U^{(n)}\|^2$$
$$+(1+\gamma_{n+1}^2c_1)\|\Theta^{(n)}\|^2+2c_2\|U^{(n)}\|\|\Theta^{(n)}\|$$
$$+\gamma_{n+1}^2(K_1\|U^{(n)}\|^2+K_2\|\Theta^{(n)}\|^2+K_3)+\gamma_{n+1}^2c_3, \tag{5.15}$$

where c_1, c_2 and c_3 are some constants.

Noting that

$$2c_2\|U^{(n)}\|\|\Theta^{(n)}\|\leqq\gamma_{n+1}k_0\|U^{(n)}\|^2+(c_2^2/k_0)(\|\Theta^{(n)}\|^2/\gamma_{n+1})$$

and using (5.15), we have

$$E[\|U^{(n+1)}\|^2|\xi^n, \alpha^n]\leqq\{1+\gamma_{n+1}^2(c_1+K_1)\}\|U^{(n)}\|^2-\gamma_{n+1}k_0\|U^{(n)}\|^2$$
$$+\{1+\gamma_{n+1}^2(c_1+K_1)\}\|\Theta^{(n)}\|^2+(c_2^2/k_0)(\|\Theta^{(n)}\|^2/\gamma_{n+1})+\gamma_{n+1}^2(K_3+c_3). \tag{5.16}$$

$U_{n+1}=\|U^{(n+1)}\|^2$, $V_n=\|U^{(n)}\|^2$, $\mu_n=\gamma_{n+1}^2(c_1+K_1)$ and $\zeta_n=\{1+\gamma_{n+1}^2(c_1+K_1)\}\|\Theta^{(n)}\|^2+(c_2^2/k_0)(\|\Theta^{(n+1)}\|^2/\gamma_{n+1})+\gamma_{n+1}^2(K_3+c_3)$ satisfy (A1)~(A5) and (i), (ii), (iii) of Lemma 1 by (5.2), (5.3), (5.10) and (5.11). Therefore, by Lemma 1, it follows that

$$P[\lim_{n\to\infty}U^{(n)}=0]=1 \quad \text{and} \quad \lim_{n\to\infty}E[\|U^{(n)}\|^{2\beta}]=0 \quad \text{for all } 0<\beta<1.$$

Furthermore, taking the unconditional expectation on both sides of (5.16), we have

$$E[\|U^{(n+1)}\|^2]\leqq\{1-\gamma_{n+1}k_0+\gamma_{n+1}^2(c_1+K_1)\}E[\|U^{(n)}\|^2]$$
$$+\{1+\gamma_{n+1}^2(c_1+K_2)\}E[\|\Theta^{(n)}\|^2]+(c_2^2/k_0)(E\|\Theta^{(n)}\|^2/\gamma_{n+1})$$
$$+\gamma_{n+1}^2(K_3+c_3). \tag{5.17}$$

$a_n=E[\|U^{(n)}\|^2]$ and $A_n=\{1+\gamma_{n+1}^2(c_1+K_2)\}E[\|\theta^{(n)}\|^2]+(c_2^2/k_0)(E[\|\theta^{(n)}\|^2/\gamma_{n+1})+\gamma_{n+1}^2(K_3+c_3)$ satisfy (2.1), (2.2), (2.3) and (2.4) by (5.3), (5.10) and (5.11).

Therefore, by Lemma 2, it follows that

$$\lim_{n\to\infty} E[\|U^{(n)}\|^2]=0 .$$

Thus, the proof of the theorem is completed.

REMARK. We can easily obtain the similar result to Theorem 4.2 concerning the order of convergence with respect to $E[\|U^{(n)}\|^2]$.

§ 6. Acknowledgment.

The author is deeply indebted to Professor T. Kitagawa for his helpful advices and critical readings of the manuscript. The author is also grateful to Professors S. Kanō and N. Furukawa for their advices and encouragements.

References

[1] AIZERMAN, M. A., BRAVERMAN, E. M. and ROZONOER, L. I.: *Theoretical foundations of the potential function method in pattern recognition learning*, Autom. and Remote Control., vol. 25, No. 6, (1964), 821-837.

[2] AIZERMAN, M. A., BRAVERMAN, E. M. and ROZONOER, L. I.: *The probability problem of pattern recognition learning and the method of potential functions*, Autom. and Remote Control., Vol. 25, No. 9, (1964), 1175-1190.

[3] AIZERMAN, M. A., BRAVERMAN, E. M. and ROZONOER, L. I.: *The method of potential functions for the problem of restoring the characteristic of a function converter from randomly observed points*, Autom. and Remote Control., Vol. 25, No. 12, (1964), 1546-1556.

[4] BLUM, J. R.: *Multidimensonal stochastic approximation methods*, Ann. Math. Stat., Vol. 25, (1954), 734-744.

[5] BRAVERMAN, E. M.: *On the method of potential functions*, Autom. and Remote Control, Vol. 26, No. 12, (1965), 2130-2138.

[6] Браверман, Э. М., и Розоноер, Л. И.: Сходимость случайных процессов в теории обучения машин. 1, Автоматика и телемеханика, No. 1, (1969), 57-77.

[7] Браверман, Э. М., и Розоноер, Л. И.: Сходимость случайных процессов в теории обучения машин. 2, Автоматика и телемеханика, No. 3, (1969), 87-103.

[8] CHUNG, K. L.: *On a stochastic approximation method*, Ann. Math. Stat., Vol. 25, (1954), 463-483.

[9] DUPAČ, V.: *A dynamic stochastic approximation method*, Ann. Math. Stat., Vol. 36, (1965), 1965-1702.

[10] KITAGAWA, T.: *Successive process of statistical controls I*, Mem. Fac. Sci. Kyushu Univ., Ser. A., Vol. 7, (1952), 13-28.

[11] KITAGAWA, T.: *Successive process of statistical controls II*, Mem. Fac. Sci. Kyushu Univ., Ser. A., Vol. 13, (1959), 1-16.

[12] TANAKA, K.: *On the pattern classification problems by learning I*, Bull. Math. Stat., Vol. 10, (1970), 31-49.

[13] TSYPKIN, Ya. Z.: *Establishing characteristics of a function transformer from randomly observed points*, Autom. and Remote Control., Vol. 26, No. 11, (1965), 1878-1881.

[14] WASAN, M. T.: "*Stochastic Approximation*", Cambridge Univ, Press, 1969.