

SOME APPLICATIONS OF STOCHASTIC APPROXIMATION METHOD

Fukamichi, Haruo
Research Institute of Fundamental Information Science, Kyushu University

<https://doi.org/10.5109/13046>

出版情報：統計数理研究. 14 (1/2), pp.75-81, 1970-03. Research Association of Statistical
Sciences
バージョン：
権利関係：



SOME APPLICATIONS OF STOCHASTIC APPROXIMATION METHOD

By

Haruo FUKAMICHI*

(Received February 7, 1970)

§ 1. Introduction.

Stochastic approximation method first introduced by Robbins-Monro [10] has been proved to be very useful for a learning system in the sense that its algorithm is very simple and that, at any given time in the learning process, the past samples are not required to retain in memory as shown by Albert and Gardner [1], Blum [3] etc.

In this paper we are concerned with its applications. After giving some preliminaries and notations in Section 2, we shall treat in Section 3 the problem of finding threshold elements, which is fundamental in pattern classification. In Section 4 we shall consider the problem of parameter identification in linear system with an additive noise.

§ 2. Preliminaries and notations.

Before going into the main argument, we state two Lemmas by which the sequence of vectors determined by our algorithm is proved to converge almost surely (abbreviated by a. s.), that is to say, with the probability one, to a solution.

Lemma 2.1 is a straightforward modification of the result given already by Blum [3], and Lemma 2.2 is an immediate consequence of the well known convergence theorem on random variables (see e. g. Chapter 4 of Chung [5]).

LEMMA 2.1. *Let $\{X_n\}$ be a sequence of real-valued random variables satisfying*

(i) $X_n \geq 0$, a. s., for $n = 1, 2, \dots$,

(ii) $\sup_n E\{X_n\} < \infty$,

(iii) $E\{X_{n+1} | X_1, X_2, \dots, X_n\} \leq (1 + \alpha_n K_1)X_n + \beta_n K_2$, for $n = 1, 2, \dots$,

where K_1 and K_2 are non-negative constants, and $\{\alpha_n\}$ and $\{\beta_n\}$ are two assigned sequences of positive numbers satisfying

$$(2.1) \quad \sum_{n=1}^{\infty} \alpha_n < \infty, \quad \sum_{n=1}^{\infty} \beta_n < \infty.$$

Then X_n converges a. s. to a random variable.

LEMMA 2.2. *Let $\{X_n\}$ be a sequence of real-valued random variables satisfying*

* Research Institute of Fundamental Information Science, Kyushu University, Fukuoka.

(i) $X_n \geq 0$, a. s., for $n=1, 2, \dots$,

(ii) $\sum_{n=1}^{\infty} \alpha_n E\{X_n\} < \infty$,

where $\{\alpha_n\}$ is an assigned sequence of positive numbers with

$$(2.2) \quad \sum_{n=1}^{\infty} \alpha_n = \infty.$$

Then there exists a strictly monotone increasing sequence of positive integers $\{n_p\}$ such that

$$(2.3) \quad \lim_{p \rightarrow \infty} X_{n_p} = 0, \quad \text{a. s.}$$

In what follows we adopt the following notations. Let \mathbf{E}_r be a real r -dimensional vector space. For any two vectors \mathbf{a} and \mathbf{b} in \mathbf{E}_r , we denote their inner product by $\mathbf{a} \cdot \mathbf{b}$, their norms by $\|\mathbf{a}\|$ and $\|\mathbf{b}\|$, and their transposes by \mathbf{a}' and \mathbf{b}' .

§ 3. Convergence algorithm to find threshold elements.

Let \mathbf{X} be an \mathbf{E}_r -valued random variable with the probability distribution $P_{\mathbf{X}}(\cdot)$ characterized by the following assumptions:

ASSUMPTION 3.1. For a certain pre-assigned positive number θ the set $\mathbf{S}_{\theta} = \{\mathbf{a} | \mathbf{a} \cdot \mathbf{X} \geq \theta \text{ a. s., } \mathbf{a} \in \mathbf{E}_r\}$ is nonempty.

ASSUMPTION 3.2. There exists a positive constant α satisfying $\|\mathbf{X}\| \leq \alpha$, a. s.

We consider the problem how to find such an \mathbf{a} ($\in \mathbf{S}_{\theta}$), which is called a vector of threshold elements in the sense of Martinez [8]. The problems of this type seem to be fundamental in pattern classification as pointed out by Horibe [7] and an analogous problem is treated in Ho and Kashyap [6], where \mathbf{X} is assumed to be a discrete random variable. Our approach to this problem is to give a sequence of vectors $\{\mathbf{a}_n\}$ determined by the algorithm stated below, and to show that $\{\mathbf{a}_n\}$ tends a. s. to the set \mathbf{S}_{θ} .

Let $\{\mathbf{X}_n\}$ be a sequence of independently observed \mathbf{E}_r -valued random vectors with the same probability distribution $P_{\mathbf{X}}(\cdot)$. We define $\{\mathbf{a}_n\}$ recursively as follows:

$$(3.3) \quad \begin{aligned} \mathbf{a}_1 &= \text{arbitrary given } (r\text{-dim.}) \text{ constant vector} \\ \mathbf{a}_{n+1} &= \mathbf{a}_n + \gamma_n(1 - \text{sgn}(\mathbf{a}_n \cdot \mathbf{X}_n - \theta))\mathbf{X}_n, \quad \text{for } n=1, 2, \dots, \end{aligned}$$

where $\{\gamma_n\}$ is an assigned sequence of positive numbers satisfying

$$(3.4) \quad \sum_{n=1}^{\infty} \gamma_n = \infty, \quad \sum_{n=1}^{\infty} \gamma_n^2 < \infty.$$

and as usually

$$(3.5) \quad \text{sgn}(\varphi) = 1, \text{ for } \varphi \geq 0 \quad \text{and} \quad \text{sgn}(\varphi) = 0, \text{ for } \varphi < 0.$$

The following Theorem shows that $\{\mathbf{a}_n\}$ tends, a. s., to a point in \mathbf{S}_{θ} .

THEOREM 3.1. Under the assumptions (3.1) and (3.2), we have,

$$(3.6) \quad \lim_{n \rightarrow \infty} \inf_{\varphi \in \vartheta} \|\mathbf{a}_n - \mathbf{a}\|^2 = 0, \quad \text{a. s.},$$

where $\{\mathbf{a}_n\}$ is defined by (3.3).

PROOF. Let us put,

$$Y_n = \inf_{\mathbf{a} \in \theta} \|\mathbf{a}_n - \mathbf{a}\|^2.$$

In view of (3.3) we have, for all $\mathbf{a} \in S_\theta$,

$$(3.7) \quad \begin{aligned} \|\mathbf{a}_{n+1} - \mathbf{a}\|^2 &= \|\mathbf{a}_n - \mathbf{a}\|^2 + 2\gamma_n(\mathbf{a}_n - \mathbf{a}) \cdot \mathbf{X}_n(1 - \operatorname{sgn}(\mathbf{a}_n \cdot \mathbf{X}_n - \theta)) \\ &\quad + \gamma_n^2 \|\mathbf{X}_n\|^2(1 - \operatorname{sgn}(\mathbf{a}_n \cdot \mathbf{X}_n - \theta))^2. \end{aligned}$$

Also, from (3.5) and the assumption on $\{\mathbf{X}_n\}$, it can be easily shown that

$$(3.8) \quad (\mathbf{a}_n - \mathbf{a}) \cdot \mathbf{X}_n(1 - \operatorname{sgn}(\mathbf{a}_n \cdot \mathbf{X}_n - \theta)) \leq 0, \quad \text{a.s., for all } \mathbf{a} \in S_\theta$$

and that

$$(3.9) \quad \|\mathbf{X}_n\|^2(1 - \operatorname{sgn}(\mathbf{a}_n \cdot \mathbf{X}_n - \theta))^2 \leq \alpha^2, \quad \text{a.s..}$$

Inserting (3.8) and (3.9) into (3.7) lead us to

$$(3.10) \quad \|\mathbf{a}_{n+1} - \mathbf{a}\|^2 \leq \|\mathbf{a}_n - \mathbf{a}\|^2 + \gamma_n^2 \alpha^2, \quad \text{a.s., for all } \mathbf{a} \in S_\theta,$$

by which it can be easily shown that

$$(3.11) \quad E\{Y_{n+1} | Y_1, Y_2, \dots, Y_n\} \leq Y_n + \gamma_n^2 \alpha^2, \quad \text{for } n = 1, 2, \dots$$

Also, from (3.10) and the assumption (3.4) on $\{\gamma_n\}$, it follows that, for an arbitrary given point $\bar{\mathbf{a}} \in S_\theta$

$$(3.12) \quad \|\mathbf{a}_{n+1} - \bar{\mathbf{a}}\|^2 \leq \|\mathbf{a}_1 - \bar{\mathbf{a}}\|^2 + \sum_{m=1}^{\infty} \gamma_m^2 \alpha^2, \quad \text{a.s., for } n = 1, 2, \dots,$$

where the right-hand side of (3.12) is a finite positive constant independent of n . From (3.12) we have

$$(3.13) \quad \sup_n E\{Y_n\} \leq \sup_n E\{\|\mathbf{a}_n - \bar{\mathbf{a}}\|^2\} < \infty.$$

In view of (3.11), (3.13) and the fact that $Y_n \geq 0$, a.s., the application of Lemma 2.1 to $\{Y_n\}$ yields the result that there is a random variable Y such that

$$(3.14) \quad \lim Y_n = Y, \quad \text{a.s..}$$

Now we shall show $Y = 0$ a.s.. Let \mathbf{a} be a point in S_θ . Then, from (3.7) and (3.9), we have

$$(3.15) \quad \|\mathbf{a}_{n+1} - \mathbf{a}\|^2 \leq \|\mathbf{a}_n - \mathbf{a}\|^2 + \gamma_n(\mathbf{a}_n - \mathbf{a}) \cdot \mathbf{X}_n(1 - \operatorname{sgn}(\mathbf{a}_n \cdot \mathbf{X}_n - \theta)) + \gamma_n^2 \alpha^2, \quad \text{a.s..}$$

and hence from the assumption on $\{\mathbf{X}_n\}$, we have

$$(3.16) \quad E\{\|\mathbf{a}_{n+1} - \mathbf{a}\|^2\} \leq E\{\|\mathbf{a}_n - \mathbf{a}\|^2\} + 2\gamma_n E\left\{\int_{\mathbf{a}_n \cdot \mathbf{x} \leq \theta} (\mathbf{a}_n - \mathbf{a}) \cdot \mathbf{x} P_{\mathbf{x}}(d\mathbf{x})\right\} + \gamma_n^2 \alpha^2.$$

The application of (3.16) from $n=1$ to $n=m$, noting that $E\{\|\mathbf{a}_{m+1} - \mathbf{a}\|^2\} \geq 0$, gives

$$(3.17) \quad 2\gamma_n \sum_{n=1}^{\infty} E\left\{\int_{\mathbf{a}_n \cdot \mathbf{x} \leq \theta} (\mathbf{a} - \mathbf{a}_n) \cdot \mathbf{x} P_{\mathbf{x}}(d\mathbf{x})\right\} \leq E\{\|\mathbf{a}_1 - \mathbf{a}\|^2\} + \sum_{n=1}^{\infty} \gamma_n^2 \alpha^2.$$

Hence, from the assumption (3.4) on $\{\gamma_n\}$, we have, by $m \rightarrow \infty$ in (3.17),

$$(3.18) \quad \sum_{n=1}^{\infty} \gamma_n E\left\{\int_{\mathbf{a}_n \cdot \mathbf{x} \leq \theta} (\mathbf{a} - \mathbf{a}_n) \cdot \mathbf{x} P_{\mathbf{x}}(d\mathbf{x})\right\} < \infty.$$

Let us put, for the moment,

$$f(\mathbf{a}) = \int_{\mathbf{a} \cdot \mathbf{x} \leq \theta} (\theta - \mathbf{a} \cdot \mathbf{x}) P_{\mathbf{X}}(d\mathbf{x}).$$

Then, from the assumption (3.1), we have

$$(3.19) \quad 0 \leq f(\mathbf{a}_n) \leq \int_{\mathbf{a}_n \cdot \mathbf{x} \leq \theta} (\mathbf{a} - \mathbf{a}_n) \cdot \mathbf{x} P_{\mathbf{X}}(d\mathbf{x})$$

and hence

$$(3.20) \quad 0 \leq E\{f(\mathbf{a}_n)\} \leq E\left\{\int_{\mathbf{a}_n \cdot \mathbf{x} \leq \theta} (\mathbf{a} - \mathbf{a}_n) \cdot \mathbf{x} P_{\mathbf{X}}(d\mathbf{x})\right\},$$

which gives, in view of (3.18),

$$(3.21) \quad \sum_{n=1}^{\infty} \gamma_n E\{f(\mathbf{a}_n)\} < \infty.$$

From (3.19), (3.21) and the assumption on $\{\gamma_n\}$ it follows, by using Lemma 2.2, that there exists a strictly monotone increasing sequence of positive integers $\{n_p\}$ such that

$$(3.22) \quad \lim_{n \rightarrow \infty} f(\mathbf{a}_{n_p}) = 0, \quad \text{a. s.},$$

Also, from (3.12), it follows that there exists a positive constant K such that

$$(3.23) \quad \|\mathbf{a}_{n_p}\|^2 < K, \quad \text{a. s. for } p = 1, 2, \dots.$$

Noting that $f(\mathbf{a}) = 0$ if and only if $\mathbf{a} \in S_{\theta}$, and that $f(\mathbf{a})$ is continuous on \mathbf{E}_r under the assumption (3.2), it follows from (3.22) and (3.23) that

$$(3.24) \quad Y_{n_p} = \inf_{\mathbf{a} \in S_{\theta}} \|\mathbf{a}_{n_p} - \mathbf{a}\|^2 \rightarrow 0, \quad \text{a. s., as } p \rightarrow \infty.$$

Consequently, from (3.14) and (3.24), it follows that $Y = 0$, a. s., which is to be proved.

§ 4. Parameter identification algorithm.

In this section we consider the problem of identifying parameters in the following linear system:

$$(4.1) \quad Y = \mathbf{a} \cdot \mathbf{X} + Z,$$

where $\mathbf{a} = (a_1, a_2, \dots, a_r)^t$ is a vector of unknown parameters to be identified, $\mathbf{X} = (X_1, \dots, X_r)^t$ is an \mathbf{E}_r -valued random variable with two conditions:

$$(4.2) \quad E\{X_i\} < \infty, \quad \text{for } i = 1, 2, \dots, r,$$

$$(4.3) \quad E\{\mathbf{X}\mathbf{X}^t\} \text{ is nonsingular,}$$

and Z is a noise such that

$$(4.4) \quad Z \text{ is stochastically independent of } \mathbf{X}$$

and

$$(4.5) \quad E\{Z\} = 0, \quad \text{Var}\{Z\} < \infty.$$

The case when the input \mathbf{X} is not a random variable is treated in Albert and

Gardner [1]. Noda [9] considers the problem analogous to our case under certain conditions on the probability distribution function of X . In our paper we give a sequence of vectors $\{\mathbf{a}_n\}$ which converges to an \mathbf{a} in (4.1) under fairly weak conditions (4.2)–(4.5). We consider a sequence $\{\mathbf{a}_n\}$ which is determined by the following algorithm.

Let $\{X_n\}$ be a sequence of observable independent random inputs in (4.1) and $\{Y_n\}$ is a sequence of observable outputs corresponding to $\{X_n\}$, that is to say, $\{Y_n\}$ is determined by the equations $Y_n = \mathbf{a} \cdot X_n + Z_n$ ($n = 1, 2, \dots$), where $\{Z_n\}$ is a sequence of unobservable independent random variables (noises) with (4.4) and (4.5). We define $\{\mathbf{a}_n\}$ recursively as follows

$$(4.6) \quad \begin{aligned} \mathbf{a}_1 &= \text{arbitrary given } (r\text{-dim.}) \text{ constant vector} \\ \mathbf{a}_{n+1} &= \mathbf{a}_n - \gamma_n X_n (\mathbf{a}_n \cdot X_n - Y_n), \quad \text{for } n = 1, 2, \dots, \end{aligned}$$

where $\{\gamma_n\}$ is a pre-assigned sequence of positive numbers satisfying (3.4) in the previous section.

Now we give the Theorem which shows that \mathbf{a}_n converges to \mathbf{a} a. s. as $n \rightarrow \infty$.

THEOREM 4.1. *Under the conditions (4.2)–(4.5), we have*

$$(4.7) \quad \lim_{n \rightarrow \infty} \mathbf{a}_n = \mathbf{a}, \quad \text{a. s.,}$$

where $\{\mathbf{a}_n\}$ is defined by (4.6).

PROOF. In view of (4.6) we have

$$(4.8) \quad \begin{aligned} \|\mathbf{a}_{n+1} - \mathbf{a}\|^2 &= \|\mathbf{a}_n - \mathbf{a}\|^2 - 2\gamma_n (\mathbf{a}_n - \mathbf{a}) \cdot X_n (\mathbf{a}_n \cdot X_n - Y_n) \\ &\quad + \gamma_n^2 \|X_n\|^2 (\mathbf{a}_n \cdot X_n - Y_n)^2. \end{aligned}$$

Inserting $Y_n = \mathbf{a} \cdot X_n + Z_n$ into (4.8) yields

$$(4.9) \quad \begin{aligned} \|\mathbf{a}_{n+1} - \mathbf{a}\|^2 &= \|\mathbf{a}_n - \mathbf{a}\|^2 - 2\gamma_n (\mathbf{a}_n - \mathbf{a}) \cdot X_n X_n^t (\mathbf{a}_n - \mathbf{a}) \\ &\quad + 2\gamma_n (\mathbf{a}_n - \mathbf{a}) \cdot X_n Z_n + \gamma_n^2 \|X_n\|^2 ((\mathbf{a}_n - \mathbf{a}) \cdot X_n - Z_n)^2. \end{aligned}$$

Hence, from the assumptions on $\{X_n\}$ and $\{Z_n\}$, it follows that

$$(4.10) \quad \begin{aligned} E\{\|\mathbf{a}_{n+1} - \mathbf{a}\|^2 | \mathbf{a}_n\} &= \|\mathbf{a}_n - \mathbf{a}\|^2 - 2\gamma_n (\mathbf{a}_n - \mathbf{a}) \cdot E\{X_n X_n^t\} (\mathbf{a}_n - \mathbf{a}) \\ &\quad + \gamma_n^2 E\{\|X_n\|^2 ((\mathbf{a}_n - \mathbf{a}) \cdot X_n - Z_n)^2 | \mathbf{a}_n\} \\ &= \|\mathbf{a}_n - \mathbf{a}\|^2 - 2\gamma_n (\mathbf{a}_n - \mathbf{a}) \cdot E\{XX^t\} (\mathbf{a}_n - \mathbf{a}) \\ &\quad + \gamma_n^2 (\mathbf{a}_n - \mathbf{a}) \cdot E\{\|X\|^2 XX^t\} (\mathbf{a}_n - \mathbf{a}) \\ &\quad + \gamma_n^2 E\{\|X\|^2\} E\{Z^2\}. \end{aligned}$$

Let $\mathbf{A} = E\{XX^t\}$ and $\mathbf{B} = E\{\|X\|^2 XX^t\}$. Since \mathbf{A} is positive definite, from the condition (4.3), and \mathbf{B} exists, from the condition (4.2), and is always positive semi-definite, it follows from the well known result in matrix analysis (see e. g. Ch. 7 of Bellman [2]) that

$$(4.11) \quad (\mathbf{a}_n - \mathbf{a}) \cdot \mathbf{A}(\mathbf{a}_n - \mathbf{a}) \geq \lambda_0(\mathbf{A}) > 0$$

and that

$$(4.12) \quad 0 \leq (\mathbf{a}_n - \mathbf{a}) \cdot \mathbf{B}(\mathbf{a}_n - \mathbf{a}) \leq \lambda_1(\mathbf{B}),$$

where $\lambda_0(\mathbf{A})$ ($\lambda_1(\mathbf{B})$) is the minimum (maximum) eigenvalue of $\mathbf{A}(\mathbf{B})$. Inserting these in equality (4.10), we get

$$(4.13) \quad E\{\|\mathbf{a}_{n+1} - \mathbf{a}\|^2 | \mathbf{a}_n\} \leq (1 - 2\gamma_n \lambda_0(\mathbf{A}) + \gamma_n^2 \lambda_1(\mathbf{B}))\|\mathbf{a}_n - \mathbf{a}\|^2 + \gamma_n^2 K,$$

where $K = E\{\|\mathbf{X}\|^2\}E\{Z^2\}$ (note that K is a finite positive constant). Let $U_n = \|\mathbf{a}_n - \mathbf{a}\|^2$. Then (4.13) gives

$$(4.14) \quad \begin{aligned} E\{U_{n+1} | U_1, U_2, \dots, U_n\} &\leq (1 - 2\gamma_n \lambda_0(\mathbf{A}) + \gamma_n^2 \lambda_1(\mathbf{B}))U_n + \gamma_n^2 K \\ &\leq (1 + \gamma_n^2 \lambda_1(\mathbf{B}))U_n + \gamma_n^2 K \end{aligned}$$

and

$$(4.15) \quad E\{U_{n+1}\} \leq (1 - 2\gamma_n \lambda_0(\mathbf{A}) + \gamma_n^2 \lambda_1(\mathbf{B}))E\{U_n\} + \gamma_n^2 K.$$

Since, in view of the assumption on $\{\gamma_n\}$, (4.11) and (4.12), there exists a positive integer n_0 such that

$$(4.16) \quad \gamma_n^2 \lambda_1(\mathbf{B}) - 2\gamma_n \lambda_0(\mathbf{A}) \leq -\gamma_n \lambda_0(\mathbf{A}), \quad \text{for all } n \geq n_0,$$

we have, from (4.15),

$$(4.17) \quad \begin{aligned} E\{U_{n+1}\} &\leq (1 - \gamma_n \lambda_0(\mathbf{A}))E\{U_n\} + \gamma_n^2 K, \\ &\leq E\{U_n\} + \gamma_n^2 K, \quad \text{for all } n \geq n_0. \end{aligned}$$

The assumption on $\{\gamma_n\}$ and (4.17) gives

$$(4.18) \quad \begin{aligned} E\{U_{n+1}\} &\leq E\{U_{n_0}\} + \sum_{m=n_0}^n \gamma_m^2 K \\ &\leq E\{U_{n_0}\} + \sum_{m=1}^{\infty} \gamma_m^2 K, \quad \text{for all } n \geq n_0. \end{aligned}$$

Noting that $E\{U_n\} < \infty$ (for $n = 1, 2, \dots, n_0$) and that $\sum_{n=1}^{\infty} \gamma_n^2 K < \infty$, it follows from (4.18) that

$$(4.19) \quad \sup_n E\{U_n\} < \infty$$

Hence, applying Lemma 2.1 to $\{U_n\}$, we have, from (4.14) and (4.19), for some random variable Y ,

$$(4.20) \quad \lim_{n \rightarrow \infty} U_n = Y, \quad \text{a. s. .}$$

Also, by an argument analogous to that used in deriving (3.18), we have, from (4.15) and (4.19),

$$(4.21) \quad \sum_{n=1}^{\infty} \gamma_n E\{U_n\} < \infty.$$

Hence, from Lemma 2.2, it follows that there exists a strictly monotone increasing sequence of positive integers $\{n_p\}$ satisfying

$$(4.22) \quad \lim_{p \rightarrow \infty} U_{n_p} = 0, \quad \text{a. s. .}$$

Finally, the combination of (4.20) and (4.22) yields

$$(4.23) \quad \lim_{n \rightarrow \infty} U_n = 0, \quad \text{a. s.},$$

which completes the proof of Theorem 4.1.

REMARK. If $\gamma_n = 1/n$ in (4.6), $K > 0$, and $\lambda_0(A) > 1$, then the application of Chung's result [4] to (4.17) gives an evaluation of the rate of convergence to the effect that

$$(4.24) \quad E\{\|a_n - a\|^2\} \leq \frac{K}{n(\lambda_0(A) - 1)} + o\left(\frac{1}{n}\right),$$

where $\alpha_n = o(\beta_n)$ means $\alpha_n/\beta_n \rightarrow 0$ as $n \rightarrow \infty$.

In conclusion, the author wishes to express his hearty thanks to Professors Tosio Kitagawa, Seigo Kanō and Akio Kudō for their continuing guidances and encouragements given to him.

References

- [1] ALBERT, A. E. and GARDNER, L. A., *Stochastic approximation and nonlinear regression*, M. I. T. press. Cambridge, Massachusetts (1967).
- [2] BELLMAN, R., *Introduction to matrix analysis*, McGraw-Hill Book Company Inc. New York (1960).
- [3] BLUM, J. R., "Multidimensional Stochastic approximation methods", *Ann. Math. Stat.*, **25** (1954), 737-744.
- [4] CHUNG, K. L., "On stochastic approximation methods", *Ann. Math. Stat.*, **25** (1954), 463-483.
- [5] CHUNG, K. L., *A course in probability theory*, Harcourt, Brace & World, Inc. New York (1968).
- [6] HO, Y. C. and KASHYAP, R. L., "A class of iterative procedures for linear inequalities", *J. SIAM. Control*, **4** (1966), 112-115.
- [7] HORIBE, Y., "On an adaptive process for learning finite patterns", *Kōdai Math. Sem. Rep.*, **19** (1967), 43-52.
- [8] MARTINEZ, H. M., "A convergence theorem for linear threshold elements", *Bull. Math. Biophysics*, **27** (1965), 153-159.
- [9] NODA, A., "Effects of noise and parameter-variation on the learning identification method", *J. Soc. Indust. Cont. Engineers*, **8** (1969), 303-312. (In Japanese)
- [10] ROBBINS, H. and MONRO, S., "A stochastic approximation method", *Ann. Math. Stat.*, **22** (1951), 400-407.