and consequently

$$\begin{aligned} P\{x\colon \|x-Q_n(x)\|\geq\delta\} &\geq P\{x\in V_n\colon \|x-Q_n(x)\|\geq\delta\} \\ &= P(V_n)-P\{x\in V_n\colon \|x-Q_n(x)\|<\delta\} \\ &\geq P(V_n)-\epsilon/4. \end{aligned}$$

By definition of V_n and the relation (6)

$$\limsup_{n\to\infty} P\{x\colon \|x-Q_n(x)\|\geq\delta\}\geq\epsilon/3-\epsilon/4=\epsilon/12$$

and therefore $\limsup_{n\to\infty} D_r(Q_n)>0$ for every $r>0$. □

ACKNOWLEDGMENT

The author wishes to thank R. Alexander for bringing the paper [1] to his attention.

REFERENCES

[1] R. Alexander, "A problem about lines and ovals," *Amer. Math. Monthly*, vol. 75, no. 5, pp. 482–487, 1968.
[2] J. A. Bucklew and G. L. Wise, "Multidimensional asymptotic quantization theory with rth distortion measures," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 239–247, 1982.
[3] S. Cambains and N. Gerr, "A simple class of asymptotically optimal quantizers," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 664–676, Sept. 1983.
[4] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 373–380, 1979.
[5] R. M. Gray, K. L. Oehler, K. O. Perlmutter, and R. A. Olshen, "Combining treestructured vector quantization with classification and regression trees," in *Proc. 27th Asilomar Conf. on Circuits Systems and Computers*, 1993.
[6] G. Lugosi and A. B. Nobel, "Consistency of data-driven histogram methods for density estimation and classification," to appear in *Annals of Statistics*, 1996.
[7] S. Na and D. L. Neuhoff, "Bennett's integral for vector quantizers," *IEEE Trans. Inform. Theory*, vol. 41, pp. 886–900, 1995.
[8] D. L. Neuhoff, "On the asymptotic distribution of the errors in vector quantization," *IEEE Trans. on Inform. Theory*, vol. 42, no. 2, pp. 461–468, Mar. 1996.
[9] A. B. Nobel, "Recursive partitioning to reduce distortion," Tech. Rep. UIUC-BI-95-01, Beckman Institute, University of Illinois, Urbana-Champaign, 1995.
[10] ——, "Histogram regression estimation using data-dependent partitions," to appear in *Annals Stat.*
[11] K. L. Oehler, P. C. Cosman, R. M. Gray, and J. May, "Classification using vector quantization," in *Proc. 25th Asilomar Conf. on Signals, Systems and Computers* (Pacific Grove, CA, Nov. 1991), pp. 439–445.
[12] K. L. Oehler and R. M. Gray, "Combining image compression and classification using vector quantizaton," *IEEE Trans. Pattern. Anal. Machine Intell.*, vol. 17, pp. 461–473, 1995.
[13] Q. Xie, C. A. Laszlo, and R. K. Ward, "Vector quantization for nonparametric classifier design," *IEEE Trans. Pattern. Anal. Machine Intell.*, vol. 15, no. 12, pp. 1326–1330, 1993.
[14] Y. Yamada, S. Tazaki, and R. M. Gray, "Asymptotic performance of block quantizers with difference distortion measures," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 6–14, 1980.
[15] P. Zador, "Asymptotic quantization of continuous random variables," unpublished memo., Bell Labs., 1966.
[16] ——, "Asymptotic quantization error of continuous signals and quantization dimension," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 139–149, 1982.

On the Convergence of Linear Stochastic Approximation Procedures

Michael A. Kouritzin

***Abstract*—Many stochastic approximation procedures result in a stochastic algorithm of the form**

$$h_{k+1}=h_k+\frac{1}{k}(b_k-A_kh_k),\quad \text{for all } k=1,2,3,\cdots. \tag{1}$$

Here, $\{b_k, k=1,2,3,\cdots\}$ is a $\mathbb{R}^d$-valued process, $\{A_k, k=1,2,3,\cdots\}$ is a symmetric, positive semidefinite $\mathbb{R}e^{d\times d}$-valued process, and $\{h_k, k=1,2,3,\cdots\}$ is a sequence of stochastic estimates which hopefully converges to

$$h\triangleq\left[\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}EA_k\right]^{-1}\cdot\left\{\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}Eb_k\right\} \tag{2}$$

(assuming everything here is well defined). In this correspondence, we give an elementary proof which relates the almost sure convergence of $\{h_k, k=1,2,3,\cdots\}$ to strong laws of large numbers for $\{b_k, k=1,2,3,\cdots\}$ and $\{A_k, k=1,2,3,\cdots\}$.

***Index Terms*— Almost sure convergence, stochastic approximation, adaptive filtering, recursive algorithms, dependent random variables, Robbins–Monro process.**

I. INTRODUCTION

Since the inception of stochastic approximation procedures many statisticians, probabilists, and engineers have strived to establish limit theorems and invariance principles for these procedures. Much of the earlier effort (see, e.g., Sacks [21], Fabian [7], McLeish [17], Gaposhkin and Krasulina [10], Heyde [13], and Ruppert [20]) was concerned with procedures which can be written in the algorithmic form

$$h_{k+1}=h_k+\frac{1}{k}(b_k-A_kh_k)\quad\text{for all}\quad k=1,2,3,\cdots \tag{1.1}$$

where h_1 is some possibly random vector $\{A_k, k=1,2,3,\cdots\}$ and $\{b_k, k=1,2,3,\cdots\}$ are, respectively, $\mathbb{R}^{d\times d}$-valued and $\mathbb{R}^d$-valued processes on some probability space $(\Omega,\mathcal{F},P)$, and A_k is constant or at least converges almost surely to some positive-definite matrix A. More recently, applications of the procedures (1.1) where A_k is symmetric and positive semidefinite but does not converge (see, e.g., Widrow and Stearns [24, ch. 6], Benveniste, Métivier, and Priouret [1, ch. 1], and the introduction of Farden [8]) have prompted many authors (see, e.g., Fritz [9], Györfi [11], Farden [8], Eweda and Macchi [5], [6], Ljung [16], and Métivier and Priouret [18]) to study strong consistency for (1.1) under less-stringent conditions on $\{A_k, k=1,2,3,\cdots\}$. In the present note, we bring forth seemingly natural almost sure convergence results for (1.1) analogous to the strong laws of large numbers for partial sums of random variables. Although our results are in some respects more general than previous results, our main contribution might be considered our elementary proof which was motivated in part by Fabian [7, Lemma 2.1] and Eweda and Macchi [5]. (There are also some similarities between

Manuscript received October 1, 1994; revised December 19, 1995. This work was done while the author was visiting the Laboratory for Research in Statistics and Probability, Carleton University, Ottawa, Ont., Canada K1S 5B6.

The author is with the Institute for Mathematics and Its Applications, University of Minnesota, Minneapolis, MN 55455-0436 USA.

Publisher Item Identifier S 0018-9448(96)03642-5.

the matrix computations in Bitmead [2] and those in the sequel.) Finally, we mention that there is substantial literature (see, e.g., the books of Kushner and Clark [14] and Benveniste, Métivier, and Priouret [1]) motivating and treating nonlinear (in h_k) stochastic approximation procedures as well as procedures with state-dependent noise. However, to apply such results to our algorithm (1.1), one would necessarily have to impose more stringent conditions than those proposed in this correspondence.

II. Notation and Results

In this section, we will define our notation and provide our results.

A. Notation List

$|x|$ is the Euclidean distance of some $\mathbb{R}^d$-vector x.

$\|A\| = \sup_{|x|=1} |Ax|$ for any $\mathbb{R}^{d\times d}$ matrix A.

$\lfloor t \rfloor \triangleq \max\{i \in \mathbb{N}_0: i \le t\}$ and

$\lceil t \rceil \triangleq \min\{i \in \mathbb{N}_0: i \ge t\}$, for any $t \ge 0$.

$a \ll b$ means that a, b are nonnegative real numbers such that $b < \infty$ implies $a < \infty$ and $b = 0$ implies $a = 0$.

$a_{i,k} \overset{i,k}{\ll} b_{i,k}$ means that there is a $c > 0$ not depending on i or k such that $|a_{i,k}| \le c|b_{i,k}|$ for all i, k.

$I^d = d \times d$ identity matrix.

$\prod_{l=p}^{q} B_l$ (with each B_l being a $\mathbb{R}^{d\times d}$-matrix) $= B_q B_{q-1} \cdots B_p$ if $q \ge p$ or I^d if $p > q$.

$a \wedge p, a \vee p$ are, respectively, the minimum and maximum of a and p.

B. The Main Result and Discussion

We will state and prove our result in a completely deterministic manner and then apply this result on a sample path by sample path basis. Therefore, we assume that d is a positive integer, $\{\overline{A}_k\}_{k=1}^{\infty}$ is a symmetric, positive semidefinite $\mathbb{R}^{d\times d}$-valued sequence, and $\{\overline{b}_k\}_{k=1}^{\infty}$ is a $\mathbb{R}^d$-valued sequence. The main result is now stated:

Propostion 1: Suppose

$$\limsup_{n\to\infty} \left\| 1/n \sum_{k=1}^{n} \overline{A}_k - A \right\| = 0$$

for some (symmetric) positive-definite A and $\{\overline{h}_k\}_{k=1}^{\infty}$ is a $\mathbb{R}^d$-valued sequence satisfying

$$\overline{h}_{k+1} = \overline{h}_k + \frac{1}{k}(\overline{b}_k - \overline{A}_k \overline{h}_k), \quad \text{for all } k = 1, 2, 3, \cdots. \tag{2.1}$$

Then, a necessary and sufficient condition for $\overline{h}_k \to h$ (with $h \in \mathbb{R}^d$) as $k \to \infty$ is that

$$\frac{1}{n} \sum_{k=1}^{n} \overline{b}_k \to b \triangleq Ah, \quad \text{as } n \to \infty.$$

To see the generality of our result, we suppose that $\{b_k, k = 1, 2, 3, \cdots\}$ is a $\mathbb{R}^d$-valued stochastic process on $(\Omega, \mathcal{F}, P)$ and $\{A_k, k = 1, 2, 3, \cdots\}$ is a symmetric, positive semidefinite $\mathbb{R}^{d\times d}$-valued process on $(\Omega, \mathcal{F}, P)$ such that

$$A \triangleq \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} E\{A_k\}$$

and

$$b \triangleq \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} E\{b_k\} \tag{2.2}$$

are well-defined and A is positive-definite. Then, Propostion 1 (with $\overline{A}_k = A_k(\omega)$ and $\overline{b}_k = b_k(\omega)$ for all k) implies $h_k(\omega)$, the solution of (1.1), converges to $h \triangleq Ab$ provided

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} (A_k - E\{A_k\}) = 0$$

and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} (b_k - E\{b_k\}) = 0. \tag{2.3}$$

Hence, the generality of our approach follows from the following remark:

Remark: In many applications

$$A_k = \frac{1}{M} \sum_{l=\max\{k-M+1,1\}}^{k} Y_l Y_l^T$$

and

$$b_k = \frac{1}{M} \sum_{l=\max\{k-M+1,1\}}^{k} \alpha_{l+1} Y_l, \quad \text{for all } k = 1, 2, 3, \cdots$$

some fixed positive integer M, $\mathbb{R}^d$-valued process $\{Y_l, l = 1, 2, 3, \cdots\}$ and $\mathbb{R}$-valued process $\{\alpha_l, l = 1, 2, 3, \cdots\}$. In fact, past values of α_{l+1} usually make up some of the components of Y_l. The result of all this is that $\{A_k, k = 1, 2, 3, \cdots\}$ often satisfies the same type of moment and dependency conditions as $\{b_k, k = 1, 2, 3, \cdots\}$ does. Hence, to verify (2.3), the same version of the strong law of large numbers can often be applied to all of the components of $\{A_k, k = 1, 2, 3, \cdots\}$ and $\{b_k, k = 1, 2, 3, \cdots\}$. For various conditions under which the strong law of large numbers holds, we refer the reader to Stout [23, pp. 165, 181, 199], Chow and Teicher [3, p. 397], Hall and Heyde [12, pp. 40–41], Phillips and Solo [19], Shao [22, Corollary 1], and Lai and Stout [15, Theorem 7]. Indeed, it is a simple exercise to show that our result combined with Lai and Stout's result generalize the results of Eweda and Macchi [5]. Moreover, as compared to Eweda and Macchi [6], one easily sees that we study a more general algorithm, do not require their moment bounds (see [6, eq. (35)]), and make use of weaker eigenvalue and law of large numbers conditions.

Remark: A reviewer has pointed out that there are interesting applications in adaptive control where $-A$ has eigenvalues in the open left half plane but may not be symmetric. However, it is the author's opinion that the symmetry assumption of A_k cannot be relaxed without imposing alternate conditions and significantly modifying our arguments (especially Lemma B, (3.15), and (3.22), (3.23)). The nonsymmetric case will not be treated here.

III. The Proof of Proposition 1

For notational convenience, we define

$$v_k \triangleq \overline{h}_k - h, \quad Y_k \triangleq \overline{A}_k - A$$

and

$$z_k \triangleq \overline{b}_k - \overline{A}_k h, \quad \text{for all } k = 1, 2, 3, \cdots. \tag{3.1}$$

Hence, observing that $z_k = \overline{b}_k - b - Y_k h$ and noting by hypothesis that

$$\frac{1}{n} \sum_{k=1}^{n} Y_k \to 0$$

we only need to show $v_k \to 0$ if and only if

$$\frac{1}{n} \sum_{k=1}^{n} z_k \to 0.$$

Indeed, we obtain from (3.1) and (2.1) that

$$\frac{1}{n}\left|\sum_{j=1}^{n} z_j\right| = \frac{1}{n}\left|\sum_{j=1}^{n} j(v_{j+1}-v_j) + \sum_{j=1}^{n} \overline{A}_j v_j\right| \le |v_{n+1}| + \frac{1}{n}\left|\sum_{j=1}^{n} \overline{A}_j v_j - \sum_{j=1}^{n} v_j\right| \quad (3.2)$$

for $n = 1, 2, 3, \cdots$. Moreover, by Lemma B (to follow) and the hypothesis it follows that

$$\sup_{n\ge 1}\sum_{j=1}^{n}\frac{1+\|\overline{A}_j\|}{n} \le \sup_{n\ge 1}\left\{1+\frac{d}{n}\left\|\sum_{j=1}^{n}\overline{A}_j\right\|\right\} < \infty. \quad (3.3)$$

Therefore, the "only if" part follows by (3.2), (3.3), and the Toeplitz lemma [23, Lemma 3.2.3 (ii), p. 120]. Next, we assume

$$\frac{1}{n}\sum_{k=1}^{n} z_k \to 0$$

and show $u_n \to 0$ and $w_n \triangleq v_n - u_n \to 0$, where

$$u_{n+1} = (I^d - n^{-1}A)u_n + n^{-1}z_n, \quad \text{for } n = 1, 2, 3, \cdots \quad (3.4)$$

subject to $u_1 = v_1$. Indeed, with the definition

$$F_{j,n} \triangleq \frac{1}{j}\prod_{r=j+1}^{n-1}\left(I^d - \frac{1}{r}A\right), \quad \text{for } j = 1, 2, \cdots, n-1, n = 2, 3, 4, \cdots \quad (3.5)$$

it follows by (3.4), (3.5), and Lemma A i) and ii) (to follow) that

$$\lim_{n\to\infty} |u_n| \le \lim_{n\to\infty}\left\|\prod_{l=1}^{n-1}(I^d - l^{-1}A)\right\|\cdot|u_1| + \lim_{n\to\infty}\left|\sum_{j=1}^{n-1}F_{j,n}z_j\right| = 0. \quad (3.6)$$

Moreover; letting $\lambda_{\min}$ be the smallest eigenvalue of A and $a > 1$ be a number small enough that

$$\log(a) < \min\left\{\frac{2}{\lambda_{\min}+\|A\|}, \frac{1}{d\|A\|}, \frac{\lambda_{\min}}{(d\|A\|\exp(1))^2}\right\} \quad (3.7)$$

and defining

$$n_k \triangleq \lceil a^k \rceil$$

and

$$I_k \triangleq \{n_k, n_k+1, \cdots, n_{k+1}-1\}, \quad \text{for } k = 0, 1, 2, \cdots \quad (3.8)$$

we observe by (3.1), (2.1), and (3.4) that

$$w_{n+1} = \left(I^d - \frac{1}{n}\overline{A}_n\right)w_n - \frac{1}{n}Y_n u_n, \quad \text{for } n = 1, 2, 3, \cdots \quad (3.9)$$

$$w_{n_{k+1}} = U_k w_{n_k} - \sum_{j\in I_k} V_{j,k}u_j, \quad \text{for } k = 0, 1, 2, \cdots \quad (3.10)$$

where

$$U_k \triangleq \prod_{l\in I_k}\left(I^d - \frac{1}{l}\overline{A}_l\right)$$

and

$$V_{j,k} \triangleq \prod_{l=j+1}^{n_{k+1}-1}\left(I^d - \frac{1}{l}\overline{A}_l\right)\frac{1}{j}Y_j. \quad (3.11)$$

Now, we use (3.11) and Lemma A v) to obtain that

$$\|V_{j,k}\| \le \prod_{l\in I_k}\left[1+\frac{\|\overline{A}_l\|}{l}\right]\frac{\|Y_j\|}{j} \overset{j,k}{\ll} \frac{\|Y_j\|}{j}, \quad \text{for } k = 0, 1, 2, \cdots \text{ and } j \in I_k \quad (3.12)$$

and (3.11), (3.1), and symmetry to obtain that

$$\begin{aligned}\|U_k\| &\le \left\|I^d - \sum_{l\in I_k}\frac{1}{l}\overline{A}_l\right\| + \sum_{\substack{l_1,l_2\in I_k\\ l_1>l_2}}\frac{\|\overline{A}_{l_1}\|}{l_1}\cdot\frac{\|\overline{A}_{l_2}\|}{l_2} + \sum_{\substack{l_1,l_2,l_3\in I_k\\ l_1>l_2>l_3}}\frac{\|\overline{A}_{l_1}\|}{l_1}\cdot\frac{\|\overline{A}_{l_2}\|}{l_2}\cdot\frac{\|\overline{A}_{l_3}\|}{l_3}\\ &\quad + \cdots + \left[\frac{\|\overline{A}_{n_k}\|}{n_k}\cdot\frac{\|\overline{A}_{n_k+1}\|}{n_k+1}\cdots\frac{\|\overline{A}_{n_{k+1}-2}\|}{n_{k+1}-2}\cdot\frac{\|\overline{A}_{n_{k+1}-1}\|}{n_{k+1}-1}\right]\\ &\le \left\|I^d - A\sum_{l\in I_k}\frac{1}{l}\right\| + \left\|\sum_{l\in I_k}\frac{1}{l}Y_l\right\| + \sum_{m=2}^{n_{k+1}-n_k}\frac{\left(\sum_{l\in I_k}\frac{\|\overline{A}_l\|}{l}\right)^m}{m!}\end{aligned} \quad (3.13)$$

for $k = 0, 1, 2, \cdots$. Moreover, by (3.8) we have that

$$\lim_{k\to\infty}\left|\sum_{l\in I_k}\frac{1}{l} - \log(a)\right| \le \lim_{k\to\infty}\max\left\{(\lceil a^k\rceil)^{-1}, \log\left(\frac{\lceil a^{k+1}\rceil}{\lceil a^k\rceil a}\right)\right\} = 0 \quad (3.14)$$

and, letting $\epsilon > 0$ be arbitrary, recalling some basic theory for symmetric matrices (see, e.g., [25, pp. 57–8]), and using the fact that $(\lambda_{\min} + \|A\|)\log(a) < 2$, we find by (3.8) that there is some K_ϵ such that

$$\left\|I^d - A\sum_{l\in I_k}\frac{1}{l}\right\| = \max\left\{\|A\|\sum_{l\in I_k}\frac{1}{l} - 1, 1 - \lambda_{\min}\sum_{l\in I_k}\frac{1}{l}\right\} \le 1 - \lambda_{\min}\log(a) + \epsilon, \quad \text{for all } k \ge K_\epsilon. \quad (3.15)$$

Hence, we can use (3.13), (3.15), Lemma A iii), (3.1), Lemma B, Taylor's theorem, (3.14), and the fact $d\|A\|\log(a) < 1$ to obtain a $K'_\epsilon \ge K_\epsilon$ such that

$$\begin{aligned}\|U_k\| &\le 1 - \lambda_{\min}\log(a) + 2\epsilon + \sum_{m=2}^{\infty}\frac{\left[d\|A\|\sum_{l\in I_k}\frac{1}{l} + d\left\|\sum_{l\in I_k}\frac{1}{l}Y_l\right\|\right]^m}{m!}\\ &\le 1 - \lambda_{\min}\log(a) + 2\epsilon + \exp\left\{d\|A\|\sum_{l\in I_k}\frac{1}{l} + \epsilon\right\} - 1 - d\|A\|\sum_{l\in I_k}\frac{1}{l} - \epsilon\\ &\le 1 - \lambda_{\min}\log(a) + 2\epsilon + \exp\{1+\epsilon\}\cdot\frac{(d\|A\|\log(a) + 2\epsilon)^2}{2}, \quad \text{for all } k \ge K'_\epsilon.\end{aligned} \quad (3.16)$$

Therefore, using the fact

$$\log(a) < \frac{\lambda_{\min}}{(d\cdot\|A\|\cdot\exp(1))^2}$$

and making $\epsilon > 0$ sufficiently small, we discover from (3.16) that there exists a $0 < \gamma < 1$ and an integer $k_1 > 0$ such that

$$\|U_k\| \le \gamma, \quad \text{for all } k \ge k_1. \quad (3.17)$$

It follows by (3.10), (3.17), (3.12), and (3.1) that

$$|w_{n_k}| \overset{k}{\ll} \gamma^{k-k_1}|w_{n_{k_1}}| + \sum_{l=k_1}^{k-1} \gamma^{k-l-1} \cdot \sum_{j\in I_l} \frac{\|A\| + \|\overline{A}_j\|}{j}|u_j|, \quad \text{for all } k \geq k_1. \tag{3.18}$$

However, by Lemma A iv), (3.14), (3.6), and the Toeplitz lemma

$$\lim_{l\to\infty} \sum_{j\in I_l} \frac{\|A\| + \|\overline{A}_j\|}{j}|u_j| = 0. \tag{3.19}$$

Furthermore, since

$$\sum_{l=k_1}^{k-1} \gamma^{k-l-1} = \frac{\gamma - \gamma^{k+1-k_1}}{1-\gamma} \overset{k}{\ll} 1, \quad \text{for all } k = k_1, k_1+1, \cdots \tag{3.20}$$

it follows by (3.19), (3.20), the Toeplitz lemma, and (3.18) that $\lim_{k\to\infty} |w_{n_k}| = 0$. Finally, using (3.9) and (3.1), one has for $n \in I_k$ that

$$\begin{aligned} |w_n| &\leq \prod_{l=n_k}^{n-1} \left\| I^d - \frac{1}{l}\overline{A}_l \right\| \cdot |w_{n_k}| \\ &\quad + \sum_{j=n_k}^{n-1} \prod_{l=j+1}^{n-1} \left\| I^d - \frac{1}{l}\overline{A}_l \right\| \cdot \frac{\|Y_j\|}{j} \cdot |u_j| \\ &\leq \prod_{l\in I_k} \left(1 + \frac{\|\overline{A}_l\|}{l}\right) \cdot \left[|w_{n_k}| + \sum_{j\in I_k} \frac{\|A\| + \|\overline{A}_j\|}{j}|u_j| \right] \end{aligned} \tag{3.21}$$

so by (3.21), Lemma A v), and (3.19) one finds that

$$\lim_{k\to\infty} \max_{n\in I_k} |w_n| = 0. \qquad \square$$

Lemma A: Suppose $\{\overline{A}_k\}_{k=1}^{\infty}$ and A are as in the statement of Propostion 1; $I_k, F_{j,k}, \{z_k\}_{k=1}^{\infty}$ and $\{Y_k\}_{k=1}^{\infty}$ are as defined in (3.8), (3.5), and (3.1) of the proof of Propostion 1; and

$$\lim_{n\to\infty} \left| \frac{1}{n} \sum_{j=1}^{n-1} z_j \right| = 0.$$

Then, the following are true:

i) $\displaystyle \lim_{n\to\infty} \left\| \prod_{l=1}^{n-1} (I^d - l^{-1}A) \right\| = 0$

ii) $\displaystyle \lim_{n\to\infty} \left| \sum_{j=1}^{n-1} F_{j,n} z_j \right| = 0$

iii) $\displaystyle \lim_{k\to\infty} \left\| \sum_{l\in I_k} \frac{1}{l} Y_l \right\| = 0$

iv) $\displaystyle \sum_{j\in I_l} \frac{\|\overline{A}_j\|}{j} \overset{l}{\ll} 1, \quad \text{for all } l = 0, 1, 2, \cdots$

v) $\displaystyle \prod_{l\in I_k} \left(1 + \frac{\|\overline{A}_l\|}{l}\right) \overset{k}{\ll} 1, \quad \text{for all } k = 0, 1, 2, \cdots.$

Proof: i) Let $\lambda, \lambda_{\min} > 0$ denote an arbitrary and the minimum eigenvalue of A. Then

$$\begin{aligned} &\left\| \prod_{l=1}^{n-1} \left(1 - \frac{\lambda}{l}\right) I^d \right\| \\ &\leq \prod_{l=1}^{(n-1)\wedge\lfloor\lambda\rfloor} \left(\frac{\lambda}{l} - 1\right) \cdot \prod_{l=\lfloor\lambda\rfloor+1}^{n-1} \left(1 - \frac{\lambda}{l}\right) \\ &\leq \left[\prod_{l:(\lambda/l)>2} \left(\frac{\lambda}{l} - 1\right) \cdot \exp\left\{ \sum_{l=1}^{\lfloor\lambda\rfloor} \frac{\lambda}{l} \right\} \right] \exp\left\{ -\sum_{l=1}^{n-1} \frac{\lambda}{l} \right\} \\ &\overset{n}{\ll} \exp\left\{ -\lambda \int_1^n t^{-1}\, dt \right\} \overset{n}{\ll} n^{-\lambda_{\min}} \\ &\qquad \text{for all } n = 1, 2, 3, \cdots \end{aligned} \tag{3.22}$$

and it follows easily from (3.22), the fact that the eigenvectors of A span $\mathbb{R}^d$ and the principle of uniform boundedness that

$$\lim_{n\to\infty} \left\| \prod_{l=1}^{n-1} (I^d - l^{-1}A) \right\| \ll \lim_{n\to\infty} n^{-\lambda_{\min}} = 0. \tag{3.23}$$

ii) Recalling definition (3.5) and following the arguments in (3.22) and (3.23), one finds that

$$\begin{aligned} \|F_{r-1,n} - F_{r,n}\| &\leq \left\| \frac{1}{r-1}\left(I^d - \frac{1}{r}A\right) - \frac{1}{r}I^d \right\| \cdot \left\| \prod_{l=r+1}^{n-1} \left(I^d - \frac{1}{l}A\right) \right\| \\ &\overset{r,n}{\ll} \frac{1}{(r-1)r} \left(\frac{r+1}{n}\right)^{\lambda_{\min}}, \\ &\qquad \text{for all } r = 2, 3, \cdots, n-1,\ n = 3, 4, \cdots. \end{aligned} \tag{3.24}$$

Hence, it follows easily by (3.24) that

$$\sum_{r=2}^{n-1} (r-1)\|F_{r-1,n} - F_{r,n}\| \overset{n}{\ll} 1, \quad \text{for all } n = 3, 4, \cdots \tag{3.25}$$

and by (3.5), (3.25), the fact that

$$\frac{1}{n} \sum_{j=1}^{n} z_j \to 0$$

as $n \to \infty$, and the Toeplitz lemma that

$$\begin{aligned} \lim_{n\to\infty} \left| \sum_{j=1}^{n-1} F_{j,n} z_j \right| &\leq \lim_{n\to\infty} \left| \sum_{j=1}^{n-1} F_{n-1,n} z_j \right| \\ &\quad + \lim_{n\to\infty} \left| \sum_{j=1}^{n-1} \sum_{r=j+1}^{n-1} (F_{r-1,n} - F_{r,n}) z_j \right| \\ &\leq \lim_{n\to\infty} \frac{1}{n-1} \left| \sum_{j=1}^{n-1} z_j \right| \\ &\quad + \lim_{n\to\infty} \sum_{r=2}^{n-1} \left[\|F_{r-1,n} - F_{r,n}\| \cdot \left| \sum_{j=1}^{r-1} z_j \right| \right] = 0. \end{aligned} \tag{3.26}$$

iii) It follows by an interchange of summation that

$$\begin{aligned} \left\| \sum_{l\in I_k} \frac{1}{l} Y_l \right\| &\leq \frac{1}{n_k} \left\| \sum_{l\in I_k} Y_l \right\| + \left\| \sum_{\substack{r,l\in I_k \\ r<l}} \left(\frac{1}{r+1} - \frac{1}{r}\right) Y_l \right\| \\ &\leq \frac{1}{n_k} \left\{ \left\| \sum_{l<n_{k+1}} Y_l \right\| + \left\| \sum_{l<n_k} Y_l \right\| \right\} \\ &\quad + \sum_{r\in I_k} \frac{1}{r^2} \left\{ \left\| \sum_{l<n_{k+1}} Y_l \right\| + \left\| \sum_{l\leq r} Y_l \right\| \right\} \end{aligned} \tag{3.27}$$

and by (3.8) and the hypothesis that

$$\lim_{k\to\infty}\max_{r\in I_k}\frac{1}{r}\left\|\sum_{l<n_{k+1}} Y_l\right\| = a\lim_{k\to\infty}\frac{1}{n_{k+1}-1}\left\|\sum_{l<n_{k+1}} Y_l\right\| = 0. \tag{3.28}$$

iii) follows by (3.27), the hypothesis, (3.28), (3.14) of Proposition 2, and the Toeplitz lemma.

iv) Using Lemma B, we find iv) follows from iii) as well as (3.1) and (3.14) of Propostion 2.

v) This follows from iv) and the fact that

$$\prod_{l\in I_k}(1+\|\overline{A}_l\|/l) \le \exp\left(\sum_{l\in I_k}\|\overline{A}_l\|/l\right)$$

for all $k = 0, 1, \cdots$. □

Lemma B: Suppose m is a positive integer and $\{M_k, k = 1, 2, 3, \cdots\}$ is a sequence of symmetric, positive semidefinite $\mathbb{R}^{m\times m}$-matrices. Then it follows that

$$\sum_{k=1}^{j}|||M_k||| \le m\left|\left|\left|\sum_{k=1}^{j} M_k\right|\right|\right|, \quad \text{for all } j = 1, 2, 3, \cdots.$$

Proof: It follows by [4, Proposition D.1.2] of Davis and Vinter that there exists a sequence of $\mathbb{R}^{m\times m}$-matrices $\{P_k\}_{k=1}^{\infty}$ such that $M_k = P_k P_k^T$ for $k = 1, 2, 3, \cdots$ and this implies that

$$\begin{aligned} m\left|\left|\left|\sum_{k=1}^{j} M_k\right|\right|\right| &\ge \sum_{k=1}^{j}\sum_{n=1}^{m} M_k^{(n,n)} = \sum_{k=1}^{j}\sum_{n=1}^{m}\sum_{o=1}^{m}(P_k^{(n,o)})^2 \\ &\ge \sum_{k=1}^{j}|||P_k|||\cdot|||P_k^T||| \ge \sum_{k=1}^{j}|||M_k||| \\ &\qquad \text{for all } j = 1, 2, 3, \cdots \end{aligned} \tag{3.29}$$

where $M_k^{(n,o)}$ and $P_k^{(n,o)}$ denote the (n, o)th components of M_k and P_k. □

REFERENCES

[1] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations.* Berlin-Heidelberg-New York: Springer-Verlag, 1990.

[2] R. R. Bitmead, "Convergence properties of LMS adaptive estimators with unbounded dependent inputs," *IEEE Trans. Automat. Contr.*, vol. AC-29, pp. 477–479, 1984.

[3] Y. S. Chow and H. Teicher, *Probability Theory*, 2nd ed. NewYork: Springer-Verlag, 1988.

[4] M. H. A. Davis and R. B. Vinter, *Stochastic Modeling and Control.* London-New York: Chapman and Hall, 1985.

[5] E. Eweda and O. Macchi, "Convergence of an adaptive linear estimation algorithm," *IEEE Trans. Automat. Contr.*, vol. AC-29, no. 2, pp. 119–127, 1984.

[6] ——, "Convergence of the rls and lms adaptive filters," *IEEE Trans. Circuits Syst.*, vol. CAS-34, no. 7, pp. 799–803, 1987.

[7] V. Fabian, "On asymptotic normality in stochastic approximation," *Ann. Math. Statist.*, vol. 39, pp. 1327–1332, 1968.

[8] D. C. Farden, "Stochastic approximation with correlated data," *IEEE Trans. Inform. Theory*, vol. IT-27, no. 1, pp. 105–113, 1981.

[9] J. Fritz, "Learning from an ergodic training sequence," in *Limit Theorems of Probability Theory*, P. Révész, Ed. Amsterdam, The Netherlands: North-Holland, 1974.

[10] V. F. Gaposhkin and T. P. Krasulina, "On the law of the iterated logarithm in stochastic approximation processes," *Theory Probab. Appl.*, vol. 19, pp. 844–850, 1974.

[11] L. Györfi, "Stochastic approximation from ergodic sample for linear regression," *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, vol. 54, pp. 47–55, 1980.

[12] P. Hall and C. C. Heyde, *Martigale Limit Theory and Its Application.* New York: Academic Press, 1980.

[13] C. C. Heyde, "On martingale limit theory and strong convergence results for stochastic approximation procedures," *Stochastic Processes Appl.*, vol. 2, pp. 359–370, 1974.

[14] H. J. Kushner and D. S. Clark, *Stochastic Approximation for Constrained and Unconstrained Systems* (Applied Math. Sci. Ser. no. 26). New York: Springer, 1978.

[15] T. L. Lai and W. Stout, "Limit theorems for sums of dependent random variables," *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, vol. 51, pp. 1–14, 1980.

[16] L. Ljung, "Analysis of stochastic gradient algorithms for linear regression problems," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 151–160, 1984.

[17] D. L. McLeish, "A maximal inequality and dependent strong laws," *Ann. Probab.*, vol. 3, no. 5, pp. 829–839, 1975.

[18] M. Métivier and P. Priouret, "Applications of a Kushner and Clark lemma to general classes of stochastic algorithms," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 140–151, 1984.

[19] P. C. B. Phillips and V. Solo, "Asymptotics for linear processes," *Ann. Statist.*, vol. 20, pp. 971–1001, 1992.

[20] D. Ruppert, "Almost sure approximations to the Robbins-Monro and Kiefer-Wolfowitz processes with dependent noise," *Ann. Probab.*, vol. 10, pp. 178–187, 1982.

[21] J. Sacks, "Asymptotic distribution of stochastic approximation procedures," *Ann. Math. Statist.*, vol. 29, pp. 373–405, 1958.

[22] Q.-M. Shao, "Complete convergence for α-mixing sequences," *Statist. and Prob. Lett.*, vol. 16, pp. 279–287, 1993.

[23] W. F. Stout, *Almost Sure Convergence.* New York: Academic Press, 1974.

[24] B. Widrow and S. D. Stearns, *Adaptive Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1985.

[25] J. H. Wilkinson, *The Algebraic Eigenvalue Problem.* Oxford, UK: Clarendon, 1965.