

# STRONG CONVERGENCE OF AN UNBOUNDED STOCHASTIC APPROXIMATION ALGORITHM UNDER GENERAL CONDITIONS

Watanabe, Masafumi  
Department of Applied Mathematics, Fukuoka University

<https://doi.org/10.5109/13426>

---

出版情報 : Bulletin of informatics and cybernetics. 25 (1/2), pp.109-123, 1992-03. Research  
Association of Statistical Sciences  
バージョン :  
権利関係 :



# **STRONG CONVERGENCE OF AN UNBOUNDED STOCHASTIC APPROXIMATION ALGORITHM UNDER GENERAL CONDITIONS**

By

**Masafumi WATANABE\***

## **Abstract**

Under general conditions on the observation processes the almost sure convergence properties of an unbounded and nonlinear Robbins–Monro type stochastic approximation algorithm are investigated.

## **1. Introduction**

In the classical studies of stochastic approximation, essentially it was assumed that the observations are independent. But many practical applications involve the use of correlated data. Consequently classical stochastic approximation convergence results are often not applicable. Recently, several efforts have been made to get rid of it, for example it can be replaced by mixing conditions or ergodic conditions. Borodin [1], Fritz [3], Györfi [3], Kushner and Clark [4], Ljung [5, 6] investigated the case when the observations may not be independent. For linear regression Fritz [2] and Györfi [3] formulate a completely deterministic problem, by which the strong consistency can be deduced if for the observations the strong law of large numbers is applied, therefore the conditions of strong consistency of linear stochastic approximation are as general as the conditions of the strong law of large numbers. Thus the effects of the randomness and the iteration were separated. By the similar motivation and for nonlinear regression Ljung [6] and Kushner and Clark [4] have obtained interesting results under the assumption that the algorithms are bounded.

Generally the bounded condition is not easily established for the family of stochastic approximation algorithms. In this paper we shall investigate the strong convergence of an unbounded and nonlinear Robbins–Monro type stochastic approximation algorithm under general conditions on the observations.

## **2. The Robins–Monro Stochastic Approximation Algorithm**

Throughout the paper  $H$  denotes a real separable Hilbert space with an inner product  $\langle \cdot, \cdot \rangle$  and a norm  $\|\cdot\|$ , and  $O$  denotes the null element of  $H$ . Let  $(\Omega, \mathcal{A}, P)$  be a probability space.

---

\* Department of Applied Mathematics, Fukuoka University Fukuoka 814–01, Japan

Let  $M(x)$  be a Borel measurable transformation from  $H$  into itself. For each  $n \geq 1$ , let a function  $Y_n(x, \omega)$  which takes value in  $H$  and is defined for  $(x, \omega) \in H \times \Omega$  be measurable with respect to  $\mathcal{B} \times \mathcal{A}$ , where  $\mathcal{B}$  is the Borel field of  $H$ . Then we consider the following Robbins-Monro stochastic approximation algorithm for finding the solution of the equation

$$M(x) = O. \quad (2.1)$$

*Algorithm:*

$$X_{n+1} = X_n - a_n Y_n, \quad n = 1, 2, \dots \quad (2.2)$$

where  $X_1$  is a constant element in  $H$ ,  $Y_n = Y_n(X_n(\omega), \omega)$ ,  $\omega \in \Omega$  and  $\{a_n\}$  is a sequence of positive numbers. Throughout the paper let us put

$$Z_n(x, \omega) = Y_n(x, \omega) - M(x), \quad n = 1, 2, \dots, (x, \omega) \in H \times \Omega. \quad (2.3)$$

Then (2.2) can be rewritten the following

$$X_{n+1} = X_n - a_n \{M(X_n) + Z_n\}, \quad n = 1, 2, \dots, \quad (2.4)$$

where

$$Z_n = Z_n(X_n(\omega), \omega), \quad \omega \in \Omega.$$

Fritz [2] and Györfi [3] have investigated the above problem under the conditions that  $M(x) = Ax + b$  and  $Y_n(x, \omega) = A_n(\omega)x + b_n(\omega)$ , where  $A, A_n$  are bounded linear operators and  $b, b_n$  are elements in  $H$ . And the *a.s.* convergence of process (2.2) has been discussed under the assumptions that  $A_n$ 's and  $b_n$ 's arithmetic means converge to  $A$  and  $b$  respectively. Kushner and Clark [4] and Ljung [6] have investigated more general case and have given *a.s.* convergence results. In [4] and [6], they assumed that

$$\sup_n \|X_n\| < \infty \quad a.s. \quad (2.5)$$

Condition (2.5) appears to be a reasonable condition but is not easily established for the family of algorithms represented by (2.2). In this paper we note that condition (2.5) will not be assumed.

### 3. Preliminaries

In this section we shall give two lemmas which will be needed the proof of the theorem in the section 5 (Theorem B).

LEMMA 1. *Let  $\{a_n\}$  and  $\{v_n\}$  be sequences of real numbers such that*

$$(i) \quad a_n > 0, \quad n = 1, 2, \dots, \quad \lim_{n \rightarrow \infty} a_n = 0,$$

$$\sum_{n=1}^{\infty} a_n = \infty, \quad \sup_n |a_n^{-1} - a_{n+1}^{-1}| < \infty,$$

$$(ii) \quad v_n \geq 0, \quad n = 1, 2, \dots, \quad \sum_{n=1}^{\infty} v_n < \infty.$$

And let  $\{x_n\}$  be a sequence of nonnegative numbers and  $\{w_n\}$  be a sequence of real numbers which satisfy the following enequality

$$(iii) \quad x_{n+1} \leq (1 - a_n + v_n)x_n + w_n, \quad n = 1, 2, \dots$$

Suppose that let  $T$  be any positive number then there exists a positive number  $W_T$  such that

$$\sup_n \max_{n \leq k < m(n, T)} \left| \sum_{i=n}^k w_i \right| \leq W_T, \quad (3.1)$$

where

$$m(n, T) = \max \left\{ k \mid \sum_{i=n}^k a_i \leq T \right\}. \quad (3.2)$$

Then it follows that there exist two positive numbers  $T_0$  and  $W_0$  such that

$$\sup_n x_n \leq (W_{T_0} + 1)W_0. \quad (3.3)$$

PROOF. It follows from (i) that there exists a positive integer  $N$  such that

$$\begin{aligned} (1 - a_i + v_i) &= (1 - a_i)\{1 + (1 - a_i)^{-1}v_i\} \\ &\leq (1 - a_i)(1 + 2v_i) \quad \text{for all } i \geq N. \end{aligned} \quad (3.4)$$

Let us put

$$A(m, n) = \begin{cases} \prod_{i=m}^n (1 - a_i)(1 + 2v_i) & \text{if } n \leq m \\ 1 & \text{if } n = m - 1. \end{cases} \quad (3.5)$$

Condition (ii) yields that there exists a positive number  $\bar{v}$  such that

$$1 \leq \prod_{i=m}^{\infty} (1 + 2v_i) \leq \bar{v} \quad \text{for all } m \geq 1. \quad (3.6)$$

Hence, by (3.5) and (3.6) we have

$$A(m, n) \leq \bar{v} \exp \left( - \sum_{i=m}^n a_i \right) \quad \text{for } N \leq m \leq n. \quad (3.7)$$

Let us define  $T_0 > 0$  by

$$a_n \leq T_0/4 \quad n = 1, 2, \dots \quad (3.8)$$

$$\bar{v}/\exp(T_0/2) \leq \frac{1}{2}. \quad (3.9)$$

Then, by (3.2) we have

$$\sum_{i=n}^{m(n, T_0)-1} a_i \leq T_0 \quad , \quad n \geq N \quad (3.10)$$

and

$$\sum_{i=n}^{m(n, T_0)+1} a_i > T_0 \quad , \quad n \geq N. \quad (3.11)$$

Further, (3.8) and (3.11) imply

$$\sum_{i=n}^{m(n, T_0)-1} a_i > T_0 - a_m(n, T_0) - a_{m(n, T_0)-1} \geq \frac{T_0}{2} \quad , \quad n \geq N. \quad (3.12)$$

Hence by (3.10) and (3.12) we obtain

$$\frac{T_0}{2} \sum_{i=n}^{m(n, T_0)-1} a_i \leq T_0, \quad , \quad n \geq N. \quad (3.13)$$

And we also have from (3.9) and (3.13) that

$$A(n, m(n, T_0) - 1) \leq \frac{1}{2} \quad , \quad n \geq N. \quad (3.14)$$

If (3.1) holds then we can easily shown from (i) and (ii) that

$$\sup_{n \geq N} \max_{n \leq k < m(n, T_0)} \left| \sum_{i=n}^k w_i A(i+1, k) \right| \leq (1 + \bar{v} \bar{A}_N) W_{T_0}, \quad (3.15)$$

where

$$\bar{A}_N = \sup_{n \geq N} \left\{ \sum_{i=N}^n a_i \prod_{j=i+1}^n (1 - a_j) \right\} + 2 \sum_{n=1}^{\infty} v_n.$$

We note that (i) yields  $\bar{A}_N < \infty$ .

Next we shall define a subsequence  $\{x_{n_r}\} \subset \{x_n\}$  by the following way.

Let  $n_0 = N$  and  $n_r = m(n_{r-1}, T_0)$ ,  $r = 1, 2, \dots$

Then we have from (iii), (3.14) and (3.15) that

$$\begin{aligned} x_{n_1} &\leq A(n_0, n_1 - 1)x_{n_0} + \left| \sum_{i=n_0}^{n_1-1} w_i A(i+1, n_1 - 1) \right| \\ &\leq x_N + 2(1 + \bar{v} \bar{A}_N) W_{T_0}, \\ x_{n_2} &\leq \frac{1}{2} x_{n_1} + (1 + \bar{v} \bar{A}_N) W_{T_0} \\ &\leq x_N + 2(1 + \bar{v} \bar{A}_N) W_{T_0}. \end{aligned}$$

Hence we have

$$\sup_r x_{n_r} \leq x_N + 2(1 + \bar{v} \bar{A}_N) W_{T_0}, \quad (3.16)$$

Let  $n$  be any integer such that  $n \geq N$ . Then there exists positive integer  $r$  such that

$$n_r \leq n < n_{r+1}.$$

Hence we have from (iii), (3.7) and (3.16) that

$$\begin{aligned}
x_n &\leq A(n_r, n-1)x_{n_r} + \max_{n_r \leq k < n_{r+1}} \left| \sum_{i=n_r}^k w_i A(i+1, k) \right| \\
&\leq \bar{v}x_{n_r} + (1 + \bar{v}\bar{A}_N)W_{T_0} \\
&\leq 3\bar{v}(1 + \bar{v}\bar{A}_N)(x_N + W_{T_0}), \quad n_r \leq n < n_{r+1}.
\end{aligned}$$

Hence we have

$$\sup_n x_n \leq (1 + W_{T_0})W_0,$$

where  $W_0 = \max\{3\bar{v}(1 + \bar{v}\bar{A}_N), 3\bar{v}(1 + \bar{v}\bar{A}_N)\bar{x}_N\}$  and  $\bar{x}_N = \max\{x_1, x_2, \dots, x_N\}$ .

Q.E.D.

LEMMA 2. Suppose that the conditions of Lemma 1 hold but with the following condition replacing (3.1),  
for any positive number  $T$  it follows that

$$\lim_{n \rightarrow \infty} \max_{n \leq k < m(n, T)} \left| \sum_{i=n}^k w_i \right| = 0. \quad (3.17)$$

Then it holds that

$$\lim_{n \rightarrow \infty} x_n = 0. \quad (3.18)$$

PROOF. It follows from Lemma 1 that there exists positive numbers  $W_1$  such that

$$\sup_n x_n \leq W_1,$$

where  $W_1 = (1 + W_{T_0})W_0$ .

Let  $N$  be a positive integer which is given in the proof of Lemma 1. Let  $\varepsilon$  be any real number such that

$$0 < \varepsilon < \frac{1}{2}.$$

And let us define a positive number  $T_1$  by the following

$$2 \leq T_1 \quad (3.19)$$

$$\bar{v}/\exp(T_1/2) < \varepsilon, \quad (3.20)$$

where  $\bar{v}$  is also given in the proof of Lemma 1.

By the same arguments of (3.14) and from (3.20) we have

$$A(n, m(n, T_1) - 1) \leq \bar{v}\exp(-T_1/2) < \varepsilon < 1/2, \quad n \geq N. \quad (3.21)$$

And (3.17) implies that there exists a positive integer  $N_1 = N_1(\varepsilon, T_1) \geq N$  such that

$$\sup_{n \geq N_1} \max_{n \leq k < m(n, T_1)} \left| \sum_{i=n}^k w_i \right| < \varepsilon.$$

Hence, by the same arguments of the proof of (3.15) we have

$$\sup_{n \geq N_1} \max_{n \leq k < m(n, T_1)} \left| \sum_{i=n}^k w_i A(i+1, k) \right| \leq \varepsilon(1 + \bar{v} \bar{A}_{N_1}). \quad (3.22)$$

Next we shall define a subsequence  $\{x_{n_r}\} \subset \{x_n\}$  by the following way. Let us take  $n_0 = N_1$  and  $n_r = m(n_{r-1}, T_1)$ ,  $r = 1, 2, \dots$ . Since  $N \leq N_1$  and  $\sup_n x_n \leq W_1$  we have from (3.21) and (3.22)

$$\begin{aligned} x_{n_1} &\leq \varepsilon(W_1 + 1 + \bar{v} \bar{A}_{N_1}) < 2\varepsilon(W_1 + 1 + \bar{v} \bar{A}_{N_1}). \\ x_{n_2} &\leq \frac{1}{2}x_{n_1} + \varepsilon(W_1 + 1 + \bar{v} \bar{A}_{N_1}) < 2\varepsilon(W_1 + 1 + \bar{v} \bar{A}_{N_1}). \end{aligned}$$

Hence we obtain

$$\sup_r x_{n_r} \leq 2\varepsilon(1 + W_1 + \bar{v} \bar{A}_{N_1}). \quad (3.23)$$

For any  $n \geq n_1 = m(N_1, T_1)$  there exists a positive integer  $r$  such that  $n_r \leq n < n_{r+1}$ . Then it follows from (iii), (3.22) and (3.23) that

$$\begin{aligned} x_n &\leq A(n_r, n)x_{n_r} + \max_{n_r \leq k < n_{r+1}} \left| \sum_{i=n}^k w_i A(i+1, k) \right| \\ &\leq \bar{v}x_{n_r} + 2\varepsilon(1 + W_1 + \bar{v} \bar{A}_{N_1}) \\ &\leq 3\varepsilon\bar{v}(1 + W_1 + \bar{v} \bar{A}_{N_1}). \end{aligned}$$

Hence we obtain

$$\overline{\lim}_{n \rightarrow \infty} x_n \leq 3\varepsilon\bar{v}(1 + W_1 + \bar{v} \bar{A}_{N_1}).$$

Q.E.D.

Let  $\{a_n\}$  and  $\{w_n\}$  be sequences of real numbers which are given in Lemma 1. And we suppose that

$$a_n > a_{n+1}, \quad n = 1, 2, \dots$$

and

$$w_n = a_n y_n, \quad n = 1, 2, \dots$$

Then we shall consider the following four conditions.

- I.  $\sum_{n=1}^{\infty} a_n y_n$  converges,
- II.  $\lim_{n \rightarrow \infty} a_n \sum_{i=1}^n y_i = 0$ ,
- III.  $\lim_{n \rightarrow \infty} \sum_{i=1}^n a_i \prod_{j=i+1}^n (1 - a_j) y_i = 0$ ,

$$\text{IV. } \lim_{n \rightarrow \infty} \max_{n \leq k < m(n, T)} \left| \sum_{i=n}^k a_i y_i \right| = 0,$$

where  $m(n, T)$  is defined by (3.2).

Since  $a_n \downarrow 0$  as  $n \rightarrow \infty$  and by using Kronecker's Lemma we have

$$\text{I} \Rightarrow \text{II}.$$

Noting  $\sup_n |a_n^{-1} - a_{n+1}^{-1}| < \infty$  we have

$$\text{II} \Rightarrow \text{III}.$$

And we can also show that

$$\text{III} \Rightarrow \text{IV}.$$

Hence it follows that

$$\text{I} \Rightarrow \text{II} \Rightarrow \text{III} \Rightarrow \text{IV}.$$

Next we suppose that

$$a_n = n^{-1}, \quad n = 1, 2, \dots$$

Then we can show that

$$(n - 2)e^T \leq m(n, T) \leq ne^T, \quad n \geq 2$$

Hence we can prove that

$$\text{IV} \Rightarrow \text{II}.$$

And we can also prove that

$$\text{II} \Leftrightarrow \text{III}.$$

In this case it follows that

$$\text{II} \Leftrightarrow \text{III} \Leftrightarrow \text{IV}.$$

Let us suppose that

$$a_n = n^{-\alpha}, \quad y_n = n^{-\beta}, \quad n = 1, 2, \dots,$$

where  $0 < \alpha, \beta$  and  $\alpha + \beta = 1$ . In this case it is easily seen that I and II do not hold. And we have

$$m(n, T) \sim \{n^{1-\alpha} + (1 - \alpha)T\}^{(1-\alpha)^{-1}}.$$

Hence it follows that

$$\sum_{i=n}^{m(n, T)} a_i y_i = \sum_{i=n}^{m(n, T)} i^{-1} \sim (1 - \alpha)^{-1} \log \{1 + (1 - \alpha)Tn^{\alpha-1}\}.$$

Hence IV holds.



Thus condition IV can be considered more general than the other conditions I to III.

#### 4. Assumptions

Let  $\theta \in H$  be the solution of the equation (2.1) and let  $M(x)$ ,  $\{a_n\}$ ,  $\{X_n\}$ ,  $\{Y_n\}$  and  $\{Z_n\}$  be given in the section 1.

Moreover let  $\Omega_0$  be a subset of  $\Omega$  such that

$$\Omega_0 \in \mathcal{A} \quad \text{and} \quad P(\Omega_0) = 1.$$

We make the following assumptions.

A0: There exists a sequence of positive random variables  $\{\delta_n(\omega)\}$  such that

- (i)  $\delta_n(\omega) \geq \delta_{n+1}(\omega)$ ,  $n = 1, 2, \dots$  and  $\lim_{n \rightarrow \infty} \delta_n(\omega) = 0$ ,  $\omega \in \Omega_0$ ,
- (ii)  $\sup_n \delta_n(\omega) \delta_{n+1}^{-1}(\omega) < \infty$ ,  $\omega \in \Omega_0$ ,
- (iii)  $\sup_n \delta_n(\omega) \|X_n(\omega)\| < \infty$ ,  $\omega \in \Omega_0$ .

A1:  $\{a_n\}$  is a decreasing sequence of positive numbers.

- (i)  $\sum_{n=1}^{\infty} a_n^2 < \infty$ ,  $\sum_{n=1}^{\infty} a_n = \infty$ ,
- (ii)  $\sup_n |a_n^{-1} - a_{n+1}^{-1}| < \infty$ .

A2: There exists a positive constant  $C$  such that

$$\|M(x)\| \leq C(\|x\| + 1), \quad x \in H.$$

A3:  $\inf_{\varepsilon < \|x - \theta\| < \varepsilon^{-1}} \langle M(x), x - \theta \rangle > 0$  for any  $0 < \varepsilon < 1$ .

A4: There exists a sequence of nonnegative random variables  $\{\alpha_n(\omega)\}$  such that

- (i)  $\|Z_n(x, \omega)\| \leq \alpha_n(\omega)(\|x\| + 1)$ ,  $n = 1, 2, \dots$ ,  $(x, \omega) \in H \times \Omega_0$ ,
- (ii)  $\sum_{n=1}^{\infty} a_n^2 \alpha_n^2(\omega) < \infty$ ,  $\omega \in \Omega_0$ .

A5: Let  $\{\delta_n(\omega)\}$  be given in A0. And there exists a nonnegative random variables  $\beta_0(\omega)$  such that

- (i)  $\sup_n a_n \delta_n^{-3}(\omega) \left\| \sum_{i=1}^n Z_i(x, \omega) \right\| \leq \beta_0(\omega)(\|x\| + 1)$ ,  $(x, \omega) \in H \times \Omega_0$ ,
- (ii)  $\sup_n a_n \delta_n^{-3}(\omega) \left\| \sum_{i=1}^n \{Z_i(x, \omega) - Z_i(y, \omega)\} \right\| \leq \beta_0(\omega) \|x - y\|$ ,  
 $(x, \omega), (y, \omega) \in H \times \Omega_0$ ,

A6:

- (i)  $\sum_{n=1}^{\infty} a_n \delta_n(\omega) < \infty, \omega \in \Omega_0,$
- (ii)  $\sum_{n=1}^{\infty} a_n \delta_n(\omega) \alpha_n(\omega) < \infty, \omega \in \Omega_0.$

Assumptions A0 to A6 will be used in Theorem A which will be given in the next section. And the following assumptions B0 to B6 will be also used in Theorem B. Let  $m(n, T)$  be defined by (3.2).

B0: Assumption A0 holds.

B1: Assumption A1 holds.

B2: Assumption A2 holds.

B3: There exists a positive number  $\lambda$  such that

$$\langle M(x), x - \theta \rangle \geq \lambda \|x - \theta\|^2, x \in H.$$

B4: Assumption A4 holds.

B5: For any positive number  $T$  there exists a positive random variable  $\beta_T(\omega)$  such that

- (i)  $\sup_n \max_{n \leq k < m(n, T)} \left\| \sum_{i=n}^k a_i \delta_i^{-2}(\omega) Z_i(x, \omega) \right\| \leq \beta_T(\omega) (\|x\| + 1), (x, \omega) \in H \times \Omega_0,$
  - (ii)  $\sup_n \max_{n \leq k < m(n, T)} \left\| \sum_{i=n}^k a_i \delta_i^{-2}(\omega) \{Z_i(x, \omega) - Z_i(y, \omega)\} \right\| \leq \beta_T(\omega) \|x - y\|, (x, \omega), (y, \omega) \in H \times \Omega_0,$
- B6:  $\sup_n \sum_{i=n}^{m(n, T)} a_i \alpha_i(\omega) < \infty, (x, \omega) \in H \times \Omega_0.$

REMARK. Let us assume that

A'0: there exist a sequence of nonnegative random variables  $\{\gamma_n(\omega)\}$  and a sequence of positive random variables  $\{\delta_n(\omega)\}$  such that

- (i) A0(i), (ii) hold
- (ii)  $\langle \theta - x, Y_n(x, \omega) \rangle \leq \gamma_n(\omega) (\|x\| + 1),$   
 $n = 1, 2, \dots, (x, \omega) \in H \times \Omega_0,$
- (iii)  $\sum_{n=1}^{\infty} a_n \delta_n(\omega) \gamma_n(\omega) < \infty, \omega \in \Omega_0.$

Then it follows from Lemma 1([9]) that A0(iii) holds.

## 5. Results

In this section we give two theorems (Theorems A and B) which are the main results of this paper.

**THEOREM A.** *Assume A0 to A6. Then it holds that*

$$\lim_{n \rightarrow \infty} \|X_n - \theta\| = 0 \quad a.s. \quad (5.1)$$

**PROOF.** Without loss of generality we can assume  $\theta = \mathbf{O}$ . Here and latter, we take  $\theta = \mathbf{O}$ . and  $K$  denotes positive random variables which are not necessarily the same.

By virtue of the algorithm (2.2) we obtain

$$\|X_{n+1}\|^2 = \|X_n\|^2 - 2a_n \langle X_n, M(X_n) + Z_n \rangle + a_n^2 \|M(X_n) + Z_n\|^2. \quad (5.2)$$

Substituting A2 and A4(i) into (5.2) gives

$$\|X_{n+1}\|^2 \leq (1 + Ka_n^2 + Ka_n^2 \alpha_n^2) \|X_n\|^2 - 2a_n \langle X_n, M(X_n) + Z_n \rangle + Ka_n^2 (\alpha_n^2 + 1). \quad (5.3)$$

First we shall show that

$$\sum_{n=1}^{\infty} a_n \langle X_n, Z_n \rangle \quad \text{converges a.s.} \quad (5.4)$$

Define  $S_n(x) = a_n \sum_{i=1}^n Z_i(x)$ ,  $S_0(x) = \mathbf{O}$  and

$$W(m, n) = \sum_{i=m}^n a_i \langle X_i, Z_i \rangle, \quad n \leq m.$$

Since  $Z_i(x) = a_i^{-1} S_i(x) - a_{i-1}^{-1} S_{i-1}(x)$ ,  $W(m, n)$  can be rewritten the following formula

$$\begin{aligned} W(m, n) &= \sum_{i=m}^n \langle X_i, S_i(X_i) - a_i a_{i-1}^{-1} S_{i-1}(X_i) \rangle \\ &= W_1(m, n) + W_2(m, n) + W_3(m, n), \quad m \leq n, \end{aligned}$$

where

$$W_1(m, n) = \sum_{i=m}^n \langle X_i, S_i(X_i) - S_{i-1}(X_i) \rangle,$$

$$W_2(m, n) = \sum_{i=m}^n \langle X_i, (1 - a_i a_{i-1}^{-1}) S_{i-1}(X_{i-1}) \rangle$$

and

$$W_3(m, n) = \sum_{i=m}^n \langle X_i, a_i a_{i-1}^{-1} \{S_{i-1}(X_{i-1}) - S_{i-1}(X_i)\} \rangle.$$

Then we have

$$\begin{aligned}
|W_1(m, n)| &\leq \left| \sum_{i=m}^{n-1} \langle X_i - X_{i+1}, S_i(X_i) \rangle \right| + |\langle X_m, S_{m-1}(X_{m-1}) \rangle| + |\langle X_n, S_n(X_n) \rangle| \\
&\leq \sum_{i=m}^n \|S_i(X_i)\| \cdot \|X_i - X_{i+1}\| + \|X_m\| \cdot \|S_{m-1}(X_{m-1})\| + \|X_n\| \cdot \|S_n(X_n)\|.
\end{aligned}$$

Since  $X_{i+1} = X_i - a_i Y_i$ , it follows from A5 that

$$\begin{aligned}
|W_1(m, n)| &\leq \sum_{i=m}^n \beta_0 a_i \delta_i^3 (\|X_i\| + 1) \|Y_i\| + \beta_0 \delta_{m-1}^3 (\|X_{m-1}\| + 1) \|X_m\| \\
&\quad + \beta_0 \delta_n^3 (\|X_n\| + 1) \|X_n\|.
\end{aligned}$$

A2 and A4(i) imply

$$\|Y_i\| \leq K \alpha_i (\|X_i\| + 1), \quad i = 1, 2, \dots \quad (5.5)$$

Hence it follows from (5.5) and A0 that

$$|W_1(m, n)| \leq K \left( \sum_{i=m}^n a_i \delta_i \alpha_i + \delta_{m-1} + \delta_n \right).$$

Hence by A0(i) and A6(ii) we obtain

$$\lim_{m, n \rightarrow \infty} |W_1(m, n)| = 0 \quad \text{on } \Omega_0.$$

Since  $\{a_n\}$  is decreasing and  $1 - a_i a_{i-1}^{-1} = a_i (a_i^{-1} - a_{i-1}^{-1})$ , it follows from A0 and A1(ii) that

$$|W_2(m, n)| \leq K \sum_{i=m}^n a_{i-1} \delta_{i-1}.$$

Hence A6(i) implies

$$\lim_{m, n \rightarrow \infty} |W_2(m, n)| = 0 \quad \text{on } \Omega_0.$$

It follows from A5(ii) that

$$|W_3(m, n)| \leq \sum_{i=m}^n \beta_0 a_i \delta_{i-1}^3 \|X_i\| \|Y_{i-1}\|.$$

Hence by A0 and (5.5) we obtain

$$|W_3(m, n)| \leq K \sum_{i=m}^n a_{i-1} \delta_{i-1} \alpha_{i-1}.$$

Hence A6 (ii) implies

$$\lim_{m, n \rightarrow \infty} |W_3(m, n)| = 0 \quad \text{on } \Omega_0.$$

Thus (5.4) has been proved.

Next we shall show (5.1). Define  $U_n = 2a_n \langle X_n, M(X_n) \rangle$ ,  $V_n = Ka_n^2(\alpha_n^2 + 1)$  and  $W_n = 2a_n \langle X_n, Z_n \rangle$ . Then (5.3) yields

$$\|X_{n+1}\|^2 \leq (1 + KV_n) \|X_n\|^2 - U_n + KV_n - W_n.$$

Noting  $U_n \geq 0$ ,  $\sum_{n=1}^{\infty} V_n < \infty$  a.s.,  $\sum_{n=1}^{\infty} W_n$  converges a.s. and using the following inequality

$$\begin{aligned} 0 \leq \|X_{n+1}\|^2 &\leq \|X_1\|^2 \prod_{i=1}^n (1 + KV_i) - \sum_{i=1}^n U_i \prod_{j=i+1}^n (1 + KV_j) \\ &\quad + \sum_{i=1}^n (KV_i - W_i) \prod_{j=i+1}^n (1 + KV_j), \end{aligned}$$

we can obtain

$$\sup_n \|X_n\| < \infty \quad a.s., \quad (5.6)$$

and

$$\sum_{n=1}^{\infty} a_n \langle X_n, M(X_n) \rangle < \infty \quad a.s. \quad (5.7)$$

Hence it follows from A1(i) and A3 that there exists a subsequence  $\{X_{n_k}\} \subset \{X_n\}$  such that

$$\lim_{k \rightarrow \infty} \|X_{n_k}\| = 0 \quad a.s.$$

Moreover (5.2) yields that

$$|\|X_n\|^2 - \|X_m\|^2| \leq \sum_{i=m}^{n-1} a_i^2 \|Y_i\|^2 + \sum_{i=n}^{n-1} U_i + \left| \sum_{i=m}^{n-1} W_i \right|.$$

We note that (5.5), A1(i) and A4(ii) imply

$$\sum_{n=1}^{\infty} a_n^2 \|Y_n\|^2 < \infty \quad a.s. \quad (5.8)$$

Hence (5.4), (5.7) and (5.8) imply

$$\lim_{m, n \rightarrow \infty} |\|X_n\|^2 - \|X_m\|^2| = 0 \quad a.s. .$$

Hence (5.1) has been proved.

Q.E.D.

**THEOREM B.** Assume B0 to B6. Then (5.1) holds.

**PROOF.** Substituting B3 into (5.2) we obtain

$$\|X_{n+1}\|^2 \leq (1 - 2\lambda_n a_n + Ka_n^2 + Ka_n^2 \alpha_n^2) \|X_n\|^2 - 2a_n \langle X_n, Z_n \rangle + Ka_n^2(\alpha_n^2 + 1). \quad (5.9)$$

First we note that B1 and B4(ii) yield that

$$\limsup_{n \rightarrow \infty} \sum_{n \leq k < m(n, T)}^{m(n, T)} \alpha_i^2 (\alpha_i^2 + 1) = 0 \quad \text{on } \Omega_0 \quad (5.10)$$

for any  $T > 0$ . Next we shall show that, for any  $T > 0$  there exists a positive random variables  $W_0(T, \omega)$  such that

$$\sup_n \max_{n \leq k < m(n, T)} \left| \sum_{i=n}^k a_i \langle X_i(\omega), Z_i(X_i(\omega), \omega) \rangle \right| \leq W_0(T, \omega), \quad \omega \in \Omega_0. \quad (5.11)$$

Define  $S_n^k(x) = \sum_{i=n}^k a_i \delta_i^{-2} Z_i(x)$ ,  $k \geq n$  and  $W(n, k) = \sum_{i=n+1}^k a_i \langle X_i, Z_i \rangle$ ,  $k \geq n + 1$ .

Since  $Z_i(x, \omega) \equiv Z_i(x) = a_i^{-1} \delta_i^2 \{S_n^i(x) - S_n^{i-1}(x)\}$ ,  $i \geq n + 1$ ,  $W(n, k)$  can be rewritten the following formula

$$W(n, k) = W_1(n, k) + W_2(n, k) + W_3(n, k), \quad k \geq n + 1,$$

where

$$W_1(n, k) = \sum_{i=n+1}^k \langle X_i, \delta_i^2 S_n^i(X_i) - \delta_{i-1}^2 S_n^{i-1}(X_{i-1}) \rangle$$

$$W_2(n, k) = \sum_{i=n+1}^k \langle X_i, \delta_{i-1}^2 \{S_n^{i-1}(X_{i-1}) - S_n^{i-1}(X_i)\} \rangle$$

and

$$W_3(n, k) = \sum_{i=n+1}^k \langle X_i, (\delta_{i-1}^2 - \delta_i^2) S_n^{i-1}(X_{i-1}) \rangle.$$

Then we have

$$\begin{aligned} |W_1(n, k)| &\leq \left| \sum_{i=n}^{k-1} \langle X_i - X_{i+1}, \delta_i^2 S_n^i(X_i) \rangle \right| + |\langle X_{n+1}, \delta_n^2 S_n^n(X_n) \rangle| \\ &\quad + |\langle X_k, \delta_k^2 S_n^k(X_k) \rangle|, \quad k \geq n + 1. \end{aligned}$$

Since  $X_{i+1} = X_i - a_i Y_i$ , it follows from B0, B5(i), (5.5) and B6 that there exists a positive random variable  $\overline{W}_1(T, \omega)$  such that

$$\begin{aligned} \max_{n \leq k < m(n, T)} |W_1(n, k)| &\leq K \max_{n \leq k < m(n, T)} \sum_{i=n}^k a_i \delta_i \alpha_i \|S_n^i(X_i)\| \\ &\leq \overline{W}_1(T) \quad \text{on } \Omega_0. \end{aligned}$$

Since  $\sup_n \delta_{n-1} \delta_n^{-1} < \infty$ , it follows from (5.5), B0(iii) and B5(ii) that there exists a positive random variable  $\overline{W}_2(T, \omega)$  such that

$$\max_{n \leq k < m(n, T)} |W_2(n, k)| \leq \overline{W}_2(T) \quad \text{on } \Omega_0.$$

And we also obtain from B0 and B5(i)

$$\max_{n \leq k < m(n, T)} |W_3(n, k)| \leq \overline{W}_3(T) \quad \text{on } \Omega_0.$$

Where  $\overline{W}_3(T) = \overline{W}_3(T, \omega)$  is a positive random variable.

Hence (5.11) has been proved. Noting B1 and B4(ii) and applying Lemma 1 to (5.9) it follows from (5.10) and (5.11) that

$$\sup_n \|X_n\| < \infty \quad \text{on } \Omega_0. \quad (5.12)$$

It is easily seen from B5(i), (ii) that there exists sequence of positive random variables  $\{\beta_n(\omega)\}$  such that

$$\lim_{n \rightarrow \infty} \beta_n(\omega) = 0 \quad , \quad \omega \in \Omega_0, \quad (5.13)$$

$$\max_{n \leq k < m(n, T)} \left\| \sum_{i=n}^k a_i Z_i(x, \omega) \right\| \leq \beta_n(\omega)(\|x\| + 1), \quad (x, \omega) \in H \times \Omega_0. \quad (5.14)$$

$$\max_{n \leq k < m(n, T)} \left\| \sum_{i=n}^k a_i \{Z_i(x, \omega) - Z_i(y, \omega)\} \right\| \leq \beta_n(\omega)\|x - y\|, \quad (x, \omega), (y, \omega) \in H \times \Omega_0. \quad (5.15)$$

Hence using (5.12) to (5.15) and by the similar arguments of the proof of (5.11) we can show that

$$\lim_{n \rightarrow \infty} \max_{n \leq k < m(n, T)} \left| \sum_{i=n}^k a_i \langle X_i, Z_i \rangle \right| = 0 \quad a.s. . \quad (5.16)$$

Hence, by (5.12) and (5.16) and using Lemma 2 we obtain (5.1).

Q.E.D.

**REMARK.** Let us consider the following assumption.

**B5:** For any  $T > 0$  there exists a sequence of positive random variables  $\{\beta_n(\omega)\}$  such that (5.13) to (5.15) hold.

And suppose that the conditions of Theorem B hold but with  $\sup_n \|X_n(\omega)\| < \infty$ ,  $\omega \in \Omega_0$  and B'5 replacing B0 and B5, respectively. Then (5.1) also holds.

## References

- [1] BORODIN, A.N.: *A stochastic approximation procedure in the case of weakly dependent observations*, Theory Prob. Appl., **24** (1979), 34–52.
- [2] FRITZ, J.: *Learning from an ergodic training sequence*, In “Limit Theorems of Probability Theory”; ed. P. Révész, North-Holland 1974, 79–91.
- [3] GYÖRFI, L.: *Stochastic approximation from ergodic sample for linear regression*, Z. Wahrscheinlichkeits theorie verw. Geb. **54** (1980), 47–55.
- [4] KUSHNER, H.J. and CLARK, D.S.: *Stochastic approximation methods for constrained and unconstrained systems*, Springer, 1978.
- [5] LJUNG, L.: *Analysis of stochastic gradient algorithms for linear regression problem*, IEEE Information Theory, **IT-30** (1984), 151–160.
- [6] LJUNG, L.: *Analysis of recursive stochastic algorithms*, IEEE Trans. Automat. Contr., **AC-22** (1977), 551–575.

- [7] ROBBINS, H. and MONRO, S.: *A stochastic approximation method*, Ann. Math. Statist., **22** (1951), 400–407.
- [8] WATANABE, M.: *A stochastic approximation from dependent observations*, Z. Wahrscheinlichkeitstheorie verw. Geb., **62** (1983), 279–292.
- [9] WATANABE, M.: *The 2<sup>nd</sup> mean convergence of a stochastic approximation algorithm with weakly dependent observations*, Fukuoka Univ. Science Rep., **16** (1986), 71–80.

*Received November 28, 1990*

*Communicated by N. Furukawa*