ANNALES DE L'I. H. P., SECTION B

E. EWEDA
O. MACCHI

Quadratic mean and almost-sure convergence of unbounded stochastic approximation algorithms with correlated observations

Annales de l'I. H. P., section B, tome 19, nº 3 (1983), p. 235-255
<http://www.numdam.org/item?id=AIHPB_1983__19_3_235_0>

Ann. Inst. Henri Poincaré, Section B :
Vol. XIX, n° 3-1983, p. 235-255. *Calcul des Probabilités et Statistique.*

Quadratic mean and almost-sure convergence of unbounded stochastic approximation algorithms with correlated observations

by

E. EWEDA (*) and O. MACCHI (**)

RÉSUMÉ. — Dans ce travail nous démontrons la convergence presque sûre et en moyenne quadratique d'un algorithme de gradient stochastique avec des pas décroissants gouvernant un estimateur linéaire adaptatif. Nous n'utilisons aucune hypothèse irréaliste telle que l'indépendance des observations successives ou la bornitude de l'algorithme comme dans la littérature antérieure. Pour les observations corrélées, nous utilisons un modèle à mémoire finie, ce qui correspond à une vaste classe d'application, étant donné que des observations suffisamment séparées dans le temps peuvent généralement être supposées indépendantes. Le modèle à mémoire finie a l'avantage, par rapport à d'autres modèles ergodiques tels que les modèles à covariance décroissante, de permettre une analyse relativement simple de la convergence de l'algorithme du gradient. De plus, le modèle permet de démontrer la convergence en moyenne quadratique, ce qui n'avait pas été fait jusqu'ici pour aucun modèle ergodique.

ABSTRACT. — In this paper we prove the almost-sure (a. s.) and quadratic mean (q. m.) convergence of a stochatic gradient algorithm with decreasing step size governing an adaptive linear estimator. We do not use unrealistic assumptions such as the independence of successive observations or the boundedness of the algorithm, as has been previously done

(*) Military Technical College, Cairo, Egypt. Mailing address : 12 A Alexandre Zaki Street, Helmiat el-Zaitoun, Cairo, Egypt.
(**) C. N. R. S.-E. S. E., Plateau du Moulon, 91190 Gif-sur-Yvette, France.

Annales de l'Institut Henri Poincaré-Section B-Vol. XIX, 0020-2347/1983/235/$ 5,00/
© Gauthier-Villars

in literature. The model we use for the correlated observations is a finite memory model. This model agrees with a wide class of applications due to the fact that sufficiently time separated observations can usually be assumed independent. The finite memory model has the advantage, with respect to other ergodic models such as models with decreasing covariance, that it allows a relatively simple convergence analysis of the gradient algorithm. Moreover, it permits the proof of the q. m. convergence which has not yet been attained with any kind of ergodic models.

I. INTRODUCTION

In this paper we prove both the almost-sure (a. s.) and quadratic mean (q. m.) convergence of the stochastic gradient algorithm

$$h_{j+1}=h_{j+}\mu_j(a_j-h_j^{\mathrm{T}}\mathrm{X}_j)\mathrm{X}_j^* \; ; \quad a_j\in\mathbb{C} \; ; \quad \mathrm{X}_j\in\mathbb{C}^{\mathrm{N}} \; ; \quad h_j\in\mathbb{C}^{\mathrm{N}} \; ; \quad \mathrm{N}\geqslant 1 \,, \qquad (1)$$

to the optimum vector h_* given by

$$\mathrm{R}h_* = \mathrm{E}(a_j\mathrm{X}_j^*) \; ; \qquad \mathrm{R} \triangleq \mathrm{E}(\mathrm{X}_j^*\mathrm{X}_j^{\mathrm{T}}) \,, \qquad (2)$$

where X^{T} and X^* denote respectively the transpose and complex conjugate of the vector X. The step size μ_j of the algorithm is a decreasing sequence of positive numbers satisfying

$$\mu_j = \frac{d_1}{j+d_2} \; ; \qquad 0 < d_1 < \infty \; ; \qquad 0 \leqslant d_2 < \infty \,. \qquad (3)$$

The algorithm (1) is intended to calculate iteratively the vector h_* that minimizes the quadratic mean error $\mathrm{E}(\,|\,a_j - h^{\mathrm{T}}\mathrm{X}_j\,|^2)$ between the message a_j and its linear estimate $h^{\mathrm{T}}\mathrm{X}_j$, based on the observation vector X_j. It is assumed that the sequence (a_j, X_j) ; $j = 1, 2, \ldots$ is stationary and that the covariance matrix R in (2) is invertible.

The algorithm (1) has been used for many years in adaptive signal processing. The experimental and simulation results have shown the convergence of h_j to h_*. However, there is not yet a satisfactory mathematical proof of that convergence under completely realistic assumptions. In earlier analysis of similar algorithms e. g. in [*1*], one finds two major assumptions, usually not fulfilled in the applications encountered in the field of signal processing. The first one is the independence of successive pairs (a_j, X_j). Now in signal processing, X_j is often made of successive time samples of the same analog signal, in such a way that successive observation vectors share $N - 1$ components and are, therefore, strongly correlated. The second unrealistic assumption often used in literature is that the vector h_j is bounded, which implies the use of a reflecting barrier that brings h_j back into a compact set every time it leaves that set. Now applications of the algorithm have shown that such a barrier is not necessary for the convergence. Afterwards, the boundedness assumption has been omitted in [*2-6*]. However, the independence assumption remains in those papers. In a recent work [*7*] Kushner and Clark have overcome the independence problem and proved a. s. convergence of algorithm (1) with correlated observations. Other papers have dealt with the correlated case e. g. Ljung [*8*]. However, the boundedness assumption reappears in these papers. Namely, it is assumed that h_j will return indefinitely within a (random) compact set. Moreover the proofs are extremely difficult to follow.

In the present paper, we overcome both barrier and independence problems and for correlated observations we prove the a. s. and q. m. convergence of h_j to h_*. The assumption we use to attain this result is that the pair (a_j, X_j) has a finite memory and finite moments. The finite memory assumption is plausible in a wide class of applications since sufficiently time separated observations can usually be assumed independent. Also the finite moments assumption is satisfied in a wide class of applications. Two practical examples in which the finite memory and finite moments assumptions are both satisfied are evoked in section III of this paper. It should be mentioned that we do not assume a linear filtering relationship between a_j and X_j. On the contrary, when applying Ljung's theorem [*8*] to the algorithm (1), one is found obliged to assume that a_j and X_j are related by linear filtering, up to some independent additive noise. This is due to the fact that Ljung uses a state space model in his theorem. In the present work we do not use a state space model and the linear filtering relation is not needed; a_k and X_k can be related through any specific non-linearity.

The restrictive independence and boundedness assumptions have also been recently suppressed by Farden [*9*] independently and at the same time as by our own previous works [*10-12*]. In both works [*9*] and [*10*], the convergence proof relies upon an assumption of rational decay of the kind $\dfrac{1}{1 + |k - j|^\beta}$ for the covariance of elements such as $|a_k X_k|$ and other second order observed variables. Admittedly, the theorem presented in this paper is on one hand more restrictive than those of [*9*] and [*10*], because it uses a less general model for the observations (i. e. finite memory). However, it is on the other hand more powerful because it reaches the result of quadratic mean convergence in addition to almost-sure convergence proved in [9], [10]. Actually, the finite memory model derives its interest from the fact that it is a model with dependence for which we prove the q. m. convergence in addition to the a. s. convergence, the former being itself based upon simple moment bounds. Similar bounds could probably be derived for the decaying covariance model, but at the expense of much more technical labor.

The convergence analysis presented in this paper concerns the real version of the algorithm (1), i. e., we assume that a_k, X_k and h_k are real-valued. However, without any additional difficulty, it can be extended to the complex case. An example of the latter case in the context of data transmission is the quadrature amplitude modulation in which the data signal a_k modulates the amplitude and phase of a carrier wave that is demodulated at the receiver by two carriers in quadrature to give the observation signal X_k.

In the following we assume implicity that the sequence (a_k, X_k) is strictly stationary and that its correlation matrix R is invertible.

II. THEOREM

THEOREM. — *If there exist a finite integer* $M' \geq 1$, *and a finite positive* M *such that*

$$\forall k \geqslant 0, \qquad \{ (a_j, X_j) : j \leqslant k \} \qquad \textit{and} \qquad \{ (a_j, X_j) : j \geqslant k + M' \} \quad (A.1)$$

are statistically independent,

$$E(|X_k|^{24M}) < \infty \; ; \qquad M \geqslant M' \; ; \qquad M > \frac{N}{12}, \qquad (A.2)$$

$$E(|a_k|^{[4+4/(6M-1)]}) < \infty \, , \qquad (A.3)$$

then the vector h_k *defined by the algorithm* (1) *tends to* h_* *in both the quadratic mean and almost-sure senses as k tends to infinity.*

The validity of the assumptions (A.1)-(A.3) in applications is emphasized in section III. The assumption (A.1) states that the sequence (a_k, X_k) has a finite strong memory M′. As already mentioned, this assumption is a realistic one since sufficiently time separated observations, can usually in practice be assumed independent. On the other hand, from a theoretical viewpoint this assumption is very convenient because it allows the computation of moments of h_k, due to factorization properties. This leads naturally to a q. m. type convergence. Such a computation of moments is not easy in a decaying covariance model, and therefore with the latter model the ergodicity will rather be used to deal with individual samples and prove a. s. convergence.

The boundedness assumption (A.2) for the moments of observations is satisfied in a wide class of practical applications. Notice that in signal processing applications, where X_k represents a sliding time-window of width N from a sampled signal, the memory M′ is at least equal to N. Hence in assumptions (A.2), (A.3), the integer M can be set at the value M′ of the memory. Therefore we see that the greater the memory, the larger the order of the $|X_k|$-moments that are assumed bounded. Such a relationship is not surprising. Neither is it isolated in the literature; e. g. in condition B.6 of [9] if appears that the required order of bounded moments increases as the rate of decay of covariance decreases, i. e., as the memory increases. The order of moments of $|a_k|$ that are assumed bounded is greater than (but very near to) 4 and less than 5. The reason why the orders of the moments assumed bounded are respectively 24M for $|X_k|$ and $4\left(1 + \frac{1}{6M - 1}\right)$ for $|a_k|$ is a technical one. It is made clear in the course of the proof that follows.

III. PRACTICAL APPLICATIONS

In this section we give two important pratical applications of the theorem in the field of data transmission. In the first example the algorithm (1) is the adaptation algorithm of a transversal equalizer. In such a case a_k is the data transmitted at time k, the observation vector X_k is made of N successive time samples of the channel output, the vector h_k is composed of the equalizer tap coefficients at time k and h_* is the optimum equalization vector. Thus two successive observation vectors share N − 1 components and are, therefore, strongly correlated. Now, a_k is bounded because it is a digital data. Thus (A.3) is satisfied. The observation vector X_k results from filtering the sequence of data by a stable filter $\mathcal{F}$, corresponding to the filtering effect of the transmission channel, to which is added a Gaussian noise. Consequently, all moments of X_k are bounded. Thus, the assumption (A.2) of the theorem is satisfied. The data sequence is usually independent and so also is the noise sequence. Hence, the memory M′ in the assumption (A.1) depends on the memory of $\mathcal{F}$ which can be assumed finite without loss of generality in practical applications. Consequently, the theorem presented in this paper can be used to prove the a. s. and q. m. convergence of the classical adaptive equalizer to the optimum one, a proof which seems to have never been completed before.

In the second practical example, the algorithm (1) is the adaptation algorithm of an echo canceller in full-duplex data transmission over two-wire telephone channels. In such a case a_k is a noisy copy of the echo at time k. Thus, apart from the additive (Gaussian) noise, a_k results from the transmitted digital data X_k by a stable filtering $\mathcal{F}'$ corresponding to the reflection properties of the channel. Hence, both a_k and X_k are bounded and thus the assumptions (A.2, 3) are satisfied. Again the memory M' in (A.1) depends on the memory of $\mathcal{F}'$ which can be assumed finite without loss of generality in practical applications.

IV. PROOF OF THE THEOREM

IV.1. Notations

The following notations are used in this work

$$v_k \triangleq h_k - h_* \tag{4}$$

$$z_k \triangleq a_k X_k - X_k X_k^T h_* \tag{5}$$

$$U_{k,t} \triangleq \prod_{i=k+1}^{t} (I - \mu_i X_i X_i^T) \triangleq \begin{cases} (I - \mu_t X_t X_t^T)(I - \mu_{t-1} X_{t-1} X_{t-1}^T) \dots \\ \qquad (I - \mu_{k+1} X_{k+1} X_{k+1}^T) \,; \quad t > k, \\ I \text{ (identity matrix)} \qquad t \leqslant k. \end{cases} \tag{6}$$

The norm of a matrix U is denoted by

$$\| U \| \triangleq \sup_{\{X : |X| = 1\}} | UX |. \tag{7}$$

In (4) v_k is the shift between h_k given by the algorithm (1) and the optimum vector h_*.

To prove the convergence of h_k to h_* we shall prove the convergence of v_k to the zero vector. Equations (2) and (5) imply that z_k is zero mean, i.e.,

$$E(z_k) = 0. \tag{8}$$

From (1), (4) and (5) it follows that

$$v_{k+1} = (I - \mu_k X_k X_k^T) v_k + \mu_k z_k. \tag{9}$$

Let us put (9) in the following form, used in the proof

$$v_{k+1} = \mathrm{W}_k + \mathrm{H}_k , \tag{10}$$

where

$$\mathrm{W}_k \triangleq \mathrm{U}_{0,k} v_1 , \tag{11}$$

$$\mathrm{H}_k \triangleq \sum_{j=1}^{k} \mu_j \mathrm{U}_{j,k} z_j . \tag{12}$$

IV.2. **Basic idea of the proof**

The idea of the proof can be figured out by the following heuristic arguments. To prove the algorithm convergence, it is sufficient to show that both terms W_k and H_k tend to zero. In that case $\mathrm{U}_{0,k}$ and all the terms $\mu_j \mathrm{U}_{j,k} z_j$ inside H_k will tend to zero. Except for the rather special case where the variable z_j is identically zero, z_j as given by (5) is a stationary non-zero random variable. Hence the product $\mu_j \mathrm{U}_{j,k}$ will tend to zero also. For j in the vicinity of k, the product matrix $\mathrm{U}_{j,k}$, (6), has very few factors and is not small. This requires that $\mu_j \to 0$ as $j \to \infty$, which is valid for the sequence (3). Now for small values of j, μ_j is not small; thus the convergence requires that $\mathrm{U}_{j,k} \ll \mathrm{I}$ or equivalently that

$$\| \mathrm{U}_{j,k} \| \to 0 \qquad \text{as} \qquad k \to \infty, \qquad j \ll k . \tag{13}$$

In this paper, an inequality similar to (13) is proved *on the average.* In fact, we show in step 3 of the proof, that for a given positive m

$$\mathrm{E}(\| \mathrm{U}_{j,k} \|^{m/3}) \leqslant \mathrm{C} . \left(\frac{j+1}{k+1} \right)^{2\beta} , \tag{14}$$

with fixed positive C and β; assume for simplicity that $k = j + n\mathrm{P}$; then due to the specific product structure of the matrix $\mathrm{U}_{j,k}$

$$\| \mathrm{U}_{j,k} \| \leqslant \prod_{l=0}^{n-1} \| \mathrm{U}_{j+l\mathrm{P},j+(l+1)\mathrm{P}} \| . \tag{15}$$

Now when j is large, μ_i is small for $i \geq j$, and it does not change significantly over the range $[j + l\mathrm{P}, j + (l + 1)\mathrm{P}]$. Thus the first order development

$$\mathrm{U}_{j+l\mathrm{P},j+(l+1)\mathrm{P}} = \mathrm{I} - \mu_{j+l\mathrm{P}+1} \sum_{i=j+l\mathrm{P}+1}^{j+(l+1)\mathrm{P}} \mathrm{X}_i \mathrm{X}_i^{\mathrm{T}} \tag{16}$$

is suitable, whence

$$\| \mathrm{U}_{j+l\mathrm{P},j+(l+1)\mathrm{P}} \| \leqslant 1 - \mu_{j+l\mathrm{P}+1}\lambda_{\min}\left\{ \sum_{i=j+l\mathrm{P}+1}^{j+(l+1)\mathrm{P}} \mathrm{X}_i\mathrm{X}_i^{\mathrm{T}} \right\}. \tag{17}$$

A basic step of the proof (Lemma 1) is to show the « mixing » property that $\exists \mathrm{P}, \delta_0 > 0$, such that

$$\mathrm{E}\left[\lambda_{\min}\left\{ \sum_{i=r+1}^{r+\mathrm{P}} \mathrm{X}_i\mathrm{X}_i^{\mathrm{T}} \right\}\right] \geqslant \delta_0 > 0, \qquad \forall r. \tag{18}$$

It is easily conceivable that taking the suitable expectation of (15) along the lines (17) (18), will yield the property (14) which expresses the decrease of $\mathrm{U}_{j,k}$ to zero.

The outline of the proof is thus

. to prove (18) for the finite memory model (A.1) – step 1;

. to deduce from (18) a bound for moments of certain order of $\| \mathrm{U}_{r,r+\mathrm{P}} \|$, – step 2;

. then to prove that moments of certain order of $\| \mathrm{U}_{j,k} \|$ decrease to zero when $k \to \infty$, $j \ll k$, – step 3;

. finally to use these results to prove the q. m. convergence of h_k – step 4; and lastly its a. s. convergence – step 5.

IV.3. **Proof of the theorem**

As just mentioned the proof proceeds in five steps. Some of the steps are stated as lemmas. We use the notation

$$\mathrm{Y}_r^{\mathrm{P}} = \sum_{i=r+1}^{r+\mathrm{P}} \mathrm{X}_i\mathrm{X}_i^{\mathrm{T}}. \tag{19}$$

Step 1 (LEMMA 1). — *Under the assumption* (A.1) *of finite memory, the condition* (18) *is satisfied, provided* $| \mathrm{X}_k |$ *has finite moments of order greater, then* 2N :

$$\exists \varepsilon > 0 \quad \text{such that} \quad \mathrm{E}(| \mathrm{X}_k |^{2\mathrm{N}+\varepsilon}) < \infty . \tag{20}$$

As we have shown, the ergodic property (18) plays a crucial role in the theorem. Therefore the proof of Lemma 1-is reported in Appendix I, although already given by the authors in a work [*11*] that deals with the constant step-size algorithm. Notice that $\mathrm{E}\{\lambda_{\min}(\mathrm{Y}_r^{\mathrm{P}})\}$ is a non-decreasing function of P. Thus one can assume that $\mathrm{P} \geqslant \mathrm{M}'$, which is useful hereafter.

Step 2. — This step derives a bound to $\mathrm{E}(\|\mathrm{U}_{r,r+\mathrm{P}}\|^m)$, where the exponent m is positive. It is based upon the following lemma.

LEMMA 2.*a.* — *If, for a given positive even integer m the sequence* X_k *satisfies the multivariate finite moments assumption* (B) :

$$\left.\begin{array}{l} \forall \mathrm{K}\,; \qquad \forall p_1, p_2, \ldots, p_{\mathrm{K}} \leqslant 2m\,; \\ \text{for all distinct indices } i_1, \ldots, i_{\mathrm{K}} \\ \mathrm{E}(|\mathrm{X}_{i_1}|^{p_1} \ldots |\mathrm{X}_{i_{\mathrm{K}}}|^{p_{\mathrm{K}}}) < \infty, \end{array}\right\} \quad (\mathrm{B})$$

then, $\forall \mathrm{P}$, *there exists a triple of positive numbers* r_m, α_m, F_m *such that*

$$\forall r \geqslant r_m, \ \mathrm{E}(\|\mathrm{U}_{r,r+\mathrm{P}}\|^m) \leqslant |1 - \alpha_m \mu_{r+1} \mathrm{E}(\lambda_{\min}(\mathrm{Y}_r^{\mathrm{P}}))| + \mathrm{F}_m \mu_{r+1}^2. \qquad (21)$$

The order $2m$ of the moments will be choosen hereafter according to technical requirements. This lemma is completed by Lemma 2.*b*:

LEMMA 2.*b.* — *If the sequence* X_k *has finite memory* M′ *according to* (A.1), *and finite univariante moments according to* (A.2)′:

$$\mathrm{E}(|\mathrm{X}_k|^{2m\mathrm{M}'}) < \infty\,, \qquad (\mathrm{A}.2)'$$

it satisfies assumption (B).

Lemma 2.*a* is proved in appendix II. It uses the decreasing nature of the sequence μ_i. To prove Lemma 2.*b*, in the product $|\mathrm{X}_{i_1}|^{p_1} \ldots |\mathrm{X}_{i_{\mathrm{K}}}|^{p_{\mathrm{K}}}$, we form M′ interlaced groups, each one containing only factors $|\mathrm{X}_i|^p$ whose indices i are exactly separated by multiples of M′. Then, using for $\mathrm{L} = \mathrm{M}'$, the classical inequality

$$|\mathrm{E}(\mathrm{Y}_1\mathrm{Y}_2 \ldots \mathrm{Y}_{\mathrm{L}})| \leqslant [\mathrm{E}(|\mathrm{Y}_1|^{\mathrm{L}}) \ldots \mathrm{E}(|\mathrm{Y}_{\mathrm{L}}|^{\mathrm{L}})]^{1/\mathrm{L}}, \qquad (22)$$

which is a direct consequence of the Hölder's inequality, one obtains an upper bound for

$$\mathrm{E}(|\mathrm{X}_{i_1}|^{p_1} \ldots |\mathrm{X}_{i_{\mathrm{K}}}|^{p_{\mathrm{K}}}) \qquad (23)$$

which contains M′ factors such as

$$[\mathrm{E}(|\mathrm{X}_{i_1}|^{p_1\mathrm{M}'} |\mathrm{X}_{i_1+\mathrm{M}'}|^{p_1'\mathrm{M}'} \ldots)]^{1/\mathrm{M}'}. \qquad (24)$$

Due to assumption (A.1), the quantities (24) can be factorized. Therefore (23) is finite, provided

$$\forall i \qquad \mathrm{E}(|\mathrm{X}_k|^{p_i\mathrm{M}'}) < \infty\,. \qquad (25)$$

In particular assumption (A.2′) implies that (B) holds.

As a consequence of Lemmas 2.a and 2.b, the moment bound (21) is fulfilled under assumptions (A.1) and (A.2)′. The combination of (21) with the result (18) of Lemma 1 provides the desirable moment bound namely

$$\exists \mathrm{P} \geqslant \mathrm{M}' \; ; \qquad \exists \alpha, r_m > 0 \qquad \text{such that} \qquad \forall r \geqslant r_m, \quad \mathrm{E}(\| \mathrm{U}_{r,r+\mathrm{P}} \|^m) \leqslant 1 - \alpha \mu_r \tag{26}$$

which was the purpose of step 2, subject to condition (20). Notice that (26) can be rewritten equivalently (due to (3))

$$\exists \mathrm{P} \geqslant \mathrm{M}' \; ; \qquad \exists \gamma, r_m > 0 \qquad \text{such that} \qquad \forall r \geqslant r_m, \quad \mathrm{E}(\| \mathrm{U}_{r,r+\mathrm{P}} \|^m) \leqslant 1 - \frac{\gamma}{r+1}. \tag{27}$$

Step 3. — This step consists of the proof of the bound

$$\mathrm{E}(\| \mathrm{U}_{j,t} \|^{m/3}) \leqslant \mathrm{C} . \left(\frac{j+1}{t+1} \right)^{2\beta} , \qquad 0 \leqslant j \leqslant t , \tag{14}$$

for a given pair of positive numbers (C, β). The result (14) is an expression of the fact that $\| \mathrm{U}_{j,t} \| \to 0$ when $t \to \infty$ with $j \ll t$, as was discussed in the previous intuitive section. In order to prove (14), let us organize the sequence of indices k within $[j, t]$ into three groups

$$\left. \begin{aligned} \Gamma_1 &= \{ j+1, \ldots, j+\mathrm{P} \} \cup \{ j+2P+1, \ldots, j+3\mathrm{P} \} \\ &\qquad \cup \ldots \cup \{ j+2(n-1)\mathrm{P}+1, \ldots, j+(2n-1)\mathrm{P} \} \\ \Gamma_2 &= \{ j+\mathrm{P}+1, \ldots, j+2\mathrm{P} \} \cup \{ j+3\mathrm{P}+1, \ldots, j+4\mathrm{P} \} \\ &\qquad \cup \ldots \cup \{ j+(2n-1)\mathrm{P}+1, \ldots, j+2n\mathrm{P} \} \\ \Gamma_3 &= \{ j+2n\mathrm{P}+1, \ldots, t \} , \end{aligned} \right\} \tag{28}$$

where n is the integer part of $(t-j)/2\mathrm{P}$. The groups Γ_1 and Γ_2 are interlaced, with an interval of $\mathrm{P} \geq \mathrm{M}'$ indices. According to the multiplicative property of the product matrix $\mathrm{U}_{j,t}$, one gets

$$\begin{aligned} \mathrm{E}(\| \mathrm{U}_{j,t} \|^{m/3}) \leqslant \mathrm{E} \Bigg\{ & \prod_{i=0}^{n-1} \| \mathrm{U}_{j+2i\mathrm{P},j+(2i+1)\mathrm{P}} \|^{m/3} \times \\ & \times \prod_{i=0}^{n-1} \| \mathrm{U}_{j+(2i+1)\mathrm{P},j+(2i+2)\mathrm{P}} \|^{m/3} \times \| \mathrm{U}_{j+2n\mathrm{P},t} \|^{m/3} \Bigg\} . \end{aligned} \tag{29}$$

Using inequality (22) with L = 3, it comes

$$\mathrm{E}\left\{ \| \mathrm{U}_{j,t} \|^{m/3} \right\} \leqslant \left\{ \mathrm{E}\left(\prod_{i=0}^{n-1} \| \mathrm{U}_{j+2i\mathrm{P},j+(2i+1)\mathrm{P}} \|^{m} \right) \right\}^{1/3} \times$$
$$\times \left\{ \mathrm{E}\left(\prod_{i=0}^{n-1} \| \mathrm{U}_{j+(2i+1)\mathrm{P},j+(2i+2)\mathrm{P}} \|^{m} \right) \right\}^{1/3} \times \left\{ \mathrm{E}(\| \mathrm{U}_{j+2n\mathrm{P},t} \|^{m}) \right\}^{1/3} . \quad (30)$$

Thanks to the independence (A.1) and to the structure (28) of the subsets Γ_1 and Γ_2, each of the first two moments in (30) can be split into n factors of the type $\mathrm{E}(\| \mathrm{U}_{r,r+\mathrm{P}} \|^{m})$ studied in Lemma 2.*a*.

Consider first the case $j \geq r_m$ for which (27) is valid. A bound to the third factor in the RHS of (30) involves a finite number of multivariate moments of order less or equal to $2m$, with respect to the variables $| \mathrm{X}_i |$. Due to Lemma 2.*b* all these moments are finite. Moreover the step-sizes μ_k that appears in $\mathrm{U}_{j+2n\mathrm{P},t}$ are bounded by μ_{r_m}. Hence this third factor is uniformly bounded. Therefore inequality (30) implies

$$\mathrm{E}(\| \mathrm{U}_{j,t} \|^{m/3}) \leqslant \mathrm{C}' \prod_{i=0}^{n-1} \left\{ \left(1 - \frac{\gamma}{j+2i\mathrm{P}+1}\right)\left(1 - \frac{\gamma}{j+(2i+1)\mathrm{P}+1}\right) \right\}^{1/3} . \quad (31)$$

Using the fact that $1 - u \leq e^{-u}$ for all real u, the inequality (14) follows from (31) after straightforward calculations.

For the case $j < r_m \leq t$, we use the decomposition $\mathrm{U}_{j,t} = \mathrm{U}_{r_m,t} \circ \mathrm{U}_{j,r_m}$ and apply a similar analysis to the factor $\mathrm{U}_{r_m,t}$ which corresponds to an increasing number of indices. The remaining factors correspond to a finite number of indices and can be dealt with like $\mathrm{U}_{j+2n\mathrm{P},t}$ in the previous case. Finally, and for the same reason, in the case $t \leq r_m$, the moments

$$\mathrm{E}(\| \mathrm{U}_{j,t} \|^{m/3})$$

are bounded. This achieves the proof of (14).

Step 4. — In this step we prove the q. m. convergence of h_t to the optimum vector h_*; on the basis of (10), it follows from the bounds

$$\mathrm{E}(| \mathrm{W}_t |^2) \leqslant \mathrm{C}_2 t^{-\beta} \ ; \quad 0 < \mathrm{C}_2 , \quad (32)$$
$$\mathrm{E}(| \mathrm{H}_t |^2) \leqslant \mathrm{C}_3 t^{-\beta/2} \ ; \quad 0 < \mathrm{C}_3 . \quad (33)$$

Inequality (32) follows directly from (11) and the a. s. boundedness of the initial vector v_1, provided $m/3 \geqslant 2$.

The detailed proof of (33) is given in appendix III. It relies upon the fact that

$$\mathrm{E}(\,|\,\mathrm{H}_t\,|^2) \leqslant \sum_{j,k=1}^{t} \mu_j \mu_k \,|\, \mathrm{E}(z_j^{\mathrm{T}} \mathrm{U}_{j,t}^{\mathrm{T}} \mathrm{U}_{k,t} z_k)\,| \,. \tag{34}$$

The inequality (22) with L = 4 is applied to individual terms in the sum of the RHS of (34), and then (14) is used with $m/3 = 4$. When t is large, most of the terms satisfy $|\,k - j\,| \geqslant \mathrm{M}'$ and due to the independence (A.1) and to the zero-mean property of z_k, they get very small. At this point, the reason why the order 24 M of the bounded moments for $|\,\mathrm{X}_k\,|$ is at least 24 M′ in the theorem, becomes apparent : q. m. convergence can be proved with $m = 12$. Thus assumption (A.2)′ of Lemma 2.*b* implies $\mathrm{M} \geq \mathrm{M}'$ in assumption (A.2) of the theorem.

Step 5. — In this step, the a. s. convergence of h_t is proved. It consists of the proofs of the a. s. convergence of (*i*) W_t to zero and (*ii*) H_t to zero. For that let us assume, which is possible, that β is less than one.

i) a. s. convergence of W_t *to* 0

Let q be a finite integer greater than $\frac{1}{\beta}$ and let t_n, $n = 1, 2, \ldots$ be a subsequence such that $t_n = n^q$; then (32) implies that

$$\sum_{n=1}^{\infty} \mathrm{E}(\,|\,\mathrm{W}_{t_n}\,|^2) < \infty \,. \tag{35}$$

The equation (35) is sufficient for

$$\mathrm{W}_{t_n} \overset{\text{a.s.}}{\rightarrow} 0, \qquad \text{as} \quad n \rightarrow \infty \,. \tag{36}$$

Due to (3), $\mu_j\,|\,\mathrm{X}_j\,|^2 \rightarrow 0$, as $j \rightarrow \infty$, which implies for large enough indices j

$$\|\,\mathrm{U}_{j,t}\,\| \leqslant 1 \qquad \text{a. s.} \tag{37}$$

Thanks to definition (11) and to (36), (37)

$$\sup_{t_n \leqslant t \leqslant t_{n+1}} |\,\mathrm{W}_t\,| \leqslant |\,\mathrm{W}_{t_n}\,| \,.\, \sup_{t_n \leqslant t \leqslant t_{n+1}} \|\,\mathrm{U}_{t_n,t}\,\| \overset{\text{a.s.}}{\rightarrow} 0, \qquad n \rightarrow \infty \,. \tag{38}$$

It follows from (38) that

$$\mathrm{W}_t \overset{\text{a.s.}}{\rightarrow} 0, \qquad t \rightarrow \infty \,. \tag{39}$$

ii) a. s. convergence of H_t *to* 0.

Let q_1 be a finite integer greater than $\frac{2}{\beta}$ and let t_n, $n = 1, 2, \ldots$ be a subsequence such that

$$t_n = n^{q_1} ; \tag{40}$$

then (33) implies that

$$\sum_{n=1}^{\infty} \mathrm{E}(\,|\,\mathrm{H}_{t_n}|^2) < \infty ,$$

from which

$$\mathrm{H}_{t_n} \overset{\text{a.s.}}{\rightarrow} 0, \qquad n \rightarrow \infty . \tag{42}$$

From (12) one has

$$\mathrm{H}_t = \mathrm{U}_{t_n,t}\mathrm{H}_{t_n} + \sum_{j=t_n+1}^{t} \mu_j \mathrm{U}_{j,t} z_j . \tag{43}$$

According to (3),

$$|\,\mathrm{H}_t\,| \leqslant \|\,\mathrm{U}_{t_n,t}\,\| \,.\, |\,\mathrm{H}_{t_n}| + d_1 \sum_{j=t_n+1}^{t_{n+1}} \frac{1}{j} \|\,\mathrm{U}_{j,t}\,\|\; |\,z_j\,| . \tag{44}$$

The inequalities (37) and (44) imply

$$\lim_{n\rightarrow\infty} \sup_{t_n \leqslant t \leqslant t_{n+1}} |\,\mathrm{H}_t\,| \leqslant \limsup_{n\rightarrow\infty} . \left(|\,\mathrm{H}_{t_n}| + d_1 \sum_{j=t_n+1}^{t_{n+1}} \frac{1}{j} |\,z_j\,| \right) \text{a. s.} \tag{45}$$

The ergodicity of $|\,z_j\,|$ which results from the assumption (A.1), implies [*14*] that

$$\lim_{n\rightarrow\infty} \sum_{j=t_n+1}^{t_{n+1}} \frac{1}{j} |\,z_j\,| \overset{\text{a.s.}}{=} \lim_{n\rightarrow\infty} . \sum_{j=t_n+1}^{t_{n+1}} \frac{1}{j} \mathrm{E}(\,|\,z_j\,|\,) = 0 , \tag{46}$$

where the last equation in (46) results from (40) and the boundedness (A.2, 3) of $\mathrm{E}(\,|\,z_j\,|\,)$. The equations (42) (45) and (46) imply that

$$\mathrm{H}_t \overset{\text{a.s.}}{\rightarrow} 0, \qquad \text{as} \quad t \rightarrow \infty \tag{47}$$

and the convergence of h_t to h_* results from (39) and (47).

End of the proof.

In the above proof, several steps make use of the finite memory model (A.1). For instance in step 3, the moment

$$\mathrm{E}\left(\prod_{i=0}^{n-1} \|\,\mathrm{U}_{j+2i\mathrm{P},j+(2i+1)\mathrm{P}}\,\|^m \right)$$

has been factorized. Although not proved here, it is presumable that the same steps can be gained for more general models of ergodicity, e. g. for the decaying covariance models.

V. CONCLUSION

The convergence analysis of the stochastic gradient algorithm with decreasing step-size presented in this paper concernes both the q. m. and a. s. convergence of the algorithm when it governs an adaptive linear estimator. Unrealistic assumptions such as independence of successive observations and boundedness of the algorithm are avoided in this paper and replaced by the two plausible assumptions of bounded moments and finite memory for the observations. One of the contributions in this paper is the q. m. convergence analysis of algorithms without barrier and with correlated observations which is completely new in the literature. The other contribution is the simplicity of the proof of the a. s. convergence which does not appeal to elaborate technical arguments.

APPENDIX I

(*Proof of Lemma 1*)

The purpose of this lemma is to establish that

$$\exists \mathrm{P}, \delta_0 > 0 \quad \text{such that} \quad \forall r, \mathrm{E}\left\{ \lambda_{\min}\left(\sum_{i=r+1}^{r+\mathrm{P}} \mathrm{X}_i \mathrm{X}_i^{\mathrm{T}} \right) \right\} \geqslant \delta_0 \tag{I.1}$$

under the assumption (A.1) of finite memory and some suitable assumption on the finiteness of the $| \mathrm{X}_k |$-moments which will turn out to be

$$\exists \varepsilon > 0 \qquad \text{such that} \quad \mathrm{E}(| \mathrm{X}_k |^{2\mathrm{N}+\varepsilon}) < \infty . \tag{I.2}$$

Using the notation (19), one has

$$\lambda_{\min}(\mathrm{Y}_r^{\mathrm{P}'}) \geqslant \lambda_{\min}(\mathrm{Y}_r^{\mathrm{P}}) \geqslant 0 \,; \qquad \forall \mathrm{P}' > \mathrm{P} , \tag{I.3}$$

and due to stationarity, the average is solely a function of P, which is non-decreasing starting from zero. Thus we are searching for an integer P at which this function departs from zero. To prove the lemma, we shall show that $\mathrm{P} = \mathrm{NM}'$ is suitable, i. e., that

$$\mathrm{E}\,[\lambda_{\min}(\mathrm{Y}_r^{\mathrm{NM}'})] \geqslant \delta_1 > 0 . \tag{I.4}$$

Now due to definition (19)

$$\mathrm{E}\,[\lambda_{\min}(\mathrm{Y}_r^{\mathrm{NM}'})] \geqslant \mathrm{E}\left[\lambda_{\min}\left(\sum_{i=0}^{\mathrm{N}-1} \mathrm{X}_{r+1+i\mathrm{M}'} \mathrm{X}_{r+1+i\mathrm{M}'}^{\mathrm{T}} \right) \right]. \tag{I.5}$$

Consider the determinant of the matrix $\left(\sum_{i=0}^{\mathrm{N}-1} \mathrm{X}_{r+1+i\mathrm{M}'} \mathrm{X}_{r+1+i\mathrm{M}'}^{\mathrm{T}} \right)$ and denote by x_i^n the n^{th} component of the vector X_i. Due to the multi-linearity of the determinant with respect to each column one gets

$$\det \left(\sum_{i=0}^{\mathrm{N}-1} \mathrm{X}_{r+1+i\mathrm{M}'} \mathrm{X}_{r+1+i\mathrm{M}'}^{\mathrm{T}} \right) = \sum_{i_1=0}^{\mathrm{N}-1} \cdots \sum_{i_{\mathrm{N}}=0}^{\mathrm{N}-1} x_{r+1+i_1\mathrm{M}'}^1 \cdots x_{r+1+i_{\mathrm{N}}\mathrm{M}'}^{\mathrm{N}} \times d(\mathrm{X}_{r+1+i_1\mathrm{M}'}, \ldots, \mathrm{X}_{r+1+i_{\mathrm{N}}\mathrm{M}'}) , \tag{I.6}$$

where $d(\mathrm{U}_1, \ldots, \mathrm{U}_{\mathrm{N}})$ denotes the determinant of the matrix U with columns $\mathrm{U}_1, \ldots, \mathrm{U}_{\mathrm{N}}$. If the indices $i_1, \ldots, i_{\mathrm{N}}$ are not all distinct, the latter determinant is zero. Consequently one has

$$\mathrm{E}\left\{ \det \left[\sum_{i=0}^{\mathrm{N}-1} \mathrm{X}_{r+1+i\mathrm{M}'} \mathrm{X}_{r+1+i\mathrm{M}'}^{\mathrm{T}} \right] \right\} = \sum_{(i_1,\ldots,i_{\mathrm{N}}) \in \mathscr{P}} \mathrm{E}\,\{ d(x_{r+1+i_1\mathrm{M}'}^1 \mathrm{X}_{r+1+i_1\mathrm{M}'}, \ldots, x_{r+1+i_{\mathrm{N}}\mathrm{M}'}^{\mathrm{N}} \mathrm{X}_{r+1+i_{\mathrm{N}}\mathrm{M}'}) \} , \tag{I.7}$$

where $\mathscr{P}$ is the set of all permutations of $[0, 1, 2, \ldots, \mathrm{N}-1]$. Using the assumption (A.1) we can show that each term on the R. H. S. of (I.7) is det (R). Therefore

$$\mathrm{E}\left(\det\left\{\sum_{i=0}^{\mathrm{N}-1} \mathrm{X}_{r+1+i\mathrm{M}'}\mathrm{X}_{r+1+i\mathrm{M}'}^{\mathrm{T}}\right\}\right) = \mathrm{N}\,!\,\det(\mathrm{R}) > 0\,. \tag{I.8}$$

Denoting by $\lambda_1, \lambda_2, \ldots, \lambda_{\mathrm{N}}$ the eigenvalues of $\sum_{i=0}^{\mathrm{N}-1} \mathrm{X}_{r+1+i\mathrm{M}'}\mathrm{X}_{r+1+i\mathrm{M}'}^{\mathrm{T}}$, arranged in increasing order, we get from (I.8)

$$\mathrm{E}(\lambda_1,\, \lambda_2 \ldots \lambda_{\mathrm{N}}) = \mathrm{N}\,!\,\det(\mathrm{R}) > 0\,. \tag{I.9}$$

Obviously one has, $\forall \alpha \in\,]0, 1[$

$$0 < \mathrm{E}(\lambda_1\lambda_2 \ldots \lambda_{\mathrm{N}}) \leqslant \mathrm{E}(\lambda_1^{\alpha}\lambda_{\mathrm{N}}^{\mathrm{N}-\alpha})\,. \tag{I.10}$$

It follows from the Holder's inequality that

$$\mathrm{E}(\lambda_1^{\alpha}\lambda_{\mathrm{N}}^{\mathrm{N}-\alpha}) \leqslant [\mathrm{E}(\lambda_1)]^{\alpha}\left[\mathrm{E}\left(\lambda_{\mathrm{N}}^{\frac{\mathrm{N}-\alpha}{1-\alpha}}\right)\right]^{1-\alpha}. \tag{I.11}$$

Suppose in addition the validity of condition (I.2)-condition (20) in section IV.3. Then the obvious bound

$$\lambda_{\mathrm{N}} \leqslant \sum_{i=0}^{\mathrm{N}-1} |\,\mathrm{X}_{r+1+i\mathrm{M}'}\,|^2 \tag{I.12}$$

together with assumptions (A.1), will result into the inequality

$$\mathrm{E}\left(\lambda_{\mathrm{N}}^{\frac{\mathrm{N}-\alpha}{1-\alpha}}\right) < \infty\,. \tag{I.13}$$

Combining the inequalities (I.10), (I.11), (I.13) we obtain $\mathrm{E}(\lambda_1) > 0$, i. e.,

$$\mathrm{E}\left[\lambda_{\min}\left(\sum_{i=0}^{\mathrm{N}-1} \mathrm{X}_{r+1+i\mathrm{M}'}\mathrm{X}_{r+1+i\mathrm{M}'}^{\mathrm{T}}\right)\right] > 0\,. \tag{I.14}$$

Together with (I.5) this achieves the lemma proof.

APPENDIX II

(*Proof of Lemma* 2.*a*)

Remember that the sequence μ_i is decreasing and consider the development of $\| U_{r,r+P} \|$ according to

$$\| U_{r,r+P} \| \leqslant \| I - \mu_{r+1} Y_r^P \| + \sum_{i=r+1}^{r+P} (\mu_{r+1} - \mu_i) | X_i |^2 + \mu_{r+1}^2 \{ \text{terms in } \mu_{i_1} \ldots \mu_{i_{p-2}} | X_{i_1} |^2 \ldots | X_{i_p} |^2 \} . \quad \text{(II.0)}$$

Notice that $\mu_{r+1} - \mu_i \leqslant \mu_{r+1}^2 \dfrac{P-1}{d_1}$ when $r + 1 \leqslant i \leqslant r + P$.
Thus (II.0) can be written

$$\| U_{r,r+P} \| \leqslant \| I - \mu_{r+1} Y_r^P \| + \mu_{r+1}^2 \{ \text{terms in } \mu_{i_1} \ldots \mu_{i_{p-2}} | X_{i_1} |^2 \ldots | X_{i_p} |^2 \} \quad \text{(II.1)}$$

where $1 \leq p \leq P$ and where $i_1, \ldots, i_p$ are distinct integers, with the agreement that the product $\mu_{i_1} \ldots \mu_{i_{p-2}}$ is 1 for $p \leq 2$. Putting (II.1) to the power m and averaging, one gets

$$E(\| U_{r,r+P} \|^m) \leqslant E(\| I - \mu_{r+1} Y_r^P \|^m) + \mu_{r+1}^2$$
$$E \{ \text{terms in } \mu_{r+1}^{2k-2} \mu_{i_1}^k \ldots \mu_{i_{p-2}}^k | X_{i_1} |^{2k} \ldots | X_{i_p} |^{2k} \| I - \mu_{r+1} Y_r^P \|^{m-k} \} , \quad \text{(II.2)}$$

where $1 \leqslant k \leqslant m$. Since

$$\| I - \mu_i X_i X_i^T \| \leqslant 1 + \mu_i | X_i |^2 , \quad \text{(II.3)}$$

the bracketted average in the second term of (II.2) is bounded by a polynomial of the type

$$\sum \mu_{r+1}^l E(| X_{i_1} |^{2k} \ldots | X_{i_p} |^{2k} | X_i |^{2n}) \quad \text{(II.4)}$$

where $0 \leqslant n \leqslant m - k$. Hence each coefficient of this polynomial is a moment of the type

$$E(| X_{i_1} |^{p_1} \ldots | X_{i_K} |^{p_K}) \, ; \qquad p_1, \ldots, p_K \leqslant 2m . \quad \text{(II.5)}$$

According to assumption (B), all these moments are finite. Since μ_{r+1} is bounded, the polynomial (II.4) itself is bounded. Hence, for some fixed positive constant D_1

$$E(\| U_{r,r+P} \|^m) \leqslant E(\| I - \mu_{r+1} Y_r^P \|^m) + D_1 \mu_{r+1}^2 . \quad \text{(II.6)}$$

Let ω and $P(d\omega)$ denote respectively an arbitrary point in the space Ω of random events and the probability measure on that space. Then

$$E(\| I - \mu_{r+1} Y_r^P \|^m) = \int_\Omega \| I - \mu_{r+1} Y_r^P \|^m P(d\omega) \leqslant$$
$$\int_{\{\omega : \mu_{r-1} \| Y_r^P \| < 1\}} [1 - \mu_{r+1} \lambda_{\min}(Y_r^P)]^m P(d\omega) + \int_{\{\omega : \mu_{r+1} \| Y_r^P \| \geqslant 1\}} (2\mu_{r+1} \| Y_r^P \|)^m P(d\omega) . \quad \text{(II.7)}$$

Because m is even, one has

$$\begin{aligned}\int_{\{\omega:\mu_{r+1}\|Y_r\|<1\}} [1-\mu_{r+1}\lambda_{\min}(Y_r^P)]^m P(d\omega) &\leqslant \int_{\{\omega:\mu_{r+1}\|Y_r\|<1\}} [1-\mu_{r+1}\lambda_{\min}(Y_r^P)]^2 P(d\omega)\,, \\ &\leqslant E[(1-\mu_{r+1}\lambda_{\min}(Y_r^P))^2]\,. \end{aligned} \tag{II.8}$$

Using assumption (B) again for the moments of Y_r^P, as was done for (II.3), one gets

$$E(\| Y_r^P \|^Q) \leqslant D_2\,, \qquad \forall Q \leqslant m\,, \tag{II.9}$$

for some positive constant D_2. It follows from (II.7), (II.8) that

$$E(\| I-\mu_{r+1}Y_r^P \|^m) \leqslant | 1-2\mu_{r+1}E(\lambda_{\min}(Y_r^P)) | + \mu_{r+1}^2[E(\| Y_r^P \|^2) + 2^m E(\| Y_r^P \|^m)]\,. \tag{II.10}$$

The combination of (II.6) and (II.10) brings the result (21) of Lemma 2.*a*.

APPENDIX III

(*A bound to* $E(|H_t|^2)$)

From (12) we have

$$E(|H_t|^2) = \sum_{j=1}^{t} \mu_j^2 E(|U_{j,t} z_j|^2) + 2 \sum_{j=1}^{t-1} \sum_{k=j+1}^{t} E(z_j^T U_{j,t}^T U_{k,t} z_k). \tag{III.0}$$

Using the fact that $|z_j| \leqslant |a_j|\,|X_j| + |h_*|\,|X_j|^2$ and the assumption (A.2), we see that a sufficient condition to achieve the inequality

$$E(|z_j|^4) \leqslant B < \infty, \tag{III.1}$$

is that

$$E[(|X_j|\,|a_j|)^4] < \infty. \tag{III.2}$$

In turn, according to the Hölder's inequality and to (A.2), (III.2) and thus (III.1) will hold thanks to assumption (A.3) on the moments of $|a_j|$.

Then, applying the Schwarz inequality to individual terms in the first sum in the R. H. S. of (III.0) and using (14) and the inequality $\beta < 1$, one obtains

$$\sum_{j=1}^{t} \mu_j^2 E(|U_{j,t} z_j|^2) \leqslant G t^{-\beta}, \tag{III.3}$$

for some finite positive constant G.

Now, consider the second term in the R. H. S. of (III.0). We have

$$\sum_{j=1}^{t-1} \sum_{k=j+1}^{t} \mu_j \mu_k E(z_j^T U_{j,t}^T U_{k,t} z_k) = \sum_{(j,k)\in S_1} \mu_j \mu_k E(z_j^T U_{j,k}^T U_{k,t} z_k) + \sum_{(j,k)\in S_2} \mu_j \mu_k E(z_j^T U_{j,t}^T U_{k,t} z_k), \tag{III.4}$$

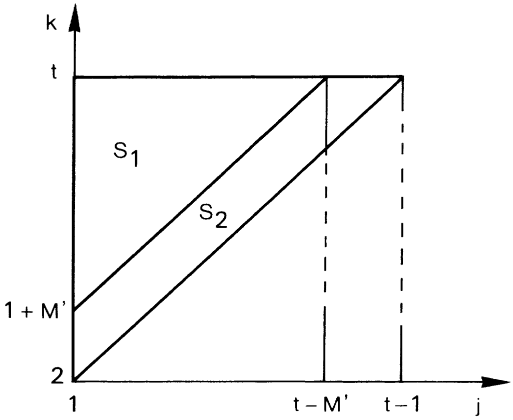

FIG. 1. — Illustration of the definitions in (III.5).

where S_1 and S_2 are defined by

$$S_1 = \{ (j, k) : 1 \leqslant j \leqslant t - M' ; \; j + M' \leqslant k \leqslant t \} \tag{III.5.a}$$

$$S_2 = \{ (j, k) : 1 \leqslant j \leqslant t - 1 ; \; j + 1 \leqslant k \leqslant \min (j + M' - 1, t) \} . \tag{III.5.b}$$

The definitions of S_1 and S_2 are illustrated by figure 1.

A bound to the first term in the R. H. S. of (III.4).

We have

$$z_j^T U_{j,t}^T U_{k,t} z_k = z_j^T U_{j,j+M'-1}^T U_{j+M'-1,t}^T U_{k,t} z_k . \tag{III.6}$$

Let P_j be defined by

$$U_{j,j+M'-1}^T \triangleq I + P_j . \tag{III.7}$$

Because in S_1, $k \geq j + M'$, then assumption (A.1) and the fact that z_j in (5) is zero-mean, both imply that

$$E(z_j^T U_{j,t}^T U_{k,t} z_k) = E(z_j^T P_j U_{j+M'-1,t}^T U_{k,t} z_k) . \tag{III.8}$$

According to Lemma 2.*b* and to (3), there exists $G_2 > 0$ such that

$$E(\| P_j \|^4) \leqslant G_2^4 \mu_{j+1}^4 . \tag{III.9}$$

Using (III.8) and (22) with $L = 4$ one obtains

$$| E(z_j^T U_{j,t}^T U_{k,t} z_k) | \leqslant \{ E(| z_j |^4 | z_k |^4) E(\| P_j \|^4) E(\| U_{j+M'-1,t} \|^4) E(\| U_{k,t} \|^4) \}^{1/4} . \tag{III.10}$$

Since z_j and z_k are independent then it follows from (III.1, 9, 10) that

$$| E(z_j^T U_{j,t}^T U_{k,t} z_k) | \leqslant \mu_{j+1} \sqrt{B} G_2 \{ E(\| U_{j+M'-1,t} \|^4) E(\| U_{k,t} \|^4) \}^{1/4} . \tag{III.11}$$

Using (14), (III.11) and (3) one obtains

$$\left| \sum_{(j,k) \in S_1} \mu_j \mu_k E(z_j^T U_{j,t}^T U_{k,t} z_k) \right| \leqslant d_1^3 G_2 \sqrt{BC} \sum_{j=1}^{t} \sum_{k=1}^{t} j^{-2} k^{-1} \left(\frac{j + M' - 1}{t+1} \right)^{\beta/2} \left(\frac{k}{t+1} \right)^{\beta/2} . \tag{III.12}$$

From (III.12) there exists a positive constant G_3 such that

$$\left| \sum_{(j,k) \in S_1} \mu_j \mu_k E(z_j^T U_{j,t}^T U_{k,t} z_k) \right| \leqslant G_3 t^{-\beta/2} . \tag{III.13}$$

A bound to the second term in the R. H. S. of (III.4)

Using the inequality (22) with $L = 4$, then (14) and (III.1), one gets

$$\left| \sum_{(j,k) \in S_2} \mu_j \mu_k E(z_j^T U_{j,k}^T U_{k,t} z_k) \right| \leqslant$$

$$d_1^2 \sqrt{BC} \sum_{j=1}^{t-1} \sum_{k=j+1}^{\min(j+M'-1,t)} j^{-1} k^{-1} \left(\frac{j}{t} \right)^{\beta/2} \left(\frac{k}{t} \right)^{\beta/2} \leqslant d_1^2 \sqrt{BC} t^{-\beta} \sum_{j=1}^{t-1} M' j^{-2+\beta} . \tag{III.14}$$

Thus, there exists a positive constant G_4 such that

$$\left| \sum_{(j,k)\in S_2} \mu_j \mu_k E(z_j^T U_{j,t}^T U_{k,t} z_k) \right| \leqslant G_4 t^{-\beta} . \tag{III.15}$$

From (III.4, 13, 15) we have

$$\left| \sum_{j=1}^{t-1} \sum_{k=j+1}^{t} \mu_j \mu_k E(z_j^T U_{j,t}^T U_{k,t} z_k) \right| \leqslant G_5 t^{-\beta/2} , \tag{III.16}$$

for some positive constant G_5. The result (33) follows immediately from (III.0, 3, 16).

REFERENCES

[*1*] H. Robbins, S. Monro, « A Stochastic Approximation Method », *The Annals of Mathematical Statistics*, t. **22**, 1951, p. 400-407.

[*2*] L. Schmetterer, « Stochastic Approximation », *Proc. of the Fourth Berkeley Symp. on Math. Stat. and Proba.*, p. 587-609.

[*3*] D. Sakrison, « Stochastic Approximation, a Recursive Method for Solving Regression Problems », *Adv. in Comm. Systems*, A. V. Balakrishnan Éditeur, Acad. Press, 1966, p. 51-106.

[*4*] C. Macchi, *Itération stochastique et Traitements numériques adaptatifs*, Thèse d'État, Paris, 1972.

[*5*] B. Widrow, J. M. McCool, M. G. Larimore, C. R. Johnson, « Stationary learning characteristics of the LMS adaptive filter » *Proc. IEEE*, t. **64**, n° 8, 1976, p. 1151-1162.

[*6*] O. Macchi, « Résolution adaptative de l'équation de Wiener-Hopf. Cas d'un canal de données affecté de gigue », *Ann. Inst. Henri Poincaré*, t. **14**, n° 3, 1978, p. 355-377.

[*7*] H. J. Kushner, D. S. Clark, « Stochastic approximation methods for constrained and unconstrained systems », *Appl. Math. Sci. Series*, n° 26, Springer-Verlag, Berlin, 1978, ch. 1.

[*8*] L. Ljung, « Analysis of a recursive stochastic algorithm », *IEEE Trans. on Automatic Control*, t. AC **22**, 1977, p. 551-575.

[*9*] D. C. Farden, « Stochastic Approximation with Correlated Data », *IEEE Trans. on Information Theory*, t. IT **27**, n° 1, 1981, p. 105-113.

[*10*] E. Eweda, O. Macchi, « Convergence of an adaptive linear estimation algorithm », *IEEE Trans. on Automatic Control*, t. AC **28**, n° 10, octobre 1983.

[*11*] O. Macchi, E. Eweda, « Second order convergence analysis of stochastic adaptive linear filtering ». *IEEE Trans. on Automatic Control*, t. **28**, n° 1, janvier 1983, p. 76-85.

[*12*] E. Eweda, *Egalisation adaptative d'un canal filtrant non stationnaire*. Thèse Ingénieur-Docteur, Orsay, 19 mars 1980.

[*13*] E. Eweda, O. Macchi, « Poursuite adaptative du filtrage optimal non stationnaire ». *C. R. Acad. Sci.*, t. **293**, série I, 1981, p. 497-500.

[*14*] H. Cramer and M. R. Leadbetter, *Stationary and Related Stochastic Processes*, New York: Wiley, 1967.

(Manuscrit reçu le 24 février 1982)