11 (1991), 3, 327—340

Acta Mathematica Scientia
数学物理学报

LMS-LIKE ESTIMATION FOR TIME VARYING PARAMETERS*

H. F. Chen (陈翰馥) *L. Guo* (郭　雷) *J. F. Zhang* (张纪峰)

Inst. of Syst. Sci., Academia Sinica, Beijing, China

Dedicated to the 70th Birthday of Professor Wu Wen-tsün

Abstract

An LMS-like algorithm is applied for estimating the time-varying parameter θ_n in the linear model $y_n=\varphi_n^T\theta_n+v_n$, which is general in the sense that none of the probabilistic properties such as stationarity, Markov property, independence and ergodicity is imposed on any of the processes $\{y_n\}$, $\{\varphi_n\}$, $\{\theta_n\}$ and $\{v_n\}$. It is shown that the α-th moment of the estimation error is of order of the α-th moment of the observation noise and the parameter variation $w_n\triangleq\theta_n-\theta_{n-1}$.

§ 1. Introduction

For linear stochastic systems with constant parameters there has been made a great progress on the parameter estimation problem. Systems with constant parameters may be viewed as a first approximation to the real processes which, as a matter of fact, mostly are time-varying in practice, and such an approximation is not alwayi satisfied by the practitioners. By this reason for recent years a considerable attention has been paid to analysing systems with unknown time-varying parameters by researchers in the areas such as control theory, signal processing and time series analysis.

It is natural to expect that the first set of results on parameter estimation (or tracking) and adaptive control for systems with time-varying parameters is obtained under some statistical law assumptions on the regressor or on the observation noise or on the parameter itself. For example, it is assumed that the regressor is stationary and independent of the observation noise (Macchi, 1986), the observation noise is Gaussian (Kitagava and Gersh, 1985) and the parameter is a Markov process (Chen and Caines, 1990; Guo and Meyn, 1989; Ji and Chizeck, 1988). Without any doubt the assumptions made in the papers mentioned above are reasonable in certain circumstances, but they are restrictive in general. For example, the stationarity assumption on the regressor excludes the feedback control system from consideration, for which the regressor cannot be independent of the noise either.

In this paper we consider the following system with unknown time-varying parameter $\{\theta_k\}$:

* Received Jan. 22, 1991. Work supported by NSFC and TWAS RG MP 898-117.

$$y_k=\varphi_k^\tau\theta_k+v_k, \tag{1.1}$$

where y_k is the one-dimensional system output, φ_k is the r-dimensional regressor and v_k is the system noise.

Clearly, in the special case where

$$\varphi_k^\tau=[y_{k-1}\cdots y_{k-s}\quad u_{k-1}\cdots u_{k-t}]$$

and v_k is a moving average process System (1.1) turns to be the ordinary ARMAX model with time-varying parameters.

The problem stated in this paper is to on-line estimate or to track the time-varying parameter θ_k based on the observed data y_i and φ_i, $i\leqslant k$.

We would like to emphasize that System (1.1) is quite general in the class of linear models: For processes $\{y_k\}$, $\{\varphi_k\}$, $\{\theta_k\}$ and $\{v_k\}$ we do not make any assumption on their statistical relationship and do not require them to be processes of a restricted class such as stationary process, Markov process etc.

We characterize the observation error $\{v_k\}$ and the parameter variation

$$w_n=\theta_n-\theta_{n-1} \tag{1.2}$$

either by

$$\sigma_\alpha\triangleq\sup_{n>0}E(|v_n|^\alpha+\|w_n\|^\alpha) \tag{1.3}$$

or by

$$s_\alpha\triangleq\lim_{n>0}\sup\frac{1}{n}\sum_{i=0}^{n}(|v_i|^\alpha+\|w_i\|^\alpha),\quad \text{a. s.} \tag{1.4}$$

for some constant $\alpha>0$.

In such a set-up of the problem, Guo (1990) provides a detailed analysis for estimation error $\tilde\theta_n\triangleq\theta_n-\hat\theta_n$, when $\hat\theta_n$ is calculated according to a Kalman-filter-like recursion for which a time-varying matrix adaptation gain is used. Guo's results are then strengthened (Zhang, Guo and Chen 1990). It is shown that α-th moment of the estimation error is of order α_α or s_α in accordance with the average taken in the sample space sense or in the time domain sense, respectively.

In this paper we establish similar results for the estimate $\hat\theta_n$ produced by an LMS-like algorithm, which is characterized by its simplicity for computation:

$$\hat\theta_{n+1}=\hat\theta_n+\mu_n\varphi_n(y_n-\varphi_n^\tau\hat\theta_n), \tag{1.5}$$

where $\{\mu_n\}$ is bounded by a positive constant with $\mu_n\|\varphi_n\|^2\leqslant1$, $\forall\ n\geqslant0$ and μ_n is measurable with respect to the σ-algebra generated by $\{v_i,\ \theta_i,\ \varphi_i,\ i\leqslant n\}$. It is easy to see that the algorithm (1.5) becomes the well-known LMS adaptive filter with step size μ considered by Widrow et al (1976), Macchi (1986), Bitmead and Anderson (1980), Benveniste et al (1987), if $\mu_k\equiv\mu\in(0,1)$, and it turns to be the projection (or gradient) algorithm (see, e.g. Anderson et al, 1986; Chen and Guo, 1987) with step size μ if $\mu_k=\frac{\mu}{1+\|\varphi_k\|^2}$.

In comparison with the previous work we note that the tracking error bounds are established when LMS is applied to tracking deterministic time-varying parameter. (Eweda and Macchi, 1985). However, they assumed that the time variation $\|\theta_n-\theta_{n-1}\|$ is bounded in k and $\{y_k,\ \varphi_k\}$ is an M-dependent sequence. For the constant parameter θ the convergence of LMS estimate is proved under the assumption that φ_k is purely nondeterminisitic and $\{\varphi_k, v_k\}$ is strictly stationary, while for the time-varying parameter an additional assumption is required that $\{\theta_n - \theta_{n-1}\}$ is stationary, zero mean and independent of $\{\varphi_k, v_k\}$ (Solo, 1990).

§ 2. Main Results

Recursively define

$$\Phi(n+1, m) = (I - \mu_n\varphi_n\varphi_n^\tau)\Phi(n, m), \ \Phi(m, m) = I, \ \forall n \geqslant m \geqslant 0. \tag{2.1}$$

From (1.1), (1.2), (1.5) and (2.1) it is easy to see that

$$\tilde{\theta}_{n+1} = (I - \mu_n\varphi_n\varphi_n^\tau)\tilde{\theta}_n + w_{n+1} - \mu_n\varphi_n v_n \tag{2.2}$$

$$= \Phi(n+1, 0)\tilde{\theta}_0 + \sum_{i=0}^{n}\Phi(n+1, i+1)\xi_{i+1} \tag{2.3}$$

where

$$\tilde{\theta}_n \triangleq \theta_n - \hat{\theta}_n \quad \text{and} \quad \xi_{n+1} \triangleq w_{n+1} - \mu_n\varphi_n v_n. \tag{2.4}$$

We always assume that

$$E\|\tilde{\theta}_0\|^\alpha < \infty.$$

From (2.3) we see that the tracking error $\{\tilde{\theta}_n\}$ strongly depends upon the behavior of $\{\Phi(n, m), \forall n \geqslant m \geqslant 0\}$. For analysing (2.3) the essential role is played by the "conditional richness" condition (Guo, 1990) and (Zhang, Guo and Chen, 1990):

There is a nondecreasing sequence $\{\mathscr{F}_n\}$ of σ-algebras such that $\mu_n \in \mathscr{F}_n$, $\varphi_n \in \mathscr{F}_n$ and

$$E\left\{\sum_{k=m+1}^{m+h}\mu_k\varphi_k\varphi_k^\tau \,\Big|\, \mathscr{F}_m\right\} \geqslant \frac{1}{\alpha_m} I \quad \text{a.s.,} \quad \forall m \geqslant 0, \tag{2.5}$$

where h is a positive integer and $\{\alpha_m, \mathscr{F}_m\}$ is an adapted nonnegative sequence satisfying $\alpha_m \geqslant 1$, and

$$\alpha_{m+1} \leqslant a\alpha_m + \eta_{m+1}, \quad \forall m \geqslant 0, \ E\alpha_0^{1+\delta} < \infty \tag{2.6}$$

where $a \in (0, 1)$ is a constant and $\{\eta_m, \mathscr{F}_m\}$ is an adapted nonnegative sequence such that

$$\sup_{m \geqslant 0} E[\eta_{m+1}^{1+\delta} \,|\, \mathscr{F}_m] \leqslant M \quad \text{a.s.} \tag{2.7}$$

with $\delta > 0$ and $M < \infty$ being constants.

Remark 1. In the case where $\mu_k = \dfrac{1}{1+\|\varphi_k\|^2}$, Condition (2.5) is an extension of the one introduced in (Guo, 1990). It is noted in (Zhang, Guo and Chen, 1990) that the conditional richness condition (2.5) is satisfied by a large class of processes. For example, if $\{\varphi_k\}$ is a ϕ-mixing process with

$$\inf_k \sup_{\|x\|=1} x^\tau E\varphi_k\varphi_k^\tau x > 0 \quad \text{and} \quad \sup_k E\|\varphi_k\|^4 < \infty,$$

then (2.5) holds. Also, if $\{\varphi_k\}$ is an output of a stable and output-controllable linear system, then (2.5) holds. Finally, (2.5) is obviously satisfied if its deterministic version is fulfilled. By the deterministic version of (2.5) we mean the one that was introduced in Lemma 2 of Chen & Guo (1987), and which is weaker than the well-known sufficient richness condition.

Theorem 1. If the conditional richness condition (2.5) is satisfied, then the parameter tracking error is estimated by

$$\limsup_{n\to\infty} E\|\tilde{\theta}_n\|^\beta \leqslant C(\sigma_a)^{\frac{\beta}{\alpha}}, \quad \forall \beta\in(0, \alpha) \tag{2.8}$$

where C is a positive constant.

Moreover, if $v_n\equiv 0$ and $w_n\equiv 0$, then

$$E\|\tilde{\theta}_{n+1}\|^\beta \xrightarrow[n\to\infty]{} 0 \quad \text{exponentially fast}, \quad \forall \beta\in(0, \alpha) \tag{2.9}$$

and

$$\tilde{\theta}_{n+1} \xrightarrow[n\to\infty]{} 0 \quad \text{exponentially fast}. \tag{2.10}$$

Theorem 2. If Condition (2.5) holds, then for any $\beta\in(0, \alpha)$ satisfying $\delta>\frac{2\beta}{\alpha-\beta}$ there exists a constant $A>0$ such that

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n} \|\tilde{\theta}_i\|^\beta \leqslant A(\varepsilon_a)^{\frac{\beta}{\alpha}} \quad \text{a. s.} \tag{2.11}$$

where δ is the constant appearing in Condition (2.5).

The idea of the proof for Theorem 1 comes from the following enlightening fact, the proof of which is given in Appendix.

Proposition 1. Assume that $r=1$, $E|\tilde{\theta}_0|<\infty$ and $\{\mu_n^{\frac{1}{2}}\varphi_n\}$ is a sequence of mutually independent vectors. Then $\sup\limits_{n\geqslant 0} E|\tilde{\theta}_n|<\infty$ for any $\{\xi_n\}\in\mathscr{M}\triangleq\{\{\xi_n\}$: $\sup\limits_{n\geqslant 0} E|\xi_n|<\infty$, $\{\xi_n\}$ is independent of $\{\mu_n^{\frac{1}{2}}\varphi_n\}\}$ if and only if there exist constants $C>0$ and $\gamma\in(0, 1)$ such that

$$E\Phi(n, k)\leqslant C\gamma^{n-k}, \quad \forall n\geqslant k\geqslant 0, \tag{2.12}$$

where $\{\tilde{\theta}_n\}$ and $\{\xi_n\}$ are related by (2.2)—(2.3)

From this Proposition it is seen that in order to prove Theorem 1 the first thing we ought to do is to investigate whether or not (2.12) is satisfied as has been done in many references (see, e.g. Eweda and Macchi, 1985; Macchi, 1986).

Lemma 1. Under Condition (2.5), the following inequality holds

$$\rho_k\leqslant 1-\frac{1}{(1+h)^2\alpha_k}, \quad \forall k\geqslant 0, \tag{2.13}$$

where

$$\rho_k=\lambda_{\max}(E[\Phi^\tau(k+h+1, k+1)\Phi(k+h+1, k+1)|\mathscr{F}_k]). \tag{2.14}$$

where $\lambda_{\max}(X)$ denotes the maximum eigenvalue of a matrix X.

Lemma 2. Suppose that $\{x_m, \mathscr{Y}_m\}$ and $\{\varDelta_m, \mathscr{Y}_m\}$ are adapted sequences such that $x_m\geqslant 1$, $\varDelta_m\geqslant 0$, $\forall m\geqslant 0$,

$$Ex_0^{1+\nu}<\infty, \ \sup_{m\geqslant 0} E[\varDelta_{m+1}^{1+\nu}|\mathscr{Y}_m]\leqslant M \quad \text{a. s.} \tag{2.15}$$

and

$$x_{m+1} \leqslant \mu x_m + \Delta_{m+1}, \quad \forall m \geqslant 0, \tag{2.16}$$

where $\mu \in [0, 1)$, $\nu > 0$ and $M < \infty$ are constants.

Then there exist adapted processes $\{y_m, \mathscr{Y}_m\}$ and $\{\delta_m, \mathscr{Y}_m\}$ such that

$$y_m \geqslant x_m \geqslant 1, \quad \delta_{m+1} \geqslant 0, \quad \forall m \geqslant 0 \tag{2.17}$$

$$\sup_{m \geqslant 0} E[\delta_{m+1}^{1+\nu} | \mathscr{Y}_m] \leqslant \overline{M} \quad \text{a.s.} \tag{2.18}$$

$$\sup_{m \geqslant 0} E[\delta_{m+1} | \mathscr{Y}_m] \leqslant b \quad \text{a.s.} \tag{2.19}$$

and

$$y_{m+1} = b y_m + \delta_{m+1}, \quad m \geqslant 0, \ E y_0^{1+\nu} < \infty, \tag{2.20}$$

where $b \in (0, 1)$ and $\overline{M} < \infty$ are constants.

Lemma 3. Let $\{a_m, \mathscr{Y}_m\}$ be an adapted random process satisfying

$$a_m \in [0, 1], \ E[a_{m+1} | \mathscr{Y}_m] \geqslant \frac{1}{x_m}, \quad \forall m \geqslant 0,$$

where $\{x_m\}$ is given by Lemma 2. Then there exist constants $C > 0$ and $\gamma \in (0, 1)$ such that

$$E \prod_{k=m}^{n} (1 - a_{k+1}) \leqslant C \gamma^{n-m+1}, \quad \forall n \geqslant m \geqslant 0.$$

Lemma 4. Under Condition (2.5) there are constants C and $\gamma \in (0, 1)$ such that

$$E \| \Phi(n, m) \|^2 \leqslant C \gamma^{n-m}, \quad \forall n \geqslant m \geqslant 0. \tag{2.21}$$

Lemma 5. Let $\{f_k, \mathscr{Y}_k\}$, $\{v_k, \mathscr{Y}_k\}$ and $\{a_k, \mathscr{Y}_k\}$ be nonnegative adapted random processes satisfying

$$f_{k+1} \leqslant (1 - a_{k+1}) f_k + v_{k+1}, \quad \forall k \geqslant 0, \ E f_0^{\alpha} < \infty \tag{2.22}$$

for some $\alpha > 0$ and let the following conditions be held

$$B \triangleq \limsup_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n} v_i^{\alpha} < \infty \quad \text{a.s.}, \quad \sup_{k \geqslant 0} E v_{k+1}^{\alpha} < \infty, \tag{2.23}$$

$$a_k \in [0, 1], \ E\{a_{k+1} | \mathscr{Y}_k\} \geqslant \frac{1}{x_k} \tag{2.24}$$

where $\{x_k\}$ is defined by Lemma 2. Then there exists $L \geqslant 0$ such that

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n} f_i^{\beta} \leqslant L B^{\frac{\beta}{\alpha}} \quad \text{a. s.}, \ \forall \beta \in (0, \alpha) \tag{2.25}$$

whenever $\nu > \frac{2\beta}{\alpha - \beta}$, where ν is the constant appearing in (2.15).

The proof for Lemmas 1, 4 and 5 is given in Section 4, while the proof for Lemmas 2 and 3 can be found in Zhang, Guo and Chen (1990).

§ 3. Proof of Theorems

Proof of Theorem 1. We first show (2.9) and (2.10).

If $\frac{\alpha\beta}{\alpha - \beta} \geqslant 2$, then from $\| \Phi(n+1, i) \| \leqslant 1$ and (2.21) it follows that

$$E\|\Phi(n+1,\ i)\|^{\frac{\alpha\beta}{\alpha-\beta}}\leqslant E\|\Phi(n+1,\ i)\|^2\leqslant C\gamma^{n+1-i},\quad \forall n\geqslant i\geqslant 0. \tag{3.1}$$

If $\frac{\alpha\beta}{\alpha-\beta}<2$, then by the Hölder inequality and (2.21), we have, for any $n\geqslant i\geqslant 0$,

$$E\|\Phi(n+1,\ i)\|^{\frac{\alpha\beta}{\alpha-\beta}}\leqslant (E\|\Phi(n+1,\ i)\|^2)^{\frac{\alpha\beta}{2(\alpha-\beta)}}\leqslant (C\gamma^{n+1-i})^{\frac{\alpha\beta}{2(\alpha-\beta)}}. \tag{3.2}$$

Hence for any $\beta\in(0,\ \alpha)$, from (3.1)—(3.2) there always exist constants $C_1\in(0,\ \infty)$ and $\gamma_1\in(0,\ 1)$ such that

$$E\|\Phi(n+1,\ i)\|^{\frac{\alpha\beta}{\alpha-\beta}}\leqslant C_1\gamma_1^{n+1-i},\quad n\geqslant i\geqslant 0. \tag{3.3}$$

In the case where $v_n\equiv 0$ and $w_n\equiv 0$, it follows from (2.3) that

$$\tilde{\theta}_{n+1}=\Phi(n+1,\ 0)\tilde{\theta}_0. \tag{3.4}$$

Applying the Hölder inequality to (3.4) $\left(p=\frac{\alpha}{\beta},\ q=\frac{\alpha}{\alpha-\beta}\right)$ from (3.3) we see that (2.9) is true, i. e. there are constants $\xi\in(0,\ 1)$ and $c_1>0$ such that

$$E\|\tilde{\theta}_n\|^\beta\xi^{-n}<c_1\quad \forall n\geqslant 0.$$

By the Borel-Cantelli lemma it is easy to see that

$$\lim_{n\to\infty}\|\tilde{\theta}_n\|^\beta\xi^{-n/2}=0\quad \text{a. s.}$$

which proves (2.10).

It remains to prove (2.8).

From the definition (2.4) of ξ_n it is easy to see that there exists a constant N such that

$$\sup_{k\geqslant 0}E\|\xi_k\|^\alpha\leqslant N\sigma_\alpha. \tag{3.5}$$

For a given $\beta\in(0,\ \alpha)$, if $\beta\in(0,\ 1]$, from (2.3) and the following elementary inequality

$$(x+y)^\beta\leqslant x^\beta+y^\beta,\quad \forall x\geqslant 0,\ y\geqslant 0,$$

it follows that

$$E\|\tilde{\theta}_{n+1}\|^\beta\leqslant E\|\Phi(n+1,\ 0)\|^\beta\|\tilde{\theta}_0\|^\beta+\sum_{i=0}^n E\|\Phi(n+1,\ i+1)\|^\beta\|\xi_{i+1}\|^\beta,$$

which by the Hölder inequality yields

$$E\|\tilde{\theta}_{n+1}\|^\beta\leqslant (E\|\Phi(n+1,0)\|^{\frac{\alpha\beta}{\alpha-\beta}})^{\frac{\alpha-\beta}{\alpha}}(E\|\tilde{\theta}_0\|^\alpha)^{\frac{\beta}{\alpha}}$$
$$+\sum_{i=0}^n (E\|\Phi(n+1,i+1)\|^{\frac{\alpha\beta}{\alpha-\beta}})^{\frac{\alpha-\beta}{\alpha}}(E\|\xi_{i+1}\|^\alpha)^{\frac{\beta}{\alpha}}. \tag{3.6}$$

If $\beta>1$, from (2.3) by the Minkowski inequality and Hölder inequality we have

$$(E\|\tilde{\theta}_{n+1}\|^\beta)^{\frac{1}{\beta}}\leqslant (E\|\Phi(n+1,\ 0)\|^\beta\|\tilde{\theta}_0\|^\beta)^{\frac{1}{\beta}}+\sum_{i=0}^n(E\|\Phi(n+1,\ i+1)\|^\beta\|\xi_{i+1}\|^\beta)^{\frac{1}{\beta}}$$
$$\leqslant\{(E\|\Phi(n+1,\ 0)\|^{\frac{\alpha\beta}{\alpha-\beta}})^{\frac{\beta-\beta}{\alpha}}(E\|\tilde{\theta}_0\|^\alpha)^{\frac{\beta}{\alpha}}\}^{\frac{1}{\beta}}$$
$$+\sum_{i=0}^n\{(E\|\Phi(n+1,\ i+1)\|^{\frac{\alpha\beta}{\alpha-\beta}})^{\frac{\alpha-\beta}{\alpha}}(E\|\xi_{i+1}\|^\alpha)^{\frac{\beta}{\alpha}}\}^{\frac{1}{\beta}}$$

$$= (E\|\Phi(n+1,\ 0)\|^{\frac{\alpha\beta}{\alpha-\beta}})^{\frac{\alpha-\beta}{\alpha\beta}}(E\|\tilde{\theta}_0\|^{\alpha})^{\frac{1}{\alpha}}$$
$$+\sum_{i=0}^{n}(E\|\Phi(n+1,\ i+1)\|^{\frac{\alpha\beta}{\alpha-\beta}})^{\frac{\alpha-\beta}{\alpha\beta}}(E\|\xi_{i+1}\|^{\alpha})^{\frac{1}{\alpha}}. \tag{3.7}$$

By (3.3) from (3.6) and (3.7) the desired result (2.8) follows immediately.

Q.E.D.

Proof of Theorem 2. For any fixed $s=0,\ 1,\ \cdots,\ h$, set

$$x_k(s)=\|\tilde{\theta}_{k(h+1)+s}\|,\quad \forall k\geqslant 0. \tag{3.8}$$

Noticing from (2.2) and (2.4)

$$\tilde{\theta}_{(k+1)(h+1)+s}=\Phi((k+1)(h+1)+s,\ k(h+1)+s)\tilde{\theta}_{k(h+1)+s}$$
$$+\sum_{i=k(h+1)+s}^{(k+1)(h+1)+s-1}\Phi((k+1)(h+1)+s,\ i+1)\xi_{i+1},$$

then by the boundedness of μ_n we have

$$x_{k+1}(s)\leqslant\|\Phi((k+1)(h+1)+s,\ k(h+1)+s)\tilde{\theta}_{k(h+1)+s}\|+\bar{\xi}_{k+1}(s), \tag{3.9}$$

where

$$\bar{\xi}_{k+1}(s)=\sum_{i=k(h+1)+s}^{(k+1)(h+1)+s-1}(|v_i|+\|w_{i+1}\|). \tag{3.10}$$

Let

$$a_{k+1}(s)=\begin{cases}1-\dfrac{\|\Phi((k+1)(h+1)+s,\ k(h+1)+s)\tilde{\theta}_{k(h+1)+s}\|}{\|\tilde{\theta}_{k(h+1)+s}\|}, & \text{if } \|\tilde{\theta}_{k(h+1)+s}\|>0\\ 1, & \text{if } \|\tilde{\theta}_{k(h+1)+s}\|=0\end{cases} \tag{3.11}$$

It is clear that

$$a_k(s)\geqslant 0,\ a_k(s)\in\mathscr{F}_{k(h+1)+s},\quad a_k(s)\in[0,\ 1]. \tag{3.12}$$

and

$$x_{k+1}(s)\leqslant(1-a_{k+1}(s))x_k(s)+\bar{\xi}_{k+1}(s) \tag{3.13}$$

We now show that

$$E[a_{k+1}(s)\mid\mathscr{F}_{k(h+1)+s}]\geqslant\frac{1}{2(1+h)^2\alpha_{k(h+1)+s}},\quad \text{a.s.} \tag{3.14}$$

Set $A_k(s)=\{w:\ \|\tilde{\theta}_{k(h+1)+s}\|=0\}$. Obviously we have $A_k(s)\in\mathscr{F}_{k(h+1)+s}$ and

$$I_{A_k(s)}E[a_{k+1}(s)\mid\mathscr{F}_{k(h+1)+s}]=E[a_{k+1}(s)I_{A_k(s)}\mid\mathscr{F}_{k(h+1)+s}]=I_{A_k(s)}, \tag{3.15}$$

which means that (3.14) is true on $A_k(s)$ since $\alpha_k\geqslant 1,\ \forall k\geqslant 0$.

From Schwarz inequality and Lemma 1 we see that

$$E[\|\Phi((k+1)(h+1)+s,\ k(h+1)+s)\tilde{\theta}_{k(h+1)+s}\|\mid\mathscr{F}_{k(h+1)+s}]$$
$$\leqslant\{\|\Phi((k+1)(h+1)+s,\ k(h+1)+s)\tilde{\theta}_{k(h+1)+s}\|^2\mid\mathscr{F}_{k(h+1)+s}\}^{\frac{1}{2}}$$
$$\leqslant\left(1-\frac{1}{(1+h)^2\alpha_{k(h+1)+s}}\right)^{\frac{1}{2}}\|\tilde{\theta}_{k(h+1)+s}\|$$
$$\leqslant\left(1-\frac{1}{2(1+h)^2\alpha_{k(h+1)+s}}\right)\|\tilde{\theta}_{k(h+1)+s}\|.$$

Hence by (3.11),

$$I_{A_k^\xi(s)}E[a_{k+1}(s)\,|\,\mathscr{F}_{k(h+1)+s}]=E[I_{A_k^\xi(s)}a_{k+1}(s)\,|\,\mathscr{F}_{k(h+1)+s}]$$
$$\geqslant I_{A_k^\xi(s)}\left(1-\left(1-\frac{1}{2(1+h)^2\alpha_{k(h+1)+s}}\right)\right)$$
$$=I_{A_k^\xi(s)}\frac{1}{2(1+h)^2\alpha_{k(h+1)+s}},$$

which together with (3.15) implies (3.14).

Finally, applying Lemma 5 to (3.13) we obtain

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n}x_i^\beta(s)\leqslant A_s(\varepsilon_a)^{\frac{\beta}{\alpha}},\quad s=0,1,\cdots,h,$$

where $A_s<\infty$ are constants. This verifies the desired result (2.11).

Q. E. D.

§ 4. Proof of Lemmas

Proof of Lemma 1. We first note that this lemma is the stochastic version of the result proved in Lemma 2 of Chen and Guo (1987 c) and the reader will find some similar ideas for their proof.

Let z_{k-1} be the unit eigenvector corresponding to the largest eigenvalue ρ_{k-1} of the matrix

$$E(\Phi^\tau(k+h,k)\Phi(k+h,k)\,|\,\mathscr{F}_{k-1}) \tag{4.1}$$

For any $j\geqslant k$, define z_j recursively by

$$z_j=(I-\mu_j\varphi_j\varphi_j^\tau)z_{j-1}. \tag{4.2}$$

It follows from (4.2) and the definition (2.1) that

$$z_{k+h-1}=\Phi(k+h,k)z_{k-1},$$

and hence,

$$E(\|z_{k+h-1}\|^2|\mathscr{F}_{k-1})=z_{k-1}^\tau E(\Phi^\tau(k+h,k)\Phi(k+h,k)\,|\,\mathscr{F}_{k-1})z_{k-1}$$
$$=\rho_{k-1}\|z_{k-1}\|^2=\rho_{k-1},$$

i. e.

$$\rho_{k-1}=E(\|z_{k+h-1}\|^2|\mathscr{F}_{k-1}). \tag{4.3}$$

We now proceed to find connections between ρ_{k-1} and α_{k-1}.

Notice (4.2), we have

$$z_j=z_{k-1}-\sum_{i=k}^{j}\mu_i\varphi_i\varphi_j^\tau z_{i-1},\quad\forall j\in[k,k+h-1],$$

which leads to

$$E(\|z_{j-1}-z_{k-1}\|^2|\mathscr{F}_{k-1})=E\left[\left\|\sum_{i=k}^{j-1}\mu_i\varphi_i\varphi_i^\tau z_{i-1}\right\|^2\,|\,\mathscr{F}_{k-1}\right]$$
$$\leqslant E\left[\left(\sum_{i=k}^{j-1}\mu_i\|\varphi_i^\tau z_{i-1}\|^2\right)\left(\sum_{i=k}^{j-1}\mu_i\|\varphi_i\|^2\right)|\mathscr{F}_{k-1}\right]$$
$$\leqslant hE\left(\sum_{i=k}^{j-1}\mu_i\|\varphi_i^\tau z_{i-1}\|^2|\mathscr{F}_{k-1}\right),\quad\forall j\in[k,k+h]. \tag{4.4}$$

From (2.5) and the Minkowski inequality we conclude that

$$\alpha_{k-1}^{-\frac{1}{2}}\leqslant\left(z_{k-1}^{\tau}E\left[\sum_{j=k}^{k+h-1}\mu_j\varphi_j\varphi_j^{\tau}\mid\mathscr{F}_{k-1}\right]z_{k-1}\right)^{\frac{1}{2}}=\left(E\left[\sum_{j=k}^{k+h-1}\mu_j|\varphi_j^{\tau}z_{k-1}|^2\mid\mathscr{F}_{k-1}\right]\right)^{\frac{1}{2}}$$

$$\leqslant\left(E\left[\sum_{j=k}^{k+h-1}\mu_j|\varphi_j^{\tau}z_{j-1}|^2\mid\mathscr{F}_{k-1}\right]\right)^{\frac{1}{2}}+\left(E\sum_{j=k}^{k+h-1}\|z_{j-1}-z_{k-1}\|^2\mid\mathscr{F}_{k-1}\right]\right)^{\frac{1}{2}}.$$

From this and (4.4) it follows that

$$\alpha_{k-1}^{-\frac{1}{2}}\leqslant(1+h)\left(E\left[\sum_{j=k}^{k+h-1}\mu_j|\varphi_j^{\tau}z_{j-1}|^2\mid\mathscr{F}_{k-1}\right]\right)^{\frac{1}{2}}$$

i. e.

$$E\left[\sum_{j=k}^{k+h-1}\mu_j|\varphi_j^{\tau}z_{j-1}|^2\mid\mathscr{F}_{k-1}\right]\geqslant\frac{1}{(1+h)^2\alpha_{k-1}}. \tag{4.5}$$

Obviously, to complete the proof we should find the relationship between ρ_{k-1} and

$$E\left[\sum_{j=k}^{k+h-1}\mu_j|\varphi_j^{\tau}z_{j-1}|^2\mid\mathscr{F}_{k-1}\right].$$

From (4.2) it is easy to see that

$$z_j^{\tau}z_j\leqslant z_{j-1}^{\tau}z_{j-1}-\mu_j|\varphi_j^{\tau}z_{j-1}|^2,$$

which implies that

$$\|z_{k+h-1}\|^2\leqslant\|z_{k-1}\|^2-\sum_{j=k}^{j+h-1}\mu_j|\varphi_j^{\tau}z_{j-1}|^2=1-\sum_{j=k}^{k+h-1}\mu_j|\varphi_j^{\tau}z_{j-1}|^2. \tag{4.6}$$

Taking conditional expectations with respect to $\mathscr{F}_{k-1}$ for both sides of (4.6) and noticing (4.3) we have

$$\rho_{k-1}=E(\|z_{k+h-1}\|^2\mid\mathscr{F}_{k-1})\leqslant1-E\left[\sum_{j=k}^{k+h-1}\mu_j|\varphi_j^{\tau}z_{j-1}|^2\mid\mathscr{F}_{k-1}\right],$$

which combining with (4.5) gives the desired result (2.13).

Q. E. D.

Proof of Lemma 4. Let

$$k_0=\min\{k\colon m\leqslant kh\leqslant n\},\qquad k_1=\max\{k\colon m\leqslant kh\leqslant n\}. \tag{4.7}$$

It is clear that, if one of k_0 and k_1 exists, then the other one also exists and $k_1\geqslant k_0\geqslant0$.

Noticing the inequalities

$$E\|\Phi(n,\ m)\|^2\leqslant E\|\Phi(k_1h,\ k_0h)\|^2\quad\text{and}\quad\gamma^{(k_1-k_0-2)h}\leqslant\gamma^{n-m} \tag{4.8}$$

which holds because $(k_1+1)h>n$ and $(k_0-1)h<m$, we find that for (2.21) it suffices to show that there exist constants $C\in(0,\ \infty)$ and $\gamma\in(0,\ 1)$ such that

$$E\|\Phi(k_1h,\ k_0h)\|^2\leqslant C\gamma^{(k_1-k_0+2)h},\quad\forall k_1\geqslant k_0. \tag{4.9}$$

We now proceed to prove (4.9).

Set

$$\rho_{k_1,k_0}=\lambda_{\max}(E[\Phi^{\tau}(k_1h,\ k_0h)\Phi(k_1h,\ k_0h)]) \tag{4.10}$$

and x_{k_1,k_0} be a unit eigenvector of the matrix $E[\Phi^{\tau}(k_1h,\ k_0h)\Phi(k_1h,\ k_0h)]$ corresponding to its largest eigenvalue ρ_{k_1,k_0}. Denote

$$z_k=\Phi(kh,\ (k-1)h)z_{k-1},\quad \forall k\in[k_0+1,\ k_1],\ z_{k_0}=x_{k_0,k_0}. \tag{4.11}$$

From (4.10)—(4.11) and Lemma 1 it follows that

$$z_k\in\mathscr{F}_{kh-1},\ \rho_{k_1,k_0}=E\|z_{k_1}\|^2, \tag{4.12}$$

$$E(\|z_k\|^2|\mathscr{F}_{(k-1)h-1})\leqslant\left(1-\frac{1}{(1+h)^2\alpha_{(k-1)h-1}}\right)\|z_{k-1}\|^2 \tag{4.13}$$

and

$$\|z_k\|^2\leqslant\|\Phi(kh,\ (k-1)h)\|^2\|z_{k-1}\|^2\leqslant\|z_{k-1}\|^2,\quad \forall k\in[k_0+1,\ k_1]. \tag{4.14}$$

Notice that from (2.6) we can get

$$(1+h)^2\alpha_{(k+1)h-1}\leqslant a^h(1+h)^2\alpha_{kh-1}+(1+h)^2\sum_{i=0}^{h-1}a^i\eta_{(k+1)h-i-1},$$

then identifying $x_m=(1+h)^2\alpha_{mh-1}$, $\mu=a^h$, $\Delta_{m+1}=(1+h)^2\sum_{i=0}^{h-1}a^i\eta_{(m+1)h-i-1}$, $\mathscr{Y}_m=\mathscr{F}_{mh-1}$ and $\nu=\delta$ in Lemma 2 we know that there exists a constant $b\in(0,\ 1)$ and two nonnegative adapted processes $\{y_k,\ \mathscr{Y}_k\}$ and $\{\delta_k,\ \mathscr{Y}_k\}$ depending only on $\{\eta_k\}$ such that

$$y_k\geqslant(1+h)^2\alpha_{kh-1}, \tag{4.15}$$

$$y_{k+1}=by_k+\delta_{k+1},\quad \forall k\geqslant0,\ Ey_0^{1+\delta}<\infty \tag{4.16}$$

and

$$\sup_{k\geqslant0}E[\delta_{k+1}^{1+\delta}|\mathscr{Y}_k]\leqslant\overline{M},\quad \sup_{k\geqslant0}E[\delta_{k+1}|\mathscr{Y}_k]\leqslant b, \tag{4.17}$$

where $\overline{M}<\infty$ is a constant.

We are now in a position to complete the proof of (4.9).

From (4.12)—(4.17) it is easy to see that

$$\begin{aligned}\rho_{k_1,k_0}&=E\|z_{k_1}\|^2\leqslant Ey_{k_1}\|z_{k_1}\|^2=E(by_{k_1-1}+\delta_{k_1})\|z_{k_1}\|^2\\&=bE[y_{k_1-1}E(\|z_{k_1}\|^2|\mathscr{F}_{(k_1-1)h-1})]+E\delta_{k_1}\|z_{k_1}\|^2\\&\leqslant bE\left[y_{k_1-1}\left(1-\frac{1}{y_{k_1-1}}\right)\|z_{k_1-1}\|^2\right]+E\delta_{k_1}\|z_{k_1-1}\|^2\\&=bE[y_{k_1-1}\|z_{k_1-1}\|^2]-bE\|z_{k_1-1}\|^2+E[\|z_{k_1-1}\|^2E(\delta_{k_1}|\mathscr{F}_{(k_1-1)h-1})]\\&=bE[y_{k_1-1}\|z_{k_1-1}\|^2]\leqslant\cdots\leqslant b^{k_1-k_0}Ey_{k_0}\\&=\left(Ey_0+\frac{b}{1-b}\right)b^{k_1-k_0},\end{aligned}$$

which implies (4.9), and hence (2.21) holds.

Q.E.D.

Proof of Lemma 5. We need the following fact: for any martingale difference sequence $\{g_n,\ \mathscr{Y}_n\}$, if

$$\sup_{n\geqslant0}E|g_n|^{1+c}<\infty,\quad\text{for some } c>0, \tag{4.18}$$

then

$$\frac{1}{n}\sum_{k=0}^{n}g_k\xrightarrow[n\to\infty]{}0\quad\text{a.s.} \tag{4.19}$$

To prove this, we note that (without loss of generality assume $c\in(0,\ 1]$),

$$E\left\{\sum_{k=1}^{\infty} E\left[\left|\frac{g_k}{k}\right|^{1+c} \mid \mathscr{Y}_{k-1}\right]\right\} = \sum_{k=1}^{\infty} \frac{E|g_k|^{1+c}}{k^{1+c}} < \infty,$$

and hence

$$\sum_{k=1}^{\infty} E\left[\left|\frac{g_k}{k}\right|^{1+c} \mathscr{Y}\,|_{k-1}\right] < \infty \quad \text{a. s.}$$

Consequently, by the martingale convergence theorem (Chow, 1965) the series $\sum_{k=1}^{\infty} \frac{g_k}{k}$ converges almost surely. From this and the Kronecker lemma (4.19) follows immediately.

We now show that for any $\varepsilon>0$ and $d>1$ there is a constant $c(\varepsilon, d)$ such that

$$(a+b)^d \leqslant (1+\varepsilon)a^d + c(\varepsilon, d)b^d, \quad \forall a\geqslant 0,\ b\geqslant 0. \tag{4.20}$$

Notice that there is x_0 such that

$$(1+x)^d \leqslant (1+\varepsilon)x^d \quad \forall x\geqslant x_0$$

which implies

$$(1+x)^d \leqslant (1+\varepsilon)x^d + c(\varepsilon, d), \quad \forall x\geqslant 0 \tag{4.21}$$

by setting $c(\varepsilon, d) = (1+x_0)^d$.

From (4.21) we readily obtain (4.20). Hence for any $\varepsilon>0$ and any $d>0$ we have

$$(a+b)^d \leqslant f(\varepsilon, d)a^d + g(\varepsilon, d)b^d, \quad \forall a\geqslant 0,\ b\geqslant 0, \tag{4.22}$$

where

$$f(\varepsilon, d) = \begin{cases} 1+\varepsilon, & \text{if } d>1, \\ 1, & \text{if } d\leqslant 1; \end{cases} \quad g(\varepsilon, d) = \begin{cases} c(\varepsilon, d), & \text{if } d>1, \\ 1, & \text{if } d\leqslant 1; \end{cases} \tag{4.23}$$

Then from (2.22) we see that, for any $\beta\in(0, \alpha)$,

$$f_{k+1}^{\frac{\alpha+\beta}{2}} \leqslant (1-ta_{k+1})f\left(\varepsilon, \frac{\alpha+\beta}{2}\right)f_k^{\frac{\alpha+\beta}{2}} + g\left(\varepsilon, \frac{\alpha+\beta}{2}\right)\nu_{k+1}^{\frac{\alpha+\beta}{2}}, \quad \forall k\geqslant 0 \tag{4.24}$$

and

$$f_{k+1}^{\beta} \leqslant (1-ta_{k+1})f(\varepsilon, \beta)f_k^{\beta} + g(\varepsilon, \beta)\nu_{k+1}^{\beta}, \quad \forall k\geqslant 0, \tag{4.25}$$

where we have used the inequality

$$(1-a_{k+1})^{\frac{\alpha+\beta}{2}} \leqslant (1-a_{k+1})^{\beta} \leqslant (1-ta_{k+1}) \quad \text{with} \quad t=\min(1, \beta).$$

It is clear that Lemma 3 still holds if $(a_m, \mathscr{Y}_m)$ is replaced by $(ta_m, \mathscr{Y}_m)$. In this case γ should be written as $\gamma(t)\in(0, 1)$ to emphasize its dependence on t.

In order that $\gamma(t)\,f\left(\varepsilon, \frac{\alpha+\beta}{2}\right)<1$, ε should belong to the following interval:

$$\varepsilon\in(0,\ (\gamma(t))^{-1}-1) \tag{4.26}$$

Then from Lemma 3, (2.23) and (4.24) it follows that

$$\sup_{k\geqslant 0} E\, f_k^{\frac{\alpha+\beta}{2}} < \infty. \tag{4.27}$$

By Lemma 2 it is easy to see that for the given number $t=\min\{1, \beta\}$, we can find an adapted sequence $\{y_n, \mathscr{Y}_n\}$ satisfying (2.18)—(2.20) and

$$y_m \geqslant \frac{1}{t} x_m, \quad \forall m \geqslant 0. \tag{4.28}$$

Now, let us set

$$\tilde{a}_k \triangleq a_k - E(a_k | \mathscr{Y}_{k-1}), \quad \tilde{\delta}_k \triangleq \delta_k - E(\delta_k | \mathscr{Y}_{k-1}), \tag{4.29}$$

where $\{\delta_k\}$ is the process satisfying (2.20). By (4.28), (2.19) and (2.24) it is easy to see that

$$a_{k+1} \geqslant \tilde{a}_{k+1} + \frac{1}{t y_k}, \ \delta_k \leqslant \tilde{\delta}_k + b. \tag{4.30}$$

By this, (4.25) and (2.20) we have

$$\begin{aligned} y_{k+1} f_{k+1}^\beta &\leqslant (b y_k + \delta_{k+1})(1 - t a_{k+1}) f(s, \beta) f_k^\beta + g(s, \beta) \nu_{k+1}^\beta y_{k+1} \\ &\leqslant b f(s, \beta)(1 + t a_{k+1}) f_k^\beta y_k + f(s, \beta) \delta_{k+1} f_k^\beta + g(s, \beta) \nu_{k+1}^\beta y_{k+1} \\ &\leqslant b f(s, \beta)\left(1 - t \tilde{a}_{k+1} - \frac{t}{t y_k}\right) f_k^\beta y_k + f(s, \beta)(\tilde{\delta}_{k+1} + b) f_k^\beta \\ &\quad + g(s, \beta) \nu_{k+1}^\beta y_{k+1} \\ &= b f(s, \beta) f_k^\beta y_k - b f(s, \beta) t \tilde{a}_{k+1} f_k^\beta y_k + f(s, \beta) \tilde{\delta}_{k+1} f_k^\beta \\ &\quad + g(s, \beta) \nu_{k+1}^\beta y_{k+1}. \end{aligned} \tag{4.31}$$

We now proceed to estimate the last three terms on the right hand side of (4.31).

By (2.18), (2.20) and the fact that $b \in (0, 1)$ it is easy to see that

$$\sup_{k \geqslant 0} E y_k^{1+\nu} < \infty, \ \sup_{k \geqslant 0} E |\tilde{\delta}_k|^{1+\nu} < \infty, \tag{4.32}$$

Notice that $|\tilde{a}_k| \leqslant 1$, from (4.32), (4.27) and the Hölder inequality $\left(\text{with } p = \frac{\alpha + 3\beta + 2\beta\nu}{2\beta(1+\nu)}, \ q = \frac{\alpha + 3\beta + 2\beta\nu}{\alpha + \beta}\right)$ we see that

$$\sup_{k \geqslant 0} E |\tilde{a}_{k+1} f_k^\beta y_k|^{1+\lambda} < \infty, \quad \forall \lambda \in \left(0, \frac{(\alpha - \beta)\nu - 2\beta}{\alpha + 3\beta + 2\beta\nu}\right] \tag{4.33}$$

and

$$\sup_{k \geqslant 0} E |\tilde{\delta}_{k+1} f_k^\beta|^{1+\lambda} < \infty, \ \forall \lambda \in \left(0, \frac{(\alpha - \beta)\nu - 2\beta}{\alpha + 3\beta + 2\beta\nu}\right]. \tag{4.34}$$

Therefore, by the fact mentioned by (4.18)—(4.19) we obtain that

$$\frac{1}{n} \sum_{k=1}^{n} (-b f(s, \beta) t \tilde{a}_{k+1} f_k^\beta y_k + f(s, \beta) \tilde{\delta}_{k+1} f_k^\beta) \xrightarrow[n \to \infty]{} 0 \quad \text{a. s.} \tag{4.35}$$

By (4.31) and $y_k \geqslant 1$ for any $k \geqslant 0$, we know that

$$\begin{aligned} \frac{1}{n} \sum_{k=0}^{n} f_k^\beta &\leqslant \frac{1}{n} \sum_{k=0}^{n} y_k f_k^\beta \\ &\leqslant \frac{1}{(1 - b f(s, \beta)) n} \Big[y_0 f_0^\beta + \sum_{k=0}^{n-1} (-b f(s, \beta) t \tilde{a}_{k+1} f_k^\beta y_k + f(s, \beta) \tilde{\delta}_{k+1} f_k^\beta) \\ &\qquad + \sum_{k=0}^{n-1} g(s, \beta) \nu_{k+1}^\beta y_{k+1} \Big] \end{aligned} \tag{4.36}$$

if

$$1 - b f(s, \beta) > 0. \tag{4.37}$$

In the case where $d \leqslant 1$ for (4.37) is obviously true, while in the case where $d>1$, $f(\varepsilon, \beta)=1+\varepsilon$ and for (4.37) it suffices to require

$$\varepsilon < b^{-1} - 1$$

which incorporating with the second interval in (4.26) yields

$$\varepsilon \in (0,\ (\max\{\gamma(t),\ b\})^{-1} - 1). \tag{4.38}$$

Therefore, (4.36) holds if ε is selected to satisfy (4.38).

To complete the proof it remains to analyse the last term on the right hand side of (4.36).

From (2.18) and the fact mentioned by (4.18)—(4.19) it is not difficult to see that for any $c \in (0, \nu)$,

$$\frac{1}{n}\sum_{k=1}^{n}\delta_k^{1+c} = \frac{1}{n}\sum_{k=1}^{n}E(\delta_k^{1+c} \mid \mathscr{Y}_{k-1}) + \frac{1}{n}\sum_{k=1}^{n}(\delta_k^{1+c} - E(\delta_k^{1+c} \mid \mathscr{Y}_{k-1})) = O(1),$$

which in conjunction with (2.20) yields that

$$\frac{1}{n}\sum_{k=1}^{n} y_k^{1+c} = O(1). \tag{4.39}$$

Finally, from (4.35) and (4.36) we see that for (2.25) it remains to show

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\nu_{k+1}^{\beta}y_{k+1} \leqslant L' B^{\frac{\beta}{\alpha}} \tag{4.40}$$

for some constant $L'>0$. By the Hölder inequality with $p=\frac{\alpha}{\beta}$, $q=\frac{\alpha}{\alpha-\beta}$ the desired result (4.40) follows from (2.23) and (4.39).

Q. E. D.

§ 5. Conclusion

An LMS-like algorithm is applied for estimating the time-varying parameter θ_n in the linear model (1.1). For the case where (2.5) holds, it is shown that the α-th moment of the estimation error is of order of the α-th moment of the observation noise and the parameter variation $w_n \triangleq \theta_n - \theta_{n-1}$. It is worth noticing that in this paper we do not require $\{y_n\}$, $\{\varphi_n\}$, $\{\theta_n\}$ and $\{v_n\}$ are stationary, Markovian, independent, ergodic and purely nondeterministic. It is worth mentioning that if we assume these processes are strictly stationary and ergodic then we can also obtain the same limit results presented in (Solo, 1990), even under the weaker conditional richness condition (2.5).

Acknowledgement The second author would like to thank Dr. S. P. Meyn for his comments on the proof of an earlier version of Lemma 1.

Appendix

Proof of the Proposition. If (2.12) holds for some constants $C>0$ and $\gamma \in (0, 1)$, then from (2.3) and $|\Phi(n, i)| \leqslant 1$ $(\forall n \geqslant i \geqslant 0)$ it is easy to see that $\sup_{n\geqslant 0} E|\tilde{\theta}_n| < \infty$ for any $\{\xi_n\} \in \mathscr{M}$.

We now show that if $\sup_{n\leqslant 0} E|\tilde{\theta}_n| < \infty$ for any $\{\xi_n\} \in \mathscr{M}$, then there must exist constants $C>0$ and $\gamma \in (0, 1)$ such that (2.12) holds. To see this, we choose the sequence $\{\xi_n \equiv 1\}$ which obviously belongs to $\mathscr{M}$.

Hence, from (2.3), $E|\tilde{\theta}_0|<\infty$ and $\sup_{n\geq 0} E|\tilde{\theta}_n|<\infty$ we conclude that

$$\sup_{n\geq 0}\sum_{i=0}^{n} E\Phi(n+1,\ i+1)\leqslant C<\infty, \tag{A. 1}$$

where $C>0$ is a constant.

Let

$$b_n=\prod_{i=0}^{n} E(1-\mu_i\varphi_i\varphi_i^\tau). \tag{A. 2}$$

Then from (A. 1) and by the independence of $\{\mu_n^{\frac{1}{2}}\varphi_n\}$ we see that

$$\sum_{i=k}^{n} E\Phi(n+1,\ i+1)=\sum_{i=k}^{n}\prod_{j=i+1}^{n} E(1-\mu_j\varphi_j\varphi_j^\tau)=b_n\sum_{i=k}^{n} b_i^{-1}\leqslant C,\qquad \forall n\geqslant k\geqslant 0$$

which implies that

$$\sum_{i=k}^{n-1} b_i^{-1}b_k\leqslant Cb_n^{-1}b_k. \tag{A. 3}$$

Adding $C\sum_{i=k}^{n-1} b_i^{-1}b_k$ to both sides of (A. 3) leads to

$$(1+C)\sum_{i=k}^{n-1} b_i^{-1}b_k\leqslant C\sum_{i=k}^{n} b_i^{-1}b_k$$

i. e.

$$\sum_{i=k}^{n} b_i^{-1}b_k\geqslant\left(1+\frac{1}{C}\right)\sum_{i=k}^{n-1} b_i^{-1}b_k\geqslant\left(1+\frac{1}{C}\right)^{n-k}.$$

From this and (A. 3) it is easy to see that

$$b_{n+1}b_k^{-1}\leqslant C\left(\frac{1}{1+C}\right)^{n-k},\quad \forall n\geqslant k\geqslant 0,$$

which implies the desired result (2.12).

Q. E. D.

References

Anderson, B. D.O. R. R. Bitmead, C. R. Johnson Jr., P. V. Kokotovic, R. L. Kosut, I. M. Y. Mareels, L. Praly and B. D. Riedle (1986), *Stability of Adaptive Systems: Passivity and Averaging Analysis*, Cambridge MA, MIT Press.

Benveniste, A., M. Metivier, P. Priouret (1987), *Algorithms Adaptifs et Approximations Stochastiques*, Paris, France: Masson.

Bitmead, R. R., B. D. O. Anderson, (1980), Performance of adaptive estimation algorithm in dependent random environments, *IEEE Trans. on Automatic Control*, **AC-25**, 788—794.

Chen,H. F., P. E. Caines (1990), On the adaptive stabilization of linear stochastic systems with jump process parameters, submitted for publication.

Chen, H. F., L. Guo (1987), Adaptive control via consistent estimation for deterministic systems, *Int. J. Control*, **45**, 2183—2202.

Chow, Y. S., (1965), Local convergence of martingalcs and the law of large numbers, *Ann. Math. Stat.*, **36**, 552—558.

Eweda, E., O. Macchi (1985), Tracking error bounds of adaptive nonstationary filtering, *Automatica*, **21**, 293—302.

Guo, L. (1990), Estimating time-varying parameters by Kalman filter based algorithm: stability and convergence, *IEEE Trans. on Automatic Control*, **35**, 141—147.

Guo, L., S. P. Meyn (1989), Adaptive control for time-varying systems: A combination of martingale and Markov chain techniques, *Int. J. Adaptive Control and Signal Processing*, **3**, 1—14.

Ji, Y. D., H. J. Chizeck (1988), Controllability, observability and discre-tetime markovian jump linear quadratic control, *Int. J. Control*, **48**, 481—498.

Kitagava, G., W. Gersh (1985), A smoothness priors time varying AR coefficient modelling of nonstationary covariance time series, *IEEE Trans. on Automatic Control*, **AC-30**, 48—56.

Macchi, O. (1986), Optimization of adaptive identification for time-varying filters, *IEEE Trans. on Automatic Control*, **AC-31**, 283—287.

Solo, V. (1990), The limiting behavior of LMS, *IEEE Trans. on Acoustics, Speech, and Signal Processing*, **37**, 1909—1922.

Widrow, B., J. M. McCool, M. G. Larimore, C. R. Johnson Jr. (1976), Stationary and nonstationary learning characteristics of the LMS adaptive filter, *Proc. IEEE*, **64**, 1151—1162.

Zhang, J. F., L. Guo, H. F. Chen(1990), L_p-stability of tracking errors for timevarying systems, Submitted for publication, January.