SIAM J. CONTROL AND OPTIMIZATION
Vol. 25, No. 4, July 1987

002

OPTIMAL ADAPTIVE CONTROL AND CONSISTENT PARAMETER ESTIMATES FOR ARMAX MODEL WITH QUADRATIC COST*

HAN-FU CHEN† AND LEI GUO†

Abstract. We consider the multidimensional ARMAX model

$$A(z)y_n = B(z)u_n + C(z)w_n$$

with loss function

$$J(u) = \overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{i=1}^{n} (y_i^T Q_1 y_i + u_i^T Q_2 u_i)$$

where the coefficients in the matrix polynomials $A(z)$, $B(z)$ and $C(z)$ are unknown. Conditions used here are: 1) stability of $A(z)$ and full rank of A_p; 2) strictly positive realness of $C(z)-\frac{1}{2}I$, and 3) controllability and observability of a matrix triple consisting of coefficients in $A(z)$, $B(z)$ and Q_1. On the basis of the estimates given by the stochastic gradient algorithm for unknown parameters an adaptive control law is recursively defined. It is proved that the parameter estimates are strongly consistent and the quadratic loss function reaches its minimum. This paper also includes some general theorems on parameter estimation, on which the results about adaptive control are essentially based.

Key words. stochastic systems, ARMAX model, stochastic adaptive control, quadratic cost, parameter estimation

AMS(MOS) subject classification. 93C40

1. Introduction and statement of problem. In recent years there has been considerable research effort on the parameter estimation and adaptive control problem for linear stochastic systems (see e.g. Goodwin et al. (1984)). Ljung (1977), Solo (1979), Chen (1981), (1982) and Lai and Wei (1982) showed various conditions guaranteeing strong consistency of parameter estimates given by different algorithms for stochastic systems without monitoring, while Goodwin et al. (1981) and Sin and Goodwin (1982) gave adaptive control making the system global stable and the tracking error minimal, but the parameter estimates given there in general, as shown by Becker et al. (1985), are inconsistent. The first step towards getting both consistency of estimates and asymptotic minimality of tracking errors was made by Caines and Lafortune (1984), Chen (1984) and Chen and Caines (1985). In their results the parameter estimates are proved strongly consistent but the tracking error is no longer minimal because of the disturbance artificially introduced to the reference signal. Recently, Chen and Guo (1985a), (1985b) have given an adaptive control under which not only the parameter estimates are strongly consistent, but also the long run average of tracking error reaches its minimum.

For stochastic adaptive control when a general quadratic loss function is considered, Kumar (1983), Hijab (1983) and Caines and Chen (1985) are concerned with the case where the unknown parameters are valued in a finite set, Chen and Caines (1984) and Chen (1985) deal with systems for which the consistent parameter estimates are available, and Samson (1983) considers bounded disturbance case. Recently for systems in state space representation with state completely observed, Chen and Guo

* Received by the editors July 1, 1985; accepted for publication (in revised form) March 20, 1986. This work was supported by the Science Fund of the Chinese Academy of Sciences.
† Institute of Systems Science, Academia Sinica, Beijing, People's Republic of China.

(1986) have given the optimal stochastic LQ control based on the least squares estimates for unknown parameters which may take arbitrary values in the Euclidean spaces of compatible dimensions.

In this paper we consider the general stochastic MIMO system (ARMAX model):

$$A(z)y_n = B(z)u_n + C(z)w_n \tag{1}$$

with quadratic loss function

$$J(u) = \overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{i=0}^{n-1} (y_i^\tau Q_1 y_i + u_i^\tau Q_2 u_i), \tag{2}$$

where $Q_1 \geqq 0$, $Q_2 > 0$ and the matrix polynomials in shift-back operator z

$$A(z) = I + A_1 z + \cdots + A_p z^p, \qquad p \geqq 0, \tag{3}$$

$$B(z) = B_1 z + B_2 z^2 + \cdots + B_q z^q, \qquad q \geqq 1, \tag{4}$$

$$C(z) = I + C_1 z + \cdots + C_r z^r, \qquad r \geqq 0 \tag{5}$$

are of known orders p, q and r, respectively, and with unknown parameter θ denoting

$$\theta^\tau = [-A_1 \cdots -A_p \; B_1 \cdots B_q \; C_1 \cdots C_r] \tag{6}$$

by definition. We emphasize that A_i, B_j, C_k $(i = 1 \cdots p, j = 1 \cdots q, k = 1 \cdots r)$ may be any matrices of compatible dimensions.

Let dimensions for y_n, u_n and w_n be m, l and m, respectively, $y_i = 0$, $u_i = 0$, $w_i = 0$ for $i < 0$, and let $\{w_n\}$ be a martingale difference sequence with respect to a family $\{\mathscr{F}_n\}$ of increasing σ-algebras, i.e., w_n is $\mathscr{F}_n$-measurable and $E(w_n | \mathscr{F}_{n-1}) = 0$. In addition, we assume that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} w_i w_i^\tau = Q > 0 \tag{7}$$

and

$$\sup_n E[\|w_n\|^2 | \mathscr{F}_{n-1}] < \infty \quad \text{a.s.} \tag{8}$$

where and hereafter $\|X\|$ denotes the maximum singular value of X.

At any time n, by use of the past input-output data $\{u_i, y_j, 0 \leqq i \leqq n-1, 0 \leqq j \leqq n\}$ we want 1) to estimate the unknown parameter θ and 2) to define adaptive control u_n^a minimizing the loss function (2). In this paper, for the case where $A(z)$ is stable, we give a complete solution of this problem in the sense that the consistency of parameter estimates and minimality of the loss function are achieved simultaneously. Although the results are established for adaptive control based on parameter estimates given by the stochastic gradient algorithm, the same results also hold for the case where the extended least squares algorithm is applied.

In § 2 we describe the optimal control for system (1) and (2) with known parameters, and in § 3 we define the algorithm for both parameter estimation and adaptive control and formulate the main theorem of this paper. For its proof we start with some general theorems on strong consistency of parameter estimates for systems without monitoring (§ 4). Then in § 5 we prove that they can be applied to the adaptive control system defined in § 3, and show that the loss function is really minimized.

2. Optimal control for systems with known parameters. The adaptive control law given later on is inspired by the optimal control for system (1), (2) with known parameters. So we first rewrite (1) in the state space form

$$x_{k+1}=Ax_k+Bu_k+Cw_{k+1}, \tag{9}$$

$$y_k=Hx_k, \qquad x_0^\tau=[y_0^\tau 0\cdots 0] \tag{10}$$

and give a solution of optimal control, where

$$A=\begin{bmatrix} -A_1 & I & 0 & \cdots & 0 \\ & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ & & & & I \\ -A_s & 0 & & \cdots & 0 \end{bmatrix}, \quad B=\begin{bmatrix} B_1 \\ \vdots \\ B_s \end{bmatrix}, \tag{11}$$

$$C^\tau=[I \quad C_1^\tau \quad \cdots \quad C_{s-1}^\tau], \qquad H=[\underbrace{I \quad 0 \quad \cdots \quad 0}_{ms}]m \tag{12}$$

with $s=p\vee q\vee(r+1)$ and $A_i=0$, $B_j=0$, $C_k=0$ for $i>p, j>q, k>r$.

We note at once that the nonzero eigenvalues of A coincide with the reciprocals of zeros of $\det A(z)$ (Chen (1985)).

All conditions used in this paper are listed here.

(a) A_p is of full rank ($A_0=I$ by definition) and $A(z)$ is stable, i.e. all zeros of $\det A(z)$ lie outside the closed unit disk.

(b) $C(z)-\frac{1}{2}I$ is strictly positive real, i.e.

$$C(e^{i\varphi})+C^\tau(e^{-i\varphi})-I>0 \quad \forall\varphi\in[0,2\pi].$$

(c) (A, B, D) is controllable and observable, where D is any matrix such that $D^\tau D=H^\tau Q_1 H$.

We first explain these conditions.

(1) The full rank of A_p is used to ensure $\deg(\det A(z))=mp$ for identifiability.

(2) For the uncorrelated noise case $r=0$, $C(z)=I$, condition (b) is automatically satisfied.

(3) Condition (c) implies that $A(z)$ and $B(z)$ have no common left factor, i.e. there are matrix polynomials $M(z)$ and $N(z)$ such that

$$A(z)M(z)+B(z)N(z)=I; \tag{13}$$

this is a consequence of Theorem 6.2-6 of Kailath (1980, p. 366). Also, condition (c) implies either A_s or B_s is not zero, which implies $r+1\leqq\max(p, q)$. So under condition (c) $s=p\vee q$.

(4) If condition (c) is fulfilled (stability of $A(z)$ is not required here), then there is a unique positive definite matrix solution S in the class of nonnegative definite matrices for the Riccati algebraic equation

$$S=A^\tau SA-A^\tau SB(Q_2+B^\tau SB)^{-1}B^\tau SA+H^\tau Q_1 H, \tag{14}$$

and the matrix $A+BL$ is stable with

$$L=-(Q_2+B^\tau SB)^{-1}B^\tau SA \tag{15}$$

(see, e.g. Anderson and Moore (1971)).

(5) Instead of condition (c), which is rather restrictive, we can directly assume (14), (15) for which the weaker conditions are sufficient and assume that $A(z)$, $B(z)$ and $C(z)$ have no common left factor which is a natural condition for identifiability of the system.

The following lemma is not concerned with adaptive control but it shows the minimal value of the loss function and hints the form of adaptive control.

Throughout the paper, the relationship between two random quantities may have an exceptional set with probability 0, but sometimes we omit to write "a.s."

LEMMA 1. *If conditions* (a) *and* (c) *hold, then*

$$J(u) = \operatorname{tr} SCQC^\tau + \overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{i=0}^{n-1} (u_i - Lx_i)^\tau (Q_2 + B^\tau SB)(u_i - Lx_i) \quad a.s. \tag{16}$$

whenever u_i *is* $\mathcal{F}_i$*-measurable and* $\{u_i\} \in U$ *with*

$$U = \left\{ u: \sum_{i=1}^{n} \|u_i\|^2 = O(n), \quad \|u_n\|^2 = o(n), \text{ as } n \to \infty \quad a.s. \right\}. \tag{17}$$

The proof is given in Appendix 1.

This lemma tells us that the optimal control is $u_n = Lx_n$ and that the lower bound to the loss is

$$\min_{u \in U} J(u) = \operatorname{tr} SCQC^\tau.$$

We now give a multidimensional version of a result from Lai and Wei (1982) which is used in the proof of Lemma 1 and will be repeatedly used in the sequel.

LEMMA 2. *Let* f_i *be* $\mathcal{F}_i$*-measurable random vectors and let* $\{w_i, \mathcal{F}_i\}$ *be a martingale difference sequence satisfying* (8). *Then as* $n \to \infty$

$$\sum_{i=1}^{n} f_i w_{i+1}^\tau = O(s_n^{1/2} \log^{(1/2)+\eta}(s_n + e)) \quad \forall \eta > 0 \quad \text{with } s_n \triangleq \sum_{i=1}^{n} \|f_i\|^2.$$

The proof is given in Appendix 1.

3. Main theorem.

For estimating the unknown parameter θ we use the stochastic gradient algorithm defined by

$$\theta_{n+1} = \theta_n + \frac{\varphi_n}{r_n}(y_{n+1}^\tau - \varphi_n^\tau \theta_n), \tag{18}$$

$$\varphi_n^\tau = [y_n^\tau, \cdots y_{n-p+1}^\tau, u_n^\tau, \cdots u_{n-q+1}^\tau, y_n^\tau - \varphi_{n-1}^\tau \theta_{n-1}, \cdots, y_{n-r+1}^\tau - \varphi_{n-r}^\tau \theta_{n-r}], \tag{19}$$

$$r_n = 1 + \sum_{i=1}^{n} \|\varphi_i\|^2, \qquad r_0 = 1. \tag{20}$$

Denote by A_{in}, B_{jn}, C_{kn} the estimates given by θ_n for A_i, B_j, C_k, respectively, $i = 1 \cdots p$, $j = 1 \cdots q$, $k = 1 \cdots r$. The state x_n is estimated by the adaptive filter

$$\begin{gathered} \hat{x}_{n+1} = \hat{A}_n \hat{x}_n + \hat{B}_n u_n + \hat{C}_n (y_{n+1} - H\hat{A}_n \hat{x}_n - H\hat{B}_n u_n), \\ \hat{x}_0 = [y_0^\tau 0 \cdots 0]^\tau \end{gathered} \tag{21}$$

where $\hat{A}_n$, $\hat{B}_n$ and $\hat{C}_n$ are defined by (11) and (12) with A_i, B_j, C_k replaced by their estimates A_{in}, B_{jn}, C_{kn}, respectively, $i = 1 \cdots p$, $j = 1 \cdots q$, $k = 1 \cdots r$.

Set

$$L_n = -(\hat{B}_n^\tau S_n \hat{B}_n + Q_2)^{-1} \hat{B}_n^\tau S_n \hat{A}_n, \tag{22}$$

where S_n is recursively defined by

$$S_n = \hat{A}_n^\tau S_{n-1} \hat{A}_n - \hat{A}_n^\tau S_{n-1} \hat{B}_n (Q_2 + \hat{B}_n^\tau S_{n-1} \hat{B}_n)^{-1} \hat{B}_n^\tau S_{n-1} \hat{A}_n + H^\tau Q_1 H, \tag{23}$$

with an arbitrary initial value $S_0 \geqq 0$.

It is natural to guess that $L_n\hat{x}_n$ is something we should take as adaptive control, but, in fact, it may lead to an inconsistent estimate for θ. To avoid this trouble we use the randomly varying truncation technique and the attenuating excitation technique similar to those used in Chen and Guo (1986).

Take an arbitrary l-dimensional i.i.d. sequence $\{\varepsilon_n\}$ independent of $\{w_n\}$ and with properties

$$E\varepsilon_1=0,\quad E\varepsilon_1\varepsilon_1^\tau=I,\quad E\|\varepsilon_1\|^3<\infty. \tag{24}$$

Without loss of generality we assume $\mathscr{F}_n=\sigma\{w_i,\, i\leqq n,\, \varepsilon_j, j\leqq n\}$.

Then the random sequence $\{v_n\}$ will serve as the source of attenuating excitation, where by definition

$$v_1=0,\quad v_n=\frac{\varepsilon_n}{\log^{\varepsilon/2} n}\quad \forall n\geqq 2,\quad \varepsilon\in\left(0,\frac{1}{4s(m+2)}\right). \tag{25}$$

From Theorem 3, which is stated later on, we shall see that for strong consistency of parameter estimates besides conditions on system structure there is a growth rate requirement for system input when the attenuating excitation is applied to the control. But $L_n\hat{x}_n$ may not meet this requirement. This is the motivation to truncate the control at randomly varying bounds which we describe right now.

We partition the time axis by a sequence of stopping times

$$1=\tau_1<\sigma_1<\tau_2<\sigma_2<\cdots$$

at which the control is truncated in order to keep the required growth rate.

From the random time τ_k we define adaptive control u_n^a as $L_n\hat{x}_n$ excited by v_n as far as $n<\sigma_k$, where σ_k is the first time when the growth rate of $1/(j-1)\sum_{i=\tau_k}^{j-1}\|L_i\hat{x}_i\|^2$ is greater, roughly speaking, than $\log^\delta(j-1)$; and from the random time σ_k we define adaptive control as a pure disturbance v_n until $n<\tau_{k+1}$ where τ_{k+1} indicates the time when $(1/n)\sum_{j=1}^n\|\hat{x}_j\|^2$ is less than $\log^{\delta/2} n$ and when some other technical conditions are satisfied. To be precise, we define

$$\sigma_k=\sup\left\{t>\tau_k\colon \sum_{i=\tau_k}^{j-1}\|L_i\hat{x}_i\|^2\leqq(j-1)\log^\delta(j-1)+\|L_{\tau_k}\hat{x}_{\tau_k}\|^2,\quad \forall j\in(\tau_k,t]\right\}, \tag{26}$$

$$\tau_{k+1}=\inf\left\{t>\sigma_k\colon \sum_{i=\tau_k}^{\sigma_k-1}\|L_i\hat{x}_i\|^2\leqq\frac{t\log^\delta t}{2^k};\ \sum_{j=1}^{t}\|\hat{x}_j\|^2\leqq t\log^{\delta/2} t;\ \frac{\|L_t\hat{x}_t\|^2}{t\log^\delta t}\leqq 1\right\} \tag{27}$$

with any but fixed δ such that

$$\delta\in\left(0,\frac{\frac{1}{4}-(m+2)s\varepsilon}{2+(m+1)s}\right). \tag{28}$$

Clearly, for any $\varepsilon\in(0,1/(4s(m+2)))$ the interval for δ is not empty and the upper bound for δ is chosen to ensure an important inequality, which will be used later on:

$$\tfrac{1}{4}-2\delta-\varepsilon-(mp+s)(\varepsilon+\delta)>0. \tag{29}$$

On the right-hand side of the inequality in definition (26) the term $\|L_{\tau_k}\hat{x}_{\tau_k}\|^2$ is added to ensure the existence of σ_k, while in definition (27) the first and the last inequalities are rather technical and are used in the proof of Lemma 4 for considering case (3).

The adaptive control is defined by

$$u_n^a=L_n^0\hat{x}_n+v_n \tag{30}$$

with

$$L_n^0=\begin{cases}L_1 & \text{if } n \text{ belongs to some } [\tau_k,\sigma_k),\\ 0 & \text{if } n \text{ belongs to some } [\sigma_k,\tau_{k+1}).\end{cases} \tag{31}$$

We note at once that u_n^a can be recursively computed in real time and this makes the results developed here practically applicable. It is not difficult to see that u_n^a is indeed $\mathcal{F}_n$-measurable, and it will be shown in § 5 that $\{u_n^a\}\in U$ defined by (17).

We now formulate our main result.

THEOREM 1. *If conditions* (a)-(c) *are satisfied, then the adaptive control* $u^a=\{u_n^a\}$ *given by* (30) *is optimal in the following sense: that for system* (1) *with* $\{u_n^a\}$ *applied the parameter estimate* θ_n *given by* (18) *is strongly consistent and the loss function* (2) *attains its minimum, i.e.,*

$$\theta_n \xrightarrow[n\to\infty]{} \theta \quad a.s.$$

and

$$J(u^a)=\operatorname{tr} SCQC^\tau \quad a.s.$$

The proof of Theorem 1 is given in § 5.

Obviously, the optimal adaptive control is not unique; it may differ first by a different choice of excitation source $\{v_n\}$, second by various estimation schemes applied to θ. For example, we can use the least squares algorithm. In this case, instead of (18)-(20) we take

$$\theta_{n+1}=\theta_n+a_nP_n\varphi_n(y_{n+1}^\tau-\varphi_n^\tau\theta_n), \tag{32}$$

$$P_{n+1}=P_n-a_nP_n\varphi_n\varphi_n^\tau P_n, \qquad a_n=(1+\varphi_n^\tau P_n\varphi_n)^{-1}, \tag{33}$$

$$\varphi_n^\tau=[y_n^\tau\cdots y_{n-p+1}^\tau, u_n^\tau\cdots u_{n-q+1}^\tau, y_n^\tau-\varphi_{n-1}^\tau\theta_n,\cdots, y_{n-r+1}^\tau-\varphi_{n-r}^\tau\theta_{n-r+1}], \tag{34}$$

and we change $\log^{\varepsilon/2} n$ in (25) to $n^{\varepsilon/2}$, $\log^\delta (j-1)$ in (26) to $(j-1)^\delta$ and finally $\log^\delta t$ and $\log^{\delta/2} t$ in (27) to t^δ and $t^{\delta/2}$, respectively, then Theorem 1 can be modified to the following.

THEOREM 1′. *Assume that conditions* (a) *and* (c) *are satisfied and* $C^{-1}(z)-\frac{1}{2}I$ *is strictly positive real. If the parameter estimates are given by* (32)-(34) *and in the definition of adaptive control* (25)-(31) $\log i$ *is replaced by* i *for all* i, *then*

$$\theta_n \xrightarrow[n\to\infty]{} \theta \quad and \quad J(u^a)=\operatorname{tr} SCQC^\tau \quad a.s.$$

The proof of this theorem can be carried out along the lines of that of Theorem 1.

In the sequel by θ_n we always mean the estimate given by (18)-(20).

4. Consistency theorems. In this section we give some theorems on the strong consistency of parameter estimates.

In the sequel we always denote, respectively, by $\lambda_{\max}(X)$ and $\lambda_{\min}(X)$ the maximum and the minimum eigenvalues of a matrix X. We first give a result on matrix production; it plays a crucial role in the proof of Theorem 2.

LEMMA 3. *Let* $\{f_i\}$ *be a sequence of deterministic vectors of dimension* d *and let* $F(n+1,i)$ *be recursively defined by*

$$F(n+1,i)=\left(I-\frac{f_nf_n^\tau}{r_n^f}\right)F(n,i), \qquad F(i,i)=I, \tag{35}$$

$$r_n^f=1+\sum_{i=1}^n \|f_i\|^2, \qquad r_0^f=1. \tag{36}$$

If $r_n^f \xrightarrow[m\to\infty]{} \infty$ *and for some* $a \in [0, \frac{1}{4}]$ *there are constants* N_0 *and* M *such that for all* $n \geqq N_0$

$$r_{n+1}^f / r_n^f \leqq M(\log r_n^f)^a,$$

and

$$\frac{\lambda_{\max}\left(\sum_{i=1}^n f_i f_i^\tau + \frac{1}{d} I\right)}{\lambda_{\min}\left(\sum_{i=1}^n f_i f_i^\tau + \frac{1}{d} I\right)} \leqq M(\log r_n^f)^{(1/4)-a}$$

then

$$F(n, 0) \xrightarrow[n\to\infty]{} 0.$$

The proof of Lemma 3 is given in Appendix 1.

Set

$$r_n^0 = 1 + \sum_{i=1}^n \|\varphi_i^0\|^2, \qquad r_0^0 = 1, \tag{37}$$

$$\varphi_n^{0\tau} = [y_n^\tau, \cdots, y_{n-p+1}^\tau, u_n^\tau, \cdots, u_{n-q+1}^\tau, w_n^\tau, \cdots, w_{n-r+1}^\tau], \tag{38}$$

which is obtained from φ_n with $y_i^\tau - \varphi_{i-1}^\tau \theta_{i-1}$ replaced by w_i^τ, $i = n \cdots n-r+1$.

THEOREM 2. *If condition* (b) *holds,* $r_n^0 \to \infty$ *and if there are* $a \in [0, \frac{1}{4}]$, N_0 *and* M *possibly depending upon* ω *such that for any* $n \geqq N_0 - 1$

$$r_{n+1}^0 / r_n^0 \leqq M(\log r_n^0)^a \quad a.s., \tag{39}$$

$$\frac{\lambda_{\max}\left(\sum_{i=1}^n \varphi_i^0 \varphi_i^{0\tau} + \frac{1}{d} I\right)}{\lambda_{\min}\left(\sum_{i=1}^n \varphi_i^0 \varphi_i^{0\tau} + \frac{1}{d} I\right)} \leqq M(\log r_n^0)^{1/4-a} \quad a.s., \tag{40}$$

with $d = mp + lq + mr$, *then*

$$\theta_n \xrightarrow[n\to\infty]{} \theta \quad a.s.$$

The theorem holds true if in its conditions φ_i^0 and r_i^0 are replaced by φ_i and r_i respectively.

Proof. We rewrite $F(n, i)$ defined in Lemma 3 to $\Phi(n, i)$ and $\Phi^0(n, i)$ if f_i is replaced by φ_i and φ_i^0 respectively. We know that $\Phi(n, 0) \to 0$ is equivalent to $\Phi^0(n, 0) \to 0$ if condition (b) holds (Chen and Guo (1985a), (1985b)). Then by Lemma 3 under the conditions of the theorem we have $\Phi^0(n, 0) \to 0$; hence $\theta_n \to \theta$ as shown in Chen and Guo (1985a), (1985b), (1987).

For consistency of parameter estimates we now give a theorem that translates conditions on φ_n and φ_n^0 to conditions on u_n alone. This is a basic step for proving our main result and is interesting by itself.

THEOREM 3. *Suppose that for system* (1) $A(z)$, $B(z)$ *and* $C(z)$ *have no common left factor and conditions* (a) *and* (b) *are satisfied and that*

$$u_n = u_n^s + v_n \tag{41}$$

and

$$\frac{1}{n} \sum_{i=1}^n \|u_i^s\|^2 = O(\log^\delta n) \tag{42}$$

for some δ satisfying (28), *where v_n is given by* (25) *and u_n^s is any $\mathcal{F}'_{n-1}$-measurable random vector with $\mathcal{F}'_{n-1}$ being σ-algebra generated by $\{w_i,\ i \leqq n,\ v_j,\ j \leqq n-1\}$, $\forall n \geqq 1$. Then θ_n is strongly consistent*:

$$\theta_n \xrightarrow[n\to\infty]{} \theta \quad a.s.$$

The proof is given in Appendix 2.

5. Proof of the main theorem. The proof of Theorem 1 is separated into several lemmas.

LEMMA 4. *Under conditions of Theorem* 1 *the estimate θ_n is strongly consistent*:

$$\theta_n \xrightarrow[n\to\infty]{} \theta \quad a.s.$$

and

$$L_n \xrightarrow[n\to\infty]{} L \quad a.s.,$$

where L and L_n are defined by (15) *and* (22) *respectively.*

Proof. We first prove consistency of θ_n.

(1) If $\tau_k < \infty$, $\sigma_k = \infty$ for some k, then $L_i^0 = L_i$ for $i \geqq \tau_k$ and by definition (26) for σ_k we have

$$\frac{1}{n}\sum_{i=1}^{n} \|L_i\hat{x}_i\|^2 = O(\log^\delta n).$$

Then by (30) and (31) we see that Theorem 3 can be applied, since $L_i^o\hat{x}_i$ is obviously $\mathcal{F}'_{i-1}$-measurable. Hence $\theta_n \xrightarrow[n\to\infty]{} \theta$ a.s.

(2) If $\sigma_k < \infty$, $\tau_{k+1} = \infty$ for some k, then by (30) and (31) $u_n^a = v_n$ for $n \geqq \sigma_k$, and again Theorem 3 leads to the conclusion of the lemma.

(3) If $\sigma_k < \infty$, $\tau_k < \infty$, for all k, then by (26), (27) and (31) we have for all $k \geqq 1$

$$\begin{aligned}
&\sup_{\tau_k \leqq n < \tau_{k+1}} \frac{1}{n\log^\delta n}\sum_{i=1}^{n}\|L_i^0\hat{x}_i\|^2 \\
&\quad= \sup_{\tau_k \leqq n \leqq \sigma_k - 1} \frac{1}{n\log^\delta n}\sum_{i=\tau_1}^{n}\|L_i^0\hat{x}_i\|^2 \\
&\quad= \sup_{\tau_k \leqq n \leqq \sigma_k - 1} \frac{1}{n\log^\delta n}\left[\left(\sum_{i=\tau_1}^{\sigma_1-1} + \sum_{i=\tau_2}^{\sigma_2-1} + \cdots + \sum_{i=\tau_{k-1}}^{\sigma_{k-1}-1} + \sum_{i=\tau_k}^{n}\right)\|L_i^0\hat{x}_i\|^2\right] \\
&\quad\leqq \frac{1}{\tau_2\log^\delta \tau_2}\sum_{i=\tau_1}^{\sigma_1-1}\|L_i\hat{x}_i\|^2 + \cdots + \frac{1}{\tau_k\log^\delta \tau_k}\sum_{i=\tau_{k-1}}^{\sigma_{k-1}-1}\|L_i\hat{x}_i\|^2 \\
&\qquad + \sup_{\tau_k \leqq n \leqq \sigma_k - 1}\frac{1}{n\log^\delta n}\sum_{i=\tau_k}^{n}\|L_i\hat{x}_i\|^2 \\
&\quad\leqq \sum_{i=1}^{k-1}\frac{1}{2^i} + \sup_{\tau_k \leqq n \leqq \sigma_k - 1}\frac{1}{n\log^\delta n}(n\log^\delta n + \|L_{\tau_k}\hat{x}_{\tau_k}\|^2) \leqq 3 \quad \forall k \geqq 1.
\end{aligned}$$

Hence in this case Theorem 3 can also be applied. Thus we have established the strong consistency of θ_n. The second assertion follows from Lemma 5.

In the proof of Lemmas 5, 6 and 7 we need the following fact; If matrices Ω_n converge to a stable matrix, then there are constants $0<\mu<1$ and c_2 such that (Chen (1985, p. 191))

$$\|\Omega_k\Omega_{k-1}\cdots\Omega_{i+1}\| \leqq c_2\mu^{k-i} \quad \forall k>i, \quad \forall i\geqq 0. \tag{43}$$

LEMMA 5. *If* $\theta_n \xrightarrow[n\to\infty]{} \theta$ *and condition* (c) *holds, then* S_n *defined by* (23) *tends to the solution* S *of* (14) *as* $n\to\infty$.

The proof is given in Appendix 1.

We now write x_n given by (9) and $\hat{x}_n$ given by (21) in the vector component forms

$$x_n=[x_n^{1\tau},\cdots,x_n^{s\tau}]^\tau, \qquad \hat{x}_n=[\hat{x}_n^{1\tau},\cdots,\hat{x}_n^{s\tau}]^\tau, \tag{44}$$

where x_n^i and $\hat{x}_n^i$ are m-dimensional, $i=1\cdots s$.

Set

$$z_n=[x_n^{2\tau}\cdots x_n^{s\tau}]^\tau, \qquad \hat{z}_n=[\hat{x}_n^{2\tau}\cdots \hat{x}_n^{s\tau}]^\tau. \tag{45}$$

From (21) we have

$$\begin{aligned}\hat{x}_{n+1}^1&=A_{1n}\hat{x}_n^1+\hat{x}_n^2+B_{1n}u_n+(HAx_n+HBu_n+w_{n+1}-H\hat{A}_n\hat{x}_n-H\hat{B}_nu_n)\\&=A_1x_n^1+x_n^2+B_1u_n+w_{n+1}=x_{n+1}^1.\end{aligned} \tag{46}$$

Then

$$\begin{aligned}&\hat{C}_n(y_{n+1}-H\hat{A}_n\hat{x}_n-H\hat{B}_nu_n)\\&\quad=\hat{C}_nHA(x_n-\hat{x}_n)+\hat{C}_nH(A-\hat{A}_n)\hat{x}_n+\hat{C}_nH(B-\hat{B}_n)u_n+\hat{C}_nw_{n+1}\\&\quad=\hat{C}_n^0(z_n-\hat{z}_n)+\hat{C}_nH(A-\hat{A}_n)\hat{x}_n+\hat{C}_nH(B-\hat{B}_n)u_n+\hat{C}_nw_{n+1},\end{aligned} \tag{47}$$

with

$$\hat{C}_n^0=[\underbrace{\hat{C}_n,0}_{(s-1)m}]\}sm.$$

Consequently, by taking $u_n=u_n^a$ we can write (21) in the following form:

$$\begin{aligned}\hat{x}_{n+1}&=(\hat{A}_n+\hat{B}_nL_n^0)\hat{x}_n+\hat{B}_nv_n+\hat{C}_n^0(z_n-\hat{z}_n)+\hat{C}_nH(A-\hat{A}_n)\hat{x}_n\\&\quad+\hat{C}_nH(B-\hat{B}_n)L_n^0\hat{x}_n+\hat{C}_nH(B-\hat{B}_n)v_n+\hat{C}_nw_{n+1}\\&=[\hat{A}_n+\hat{B}_nL_n^0+\hat{C}_nH(A-\hat{A}_n)+\hat{C}_nH(B-\hat{B}_n)L_n^0]\hat{x}_n\\&\quad+\hat{C}_n^0(z_n-\hat{z}_n)+[\hat{B}_n+\hat{C}_nH(B-\hat{B}_n)]v_n+\hat{C}_nw_{n+1}.\end{aligned} \tag{48}$$

From (9), (21) and (47) we obtain

$$\begin{aligned}x_{n+1}-\hat{x}_{n+1}&=A(x_n-\hat{x}_n)+(A-\hat{A}_n)\hat{x}_n+(B-\hat{B}_n)u_n^a+(C-\hat{C}_n)w_{n+1}\\&\quad-\hat{C}_n^0(z_n-\hat{z}_n)-\hat{C}_nH(A-\hat{A}_n)\hat{x}_n-\hat{C}_nH(B-\hat{B}_n)u_n^a,\end{aligned}$$

and from here and (46)

$$\begin{aligned}z_{n+1}-\hat{z}_{n+1}&=G_n(z_n-\hat{z}_n)+(A'-\hat{A}_n')\hat{x}_n+(B'-\hat{B}_n')u_n^a+(C'-\hat{C}_n')w_{n+1}\\&\quad-\hat{C}_n'H(A-\hat{A}_n)\hat{x}_n-C_n'H(B-\hat{B}_n)u_n^a\\&=G_n(z_n-\hat{z}_n)+[A'-\hat{A}_n'+(B'-\hat{B}_n')L_n^0-\hat{C}_n'H(A-\hat{A}_n)\\&\qquad\qquad-\hat{C}_n'H(B-\hat{B}_n)L_n^0]\hat{x}_n\\&\quad+[B'-\hat{B}_n'-\hat{C}_n'H(B-\hat{B}_n)]v_n+(C'-\hat{C}_n')w_{n+1},\end{aligned} \tag{49}$$

where

$$G_n = \begin{bmatrix} -\hat{C}'_n & \begin{matrix} I \\ 0 \end{matrix} \end{bmatrix} \begin{matrix} \}(s-2)m \\ \}m \end{matrix}$$

and X' denotes the matrix obtained from X by deleting its first m rows, for example, $B' = [B_2^\tau \cdots B_s^\tau]^\tau$.

Finally, (48) and (49) give us a useful representation:

$$(50)\quad \begin{pmatrix} \hat{x}_{n+1} \\ z_{n+1} - \hat{z}_{n+1} \end{pmatrix} = \Phi_n \begin{pmatrix} \hat{x}_n \\ z_n - \hat{z}_n \end{pmatrix} + \begin{pmatrix} \hat{B}_n + \hat{C}_n H(B - \hat{B}_n) \\ B' - \hat{B}'_n - \hat{C}'_n H(B - \hat{B}_n) \end{pmatrix} v_n + \begin{pmatrix} \hat{C}_n \\ C - \hat{C}_n \end{pmatrix} w_{n+1},$$

where

$$(51)\quad \Phi_n = \begin{pmatrix} \hat{A}_n + \hat{B}_n L_n^0 + \hat{C}_n H(A - \hat{A}_n) + \hat{C}_n H(B - \hat{B}_n) L_n^0 & \hat{C}_n^0 \\ A' - \hat{A}'_n + (B' - \hat{B}'_n) L_n^0 - \hat{C}'_n H(A - \hat{A}_n) - \hat{C}'_n H(B - \hat{B}_n) L_n^0 & G_n \end{pmatrix}.$$

LEMMA 6. *If conditions of Theorem* 1 *hold then there is a k such that*

$$\tau_k < \infty, \qquad \sigma_k = \infty.$$

Proof. Since $1 = \tau_1 < \sigma_1 < \tau_2 < \sigma_2 < \cdots$, we only need to prove the impossibility of the following two cases:

(1) $\sigma_k < \infty$, $\tau_{k+1} = \infty$ for some k;

(2) $\tau_k < \infty$, $\sigma_k < \infty$ for all k.

By (50) we have for $n \geqq \sigma_k$

$$(52)\quad \begin{aligned} \begin{pmatrix} \hat{x}_{n+1} \\ z_{n+1} - \hat{z}_{n+1} \end{pmatrix} &= \prod_{j=\sigma_k}^{n} \Phi_j \begin{pmatrix} \hat{x}_{\sigma_k} \\ z_{\sigma_k} - \hat{z}_{\sigma_k} \end{pmatrix} \\ &\quad + \sum_{i=\sigma_k}^{n} \prod_{j=i+1}^{n} \Phi_j \left\{ \begin{pmatrix} \hat{B}_i + \hat{C}_i H(B - \hat{B}_i) \\ B' - \hat{B}'_i - \hat{C}'_i H(B - \hat{B}_i) \end{pmatrix} v_i + \begin{pmatrix} \hat{C}_i \\ C - \hat{C}_i \end{pmatrix} w_{i+1} \right\} \end{aligned}$$

where by definition

$$\prod_{j=i+1}^{n} \Phi_j = \begin{cases} \Phi_n \cdots \Phi_{i+1} & \text{for } n > i, \\ I & \text{for } n = i. \end{cases}$$

In case (1) $L_n^0 = 0$ for $n \geqq \sigma_k$ by (32); then by Lemma 4 we have

$$(53)\quad \Phi_n \xrightarrow[n\to\infty]{} \begin{pmatrix} A & C^0 \\ 0 & G \end{pmatrix} \triangleq \Phi \quad \text{with } C^0 = [C, 0]\}sm, \qquad G = \begin{pmatrix} -C' & \begin{matrix} I \\ 0 \end{matrix} \end{pmatrix}.$$

Notice that $A(z)$ is stable by condition (a), and $C(z)$ is also stable since $C(z)$ is strictly positive real by condition (b), so Φ is a stable matrix.

From (43), (52) and Lemma 4 we obtain for all $n \geqq \sigma_k$

$$\begin{aligned} &\frac{1}{n} \sum_{i=\sigma_k}^{n} (\|\hat{x}_{i+1}\|^2 + \|z_{i+1} - \hat{z}_{i+1}\|^2) \\ &\qquad = O\left(\frac{1}{n} \sum_{i=\sigma_k}^{n} \mu^{i-\sigma_k}\right) + O\left(\frac{1}{n} \sum_{i=\sigma_k}^{n} \sum_{j=\sigma_k}^{i} \mu^{i-j} (\|w_{j+1}\|^2 + \|v_j\|^2)\right) = O(1). \end{aligned}$$

Then

$$\sum_{k=1}^{n} \|\hat{x}_k\|^2 = O(n) \quad \text{as } n \to \infty.$$

This means that τ_{k+1} must be finite by its definition (27) since $L_n \xrightarrow[n\to\infty]{} L$ by Lemma 4. Therefore case (1) cannot occur.

Now assume that $\tau_k<\infty$, $\sigma_k<\infty$ for all k. By definition τ_k is a sequence of monotonically increasing integers; then $\tau_k \xrightarrow[k\to\infty]{} \infty$.

By (31) $L_n^0=L_n$ for $n\in[\tau_k, \sigma_k)$, and then by (51) and Lemma 4 we have

$$\Phi_n \xrightarrow[\substack{n\in[\tau_k,\sigma_k)\\ k\to\infty}]{} \begin{bmatrix} A+BL & C^0 \\ 0 & G \end{bmatrix}, \tag{54}$$

where C^0 and G are defined in (53).

Since $A+BL$ is stable then for $n\in(\tau_k, \sigma_k-1]$ by (43), (54) and Lemma 4 it immediately follows from (50) that

$$\sum_{i=\tau_k}^{n} \|\hat{x}_{i+1}\|^2 \leqq c_3(\|\hat{x}_{\tau_k}\|^2+\|z_{\tau_k}-\hat{z}_{\tau_k}\|^2)+c_4\sigma_k, \tag{55}$$

where here and hereafter c_i, $i=3, 4, \cdots$, denote constants free of k.

Similarly, from (49) we know that

$$\begin{aligned} \|z_{\tau_k}-\hat{z}_{\tau_k}\|^2 &= O(1)+O\left(\sum_{i=0}^{\tau_k} \|\hat{x}_i\|^2\right)+\left(\sum_{i=0}^{\tau_k} (\|w_{i+1}\|^2+\|v_i\|^2\right) \\ &\leqq c_5\tau_k+c_6\tau_k \log^{\delta/2} \tau_k, \end{aligned} \tag{56}$$

where for the last inequality (27) is invoked.

Putting (56) into (55) and noticing the boundedness of L_i, we conclude that for sufficiently large k

$$\sum_{i=\tau_k}^{\sigma_k} \|L_i\hat{x}_i\|^2 \leqq c_7\tau_k \log^{\delta/2} \tau_k + c_8\sigma_k \leqq c_9\sigma_k \log^{\delta/2} \sigma_k < \sigma_k \log^{\delta} \sigma_k + \|L_{\tau_k}\hat{x}_{\tau_k}\|^2.$$

On the other hand, by definition (26) we have the converse inequality

$$\sum_{i=\tau_k}^{\sigma_k} \|L_i\hat{x}_i\|^2 > \sigma_k \log^{\delta} \sigma_k + \|L_{\tau_k}\hat{x}_{\tau_k}\|^2$$

since $\sigma_k<\infty$.

The obtained contradiction shows that case (2) cannot take place as well.

To finish the proof of Theorem 1 it remains to show that the loss function reaches its minimum when u_n^a given by (30) is applied. It is done in the next lemma.

LEMMA 7. *If conditions of Theorem* 1 *hold, then* $\{u_n^a\}\in U$ *defined by* (17) *and*

$$J(u_n^a)=\operatorname{tr} SCQC^{\tau}.$$

Proof. By Lemma 6 and (31) there exists some k_0 such that

$$L_n^0=L_n \quad \forall n \geqq \tau_{k_0}.$$

By Lemma 4 we know that $\{\Phi_n\}$ converges to the matrix stated at the right-hand side of (54). Then by (43) from (50) it is easy to see that

$$\|\hat{x}_{k+1}\|^2+\|z_{k+1}-\hat{z}_{k+1}\|^2 = O(1)+O\left(\sum_{i=1}^{k} \mu^{k-i}(\|w_{i+1}\|^2+\|v_i\|^2)\right), \tag{57}$$

and

$$\frac{1}{n}\sum_{k=1}^{n}(\|\hat{x}_{k+1}\|^2+\|z_{k+1}-\hat{z}_{k+1}\|^2) = O\left(\frac{1}{n}\sum_{k=1}^{n}\mu^k\right)+O\left(\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{k}\mu^{k-i}(\|w_{i+1}\|^2+\|v_i\|^2)\right)=O(1). \tag{58}$$

Then by (A2), (25), (57) and (58) it follows that

$$\|\hat{x}_n\|^2=o(n) \quad \text{and} \quad \sum_{k=1}^{n}\|\hat{x}_k\|^2=O(n) \quad \text{a.s.} \tag{59}$$

Hence $\sum_{i=1}^{n}\|u_i^a\|^2=O(n)$, $\|u_n^a\|^2=o(n)$, and thus $\{u^a\}\in U$.

Using (59) and the consistency of θ_n and noticing that G_n in (49) converges to a stable matrix, then from (49) we are easily convinced of

$$\frac{1}{n}\sum_{k=1}^{n}\|z_{k+1}-\hat{z}_{k+1}\|^2=o(1) \quad \text{a.s.}, \tag{60}$$

which together with (46) yields

$$\frac{1}{n}\sum_{i=1}^{n}\|x_i-\hat{x}_i\|^2 \xrightarrow[n\to\infty]{} 0. \tag{61}$$

From (59) and (61) we see

$$\frac{1}{n}\sum_{i=1}^{n}\|x_i\|^2=O(1). \tag{62}$$

Finally, putting u_n^a into (16) and using (25), (61), (62) and the fact $L_n^0 \xrightarrow[n\to\infty]{} L$ we conclude that

$$\begin{aligned} J(u^a) &= \operatorname{tr} SCQC^\tau + \overline{\lim_{n\to\infty}}\frac{1}{n}\sum_{i=0}^{n-1}[(L_i^0-L)x_i+L_i^0(\hat{x}_i-x_i)+v_i]^\tau(Q_2+B^\tau SB) \\ &\quad \cdot[(L_i^0-L)x_i+L_i^0(\hat{x}_i-x_i)+v_i] \\ &= \operatorname{tr} SCQC^\tau. \end{aligned}$$

This completes the proof for Lemma 7 as well as for Theorem 1.

6. Conclusion remark. 1) In Chen and Guo (1987) the authors have given the optimal stochastic control minimizing the tracking error and leading to consistency of estimates given by the stochastic gradient algorithm. It is natural to ask: Is it possible to give a unified adaptive control applicable to both problems of tracking and quadratic cost. This requires further consideration.

2) The stability condition on $A(z)$ is rather restrictive. It is desirable to weaken it.

Appendix 1.

Proof of Lemma 1. By a standard treatment (see e.g. Chen (1985)) from (9), (10), (14) and (15) we have

$$\begin{aligned} \sum_{i=0}^{n-1}(y_i^\tau Q_1 y_i+u_i^\tau Q_2 u_i) &= x_0^\tau S x_0 - x_n^\tau S x_n + \sum_{i=0}^{n-1} w_{i+1}^\tau C^\tau S C w_{i+1} \\ &\quad +2\sum_{i=0}^{n-1}(Ax_i+Bu_i)^\tau S C w_{i+1} \\ &\quad +\sum_{i=0}^{n-1}(u_i-Lx_i)^\tau(Q_2+B^\tau SB)(u_i-Lx_i). \end{aligned} \tag{A.1}$$

From (7) it is clear that

$$\text{(A.2)}\qquad \frac{\|w_n\|^2}{n}=\operatorname{tr}\frac{w_n w_n^\tau}{n}=\operatorname{tr}\left(\frac{\sum_{i=1}^n w_i w_i^\tau}{n}-\frac{\sum_{i=1}^{n-1} w_i w_i^\tau}{n}\right)\xrightarrow[n\to\infty]{} 0.$$

By stability of A there are constants c_0 and $\rho\in(0,1)$ such that

$$\text{(A.3)}\qquad \|A^k\|\leqq c_0\rho^k \quad \forall k\geqq 0.$$

Then by (9) and (A.3) it follows that

$$\|x_n\|^2\leqq 3c_0^2\rho^{2n}\|x_0\|^2+3c_0^2\frac{\|B\|^2+\|C\|^2}{1-\rho}\sum_{j=0}^{n-1}\rho^{n-j}(\|u_j\|^2+\|w_j\|^2).$$

Therefore by (7), (17) and (A.2) from here it is concluded that

$$\text{(A.4)}\qquad \frac{\|x_n\|^2}{n}=o(1),\qquad \sum_{i=1}^n\|x_i\|^2=O(n)\quad \text{a.s.}$$

From (17), (A.1) and (A.4) the conclusion of the lemma will follow immediately if we can show that

$$\sum_{i=0}^{n-1}(Ax_i+Bu_i)^\tau SCw_{i+1}=O\left(\left[\sum_{i=1}^{n-1}(\|x_i\|^2+\|u_i\|^2)\right]^{1/2}\log^{1/2+\eta}\left(\sum_{i=1}^{n-1}(\|x_i\|^2+\|u_i\|^2)+e\right)\right)$$

$$\forall\eta>0.$$

But this is a direct consequence of Lemma 2.

Proof of Lemma 2. By the martingale convergence theorem (Chow (1965)) $\sum_{i=1}^n f_i w_{i+1}^\tau$ is convergent on the set $V=\{\omega: s_\infty<\infty\}$; hence Lemma 2 obviously holds on V.

Further, for $\omega\in V^c$, without loss of generality we assume $f_1\neq 0$; then we have

$$\begin{aligned}\sum_{i=2}^\infty E\left[\left\|\frac{f_i w_{i+1}^\tau}{s_i^{1/2}\log^{1/2+\eta}(s_i+e)}\right\|^2\middle|\mathcal{F}_i\right]&\leqq\sigma^2\sum_{i=2}^\infty\frac{\|f_i\|^2}{s_i\log^{1+2\eta}(s_i+e)}\\ &\leqq\sigma^2\sum_{i=2}^\infty\left(\int_{s_{i-1}}^{s_i}dx\right)\Big/ s_i\log^{1+2\eta}(s_i+e)\\ &\leqq\sigma^2\sum_{i=2}^\infty\left(\int_{s_{i-1}}^{s_i}\frac{dx}{x\log^{1+2\eta}(x+e)}\right)\\ &=\sigma^2\int_{s_1}^\infty\frac{dx}{x\log^{1+2\eta}(x+e)}\\ &<\infty,\end{aligned}$$

where $\sigma^2=\sup_n E[\|w_{n+1}\|^2|\mathcal{F}_n]$ by definition. Again by the martingale convergence theorem we see that

$$\sum_{i=2}^\infty f_i w_{i+1}^\tau/s_i^{1/2}\log^{1/2+\eta}(s_i+e)$$

is convergent on V^c. Then the Kronecker Lemma guarantees validity of Lemma 2 on V^c.

Proof of Lemma 3. Set

$$\text{(A.5)}\qquad m(t)=\max[n: t_n\leqq t],$$

$$\text{(A.6)}\qquad t_n\triangleq\sum_{i=N_0}^{n-1}\frac{\|f_i\|^2}{r_i^f(\log r_{i-1}^f)^{1/4}}.$$

We note that $m(t)$ is nothing but the inverse function of t_n, which is defined such that it diverges in an appropriate rate.

It is easy to see that

$$t_n \geqq \frac{1}{M}\sum_{i=N_0}^{n-1}\frac{\|f_i\|^2}{r_{i-1}^f(\log r_{i-1}^f)^{1/4+a}} \geqq \frac{1}{M}\sum_{i=N_0}^{n-1}\int_{r_{i-1}^f}^{r_i^f}\frac{dt}{t(\log t)^{1/4+a}}$$
$$=\frac{4}{(3-4a)M}(\log^{3/4-a} r_{n-1}^f - \log^{3/4-a} r_{N_0-1}^f),$$

which via (A.5) implies $t_n \to \infty$, $m(t) < \infty$ for all t, and

$$\text{(A.7)} \quad \log r_{m(N+k\alpha)-1}^f \leqq \left[\frac{(3-4a)M}{4}(N+k\alpha)+\log^{3/4-a} r_{N_0-1}^f\right]^{4/(3-4a)} \quad \forall N \geqq 1.$$

For sufficiently large N_0 we have

$$\text{(A.8)} \quad \log r_i^f \leqq \log r_{i-1}^f + \log M + a \log\log r_{i-1}^f \leqq 2\log r_{i-1}^f \quad \forall i \geqq N_0.$$

then

$$t_n \leqq 2\sum_{i=N_0}^{n-1}\frac{\|f_i\|^2}{r_i^f(\log r_i^f)^{1/4}} \leqq \tfrac{8}{3}(\log^{3/4} r_{n-1}^f - \log^{3/4} r_{N_0-1}^f)$$

and hence

$$t \leqq t_{m(t)+1} \leqq \tfrac{8}{3}(\log^{3/4} r_{m(t)}^f - \log^{3/4} r_{N_0-1}^f)$$

or

$$\text{(A.9)} \quad \log^a r_{m(N+(k-1)\alpha)}^f \geqq [\tfrac{3}{8}(N+(k-1)\alpha)+\log^{3/4} r_{N_0-1}^f]^{4a/3}.$$

Since $m(t) < \infty$ for all t, there exists N such that $m(N) \geqq N_0$ and

$$\text{(A.10)} \quad \frac{(\log r_i^f)^{1/4-a}}{r_i^f} \leqq \frac{1}{2M} \quad \forall i \geqq m(N).$$

For any $k \geqq 1$ by summation by parts and using (A.9) we obtain

$$\sum_{i=m(N+(k-1)\alpha)}^{m(N+k\alpha)-1}\frac{f_i f_i^\tau}{r_i^f} \geqq \sum_{i=m(N+(k-1)\alpha)}^{m(N+k\alpha)}\frac{1}{r_i^f}\left(\sum_{j=1}^{i} f_j f_j^\tau - \sum_{j=1}^{i-1} f_j f_j^\tau\right) - I$$
$$\geqq \sum_{i=m(N+(k-1)\alpha)+1}^{m(N+k\alpha)}\sum_{j=1}^{i-1} f_j f_j^\tau \frac{\|f_i\|^2}{r_{i-1}^f r_i^f} - 2I$$
$$\geqq \sum_{i=m(N+(k-1)\alpha)+1}^{m(N+k\alpha)}\left[\frac{\lambda_{\max}(\sum_{j=1}^{i-1} f_j f_j^\tau + (1/d)I)}{M(\log r_{i-1}^f)^{1/4-a}} - \frac{1}{d}\right]\frac{\|f_i\|^2}{r_i^f r_{i-1}^f} I - 2I$$
$$\geqq \frac{1}{d}(\log r_{m(N+(k-1)\alpha)}^f)^a \cdot \sum_{i=m(N+(k-1)\alpha)+1}^{m(N+k\alpha)}$$
$$\cdot\left(\frac{1}{M} - \frac{(\log r_{i-1}^f)^{1/4-a}}{r_{i-1}^f}\right)\frac{\|f_i\|^2}{r_i^f(\log r_{i-1}^f)^{1/4}} I - 2I$$
$$\geqq \frac{1}{2Md}\log^a r_{m(N+(k-1)\alpha)}^f (t_{m(N+k\alpha)+1} - t_{m(N+(k-1)\alpha)+1}) I - 2I$$
$$\geqq \left[\frac{\alpha-1}{2Md}\log^a r_{m(N+(k-1)\alpha)}^f - 2\right] I$$
$$\geqq \left[\frac{\alpha-1}{2Md}\left(\frac{3}{8}(N-\alpha)+\frac{3\alpha}{8}k+\log^{3/4} r_{N-1}^f\right)^{4a/3} - 2\right] I.$$

We take N, α large enough so that $N>\alpha$ and

$$b=\frac{\alpha-1}{2Md}\left(\frac{3\alpha}{8}\right)^{4a/3}-2>0;$$

then

$$\sum_{i=m(N+(k-1)\alpha)}^{m(N+k\alpha)-1}\frac{f_if_i^\tau}{r_i^f}\geqq b(k^{4a/3})I \quad \forall k\geqq 1. \tag{A.11}$$

Let ρ_k be the maximum eigenvalue of the matrix

$$F^\tau(m(N+k\alpha), m(N+(k-1)\alpha))F(m(N+k\alpha), m(N+(k-1)\alpha))$$

and let $x_{m(N+(k-1)\alpha)}$ be the corresponding normalized eigenvector. For $i\in[m(N+(k-1)\alpha), m(N+k\alpha)-1]$ recursively define x_i

$$x_{i+1}=\left(I-\frac{f_if_i^\tau}{r_i^f}\right)x_i. \tag{A.12}$$

Then we have

$$\begin{aligned} x^\tau_{m(N+k\alpha)}x_{m(N+k\alpha)}&=x^\tau_{m(N+(k-1)\alpha)}F^\tau(m(N+k\alpha), m(N+(k-1)\alpha))\\ &\quad\cdot F(m(N+k\alpha), m(N+(k-1)\alpha))x_{m(N+(k-1)\alpha)}\\ &=x^\tau_{m(N+(k-1)\alpha)}\rho_k x_{m(N+(k-1)\alpha)} \end{aligned} \tag{A.13}$$

and

$$x^\tau_{i+1}x_{i+1}\leqq x_i^\tau x_i-x_i^\tau\frac{f_if_i^\tau}{r_i^f}x_i. \tag{A.14}$$

Summing up both sides of (A.14) we obtain that

$$\sum_{i=m(N+(k-1)\alpha)}^{m(N+k\alpha)-1}\frac{\|f_i^\tau x_i\|^2}{r_i^f}\leqq\|x_{m(N+(k-1)\alpha)}\|^2-\|x_{m(N+k\alpha)}\|^2=1-\rho_k. \tag{A.15}$$

For $i\in[m(N+(k-1)\alpha), m(N+k\alpha)-1]$ from (A.12) by Schwarz inequality and (A.6), (A.15) we see that

$$\begin{aligned} \|x_i-x_{m(N+(k-1)\alpha)}\|&=\left\|\sum_{j=m(N+k(-1)\alpha)}^{i-1}\frac{f_jf_j^\tau}{r_j^f}x_j\right\|\\ &\leqq\{\log r^f_{m(N+k\alpha)-1}\}^{1/8}\sum_{j=m(N+(k-1)\alpha)}^{m(N+k\alpha)-1}\frac{\|f_j\|}{(r_j^f)^{1/2}(\log r_{j-1}^f)^{1/8}}\cdot\frac{\|f_j^\tau x_j\|}{(r_j^f)^{1/2}}\\ &\leqq\{\log r^f_{m(N+k\alpha)-1}\}^{1/8}\sqrt{1+\alpha}\cdot\sqrt{1-\rho_k}. \end{aligned} \tag{A.16}$$

Finally, by (A.7), (A.11), (A.15) and (A.16) we conclude that

$$\begin{aligned} bk^{4a/3}&\leqq x^\tau_{m(N+(k-1)\alpha)}\sum_{i=m(N+(k-1)\alpha)}^{m(N+k\alpha)-1}\frac{f_if_i^\tau}{r_i^f}(x_{m(N+(k-1)\alpha)}-x_i+x_i)\\ &\leqq(\log r^f_{m(N+k\alpha)-1})^{1/4}\sum_{i=m(N+(k-1)\alpha)}^{m(N+k\alpha)-1}\frac{\|f_i\|^2}{r_i^f(\log r_{i-1}^f)^{1/4}}\|x_{m(N+(k-1)\alpha)}-x_i\|\\ &\quad+\{\log r^f_{m(N+k\alpha)-1}\}^{1/8}\sum_{i=m(N+(k-1)\alpha)}^{m(N+k\alpha)-1}\frac{\|f_i\|}{(r_i^f)^{1/2}(\log r_{i-1}^f)^{1/8}}\cdot\frac{\|f_i^\tau x_i\|}{(r_i^f)^{1/2}}\\ &\leqq\{(\log r^f_{m(N+k\alpha)-1})^{3/8}(\alpha+1)^{3/2}+(\log r^f_{m(N+k\alpha)-1})^{1/8}(\alpha+1)^{1/2}\}\sqrt{1-\rho_k} \end{aligned}$$

$$\leqq \left\{ (\alpha+1)^{3/2} + (\alpha+1)^{1/2} \left[\frac{(3-4a)M}{4}(N+k\alpha) + \log^{3/4-a} r^f_{N_0-1} \right]^{-1/(3-4a)} \right\}$$
$$\times \left\{ \frac{(3-4a)M}{4}(N+k\alpha) + \log^{3/4-a} r^f_{N_0-1} \right\}^{3/2(3-4a)} \cdot \sqrt{1-\rho_k}.$$

It is clear that there is a constant $c_1 > 0$ such that

$$bk^{4a/3} \leqq c_1 k^{3/2(3-4a)}(1-\rho_k)^{1/2} \quad \forall k \geqq 1,$$

or

$$\rho_k \leqq 1 - \frac{b^2}{c_1^2} \cdot \frac{1}{k^{3/(3-4a)-(8a/3)}}.$$

Then

$$\|F(m(N+k\alpha), 0)\| \leqq \prod_{i=1}^{k} \|F(m(N+i\alpha), m(N+(i-1)\alpha))\| \cdot \|F(m(N), 0)\|$$
$$\leqq \prod_{i=1}^{k} \sqrt{\rho_i} \xrightarrow[k\to\infty]{} 0$$

since

$$\frac{5}{6} \leqq \frac{3}{3-4a} - \frac{8}{3}a \leqq 1 \quad \text{for } a \in [0, \tfrac{1}{4}].$$

Notice that $\|F(n, 0)\|$ is nonincreasing; then the lemma follows immediately.

Proof of Lemma 5. For simplicity we denote by $P(A, B, S)$ the right-hand side of (14). By Theorem 14.3 of Lipster and Shiryayev (1978) equation (14) can be solved recursively

$$\Gamma_{n+1} = P(A, B, \Gamma_n) \tag{A.17}$$

and $\Gamma_n \to S$ for any $\Gamma_0 \geqq 0$. Γ_n with initial value $\Gamma_0 = 0$ is denoted by Γ_n^0. In this theorem it is proved that for any vector x of compatible dimension

$$x^\tau \Gamma_n^0 x \leqq x^\tau \Gamma_n x \leqq x^\tau S x + \bar{x}_n^\tau (\Gamma_0 - S)\bar{x}_n, \tag{A.18}$$

or equivalently,

$$x^\tau(\Gamma_n^0 - S)x \leqq x^\tau(\Gamma_n - S)x \leqq \bar{x}_n^\tau(\Gamma_0 - S)\bar{x}_n,$$

where $\bar{x}_n \xrightarrow[n\to\infty]{} 0$ and $\Gamma_n^0 \to S$ and both $\bar{x}_n$ and Γ_n^0 are independent of Γ_0. Hence from (A.18) we see that the convergence $\Gamma_n \to S$ is uniform in Γ_0 for $\|\Gamma_0\| \leqq c$ with c being any fixed constant.

From (23) we know that

$$S_n \leqq \hat{A}_n^\tau S_{n-1} \hat{A}_n + H^\tau Q_1 H \quad \forall n \geqq 1.$$

Then, taking into account (43) we have the boundedness of S_n:

$$\begin{aligned}
\|S_n\| &\leqq \|\hat{A}_n^\tau S_{n-1}\hat{A}_n + H^\tau Q_1 H\| \leqq \cdots \\
&\leqq \left\| \sum_{i=2}^{n} (\hat{A}_i \hat{A}_{i+1} \cdots \hat{A}_n)^\tau H^\tau Q_1 H (\hat{A}_i \hat{A}_{i+1} \cdots \hat{A}_n) \right. \\
&\qquad \left. + (\hat{A}_1 \cdots \hat{A}_n)^\tau S_0 (\hat{A}_1 \cdots \hat{A}_n) + H^\tau Q_1 H \right\| \\
&\leqq \|H^\tau Q_1 H\| + c_2^2(\|H^\tau Q_1 H\| + \|S_0\|) \frac{1}{1-\mu} \triangleq c \quad \forall n \geqq 1.
\end{aligned} \tag{A.19}$$

By strong consistency of θ_n and by boundedness of S_n it is easy to see

$$P(A, B, S_n) - P(\hat{A}_{n+1}, \hat{B}_{n+1}, S_n) \xrightarrow[n\to\infty]{} 0.$$

Hence for any $\varepsilon > 0$ we can find $N > 0$ such that

$$\|\Delta S_{n+k}\| \leqq \varepsilon \quad \forall k \geqq 0, \quad \forall n \geqq N, \tag{A.20}$$

where

$$\Delta S_{n+k} = S_{n+k} - P(A, B, S_{n+k-1}). \tag{A.21}$$

For simplicity we set

$$P_1(\Gamma) \triangleq P(A, B, \Gamma), \qquad P_n(\Gamma) \triangleq P_1(P_{n-1}(\Gamma)).$$

It is easy to show that there is a constant ζ such that

$$P_1(\Gamma + \Delta\Gamma) = P_1(\Gamma) + \overline{\Delta\Gamma}, \qquad \|\overline{\Delta\Gamma}\| \leqq \zeta\varepsilon \tag{A.22}$$

for matrices $\Gamma \geqq 0$ with $\|\Gamma\| \leqq c$ and $\Delta\Gamma$ with $\|\Delta\Gamma\| \leqq \varepsilon$.

We now by induction prove that for any $n \geqq N$ and $k \geqq 1$

$$S_{n+k} = P_k(S_n) + Z_{nk}(\varepsilon) \quad \text{with } \|Z_{nk}(\varepsilon)\| \leqq c_k\varepsilon, \tag{A.23}$$

where c_k is a real number independent of n.

By (A.20), (A.21), we see that (A.23) is true for $k = 1$. Now assume (A.23) holds for k. By boundedness of $\|S_n\| \leqq c$ for all n, the same argument as that used in (A.19) leads to the conclusion that $P_k(S_n)$ is uniformly bounded in $n \geqq 0$ and $k \geqq 1$. Then by (A.21)-(A.23) it follows that

$$\begin{aligned} S_{n+k+1} &= P_1(S_{n+k}) + \Delta S_{n+k+1} = P_1(P_k(S_n) + Z_{nk}(\varepsilon)) + \Delta S_{n+k+1} \\ &= P_{k+1}(S_n) + \overline{Z_{nk}}(\varepsilon) + \Delta S_{n+k+1} = P_{k+1}(S_n) + Z_{nk+1}(\varepsilon), \end{aligned}$$

where, obviously, $\|Z_{nk+1}(\varepsilon)\| \leqq c_{k+1} \cdot \varepsilon$ with $c_{k+1} = \zeta c_k + 1$. Hence (A.23) holds for $k+1$.

In the present notation

$$\Gamma_n = P_n(\Gamma_0)$$

where Γ_n is defined by (A.17). By the uniform convergence of Γ_n for any $\delta > 0$ we can take k_0 large enough such that

$$\|P_{k_0}(\Gamma_0) - S\| \leqq \delta \quad \forall \Gamma_0 : \|\Gamma_0\| \leqq c. \tag{A.24}$$

For $\varepsilon \triangleq \delta / c_{k_0}$ take N such that

$$\|\Delta S_{n+k_0}\| \leqq \varepsilon \quad \forall n \geqq N. \tag{A.25}$$

Then from (A.23) we have

$$S_{n+k_0} = P_{k_0}(S_n) + Z_{nk_0}(\varepsilon), \qquad \|Z_{nk_0}(\varepsilon)\| \leqq c_{k_0}\varepsilon = \delta$$

and by (A.24) for all $n \geqq N$

$$\|S_{n+k_0} - S\| \leqq \|P_{k_0}(S_n) - S\| + \|Z_{nk_0}(\varepsilon)\| \leqq 2\delta,$$

which yields the conclusion of the lemma.

Appendix 2.

Proof of Theorem 3. First we note that $\{v_n, \mathcal{F}'_n\}$ is a martingale difference sequence. Then by Lemma 2 we have

$$\sum_{i=1}^{n} u_i^s v_i^\tau = O\left(\left(\sum_{i=1}^{n} \|u_i^s\|^2\right)^{1/2} \log^{1/2+\eta}\left(\sum_{i=1}^{n} \|u_i^s\|^2 + e\right)\right). \tag{A.26}$$

Further, by (24), (25) we know that for $\gamma \in (\frac{2}{3}, 1)$

$$\sum_{i=1}^{\infty} E\left[\frac{\|v_i v_i^\tau - I/\log^\varepsilon i\|^{3/2}}{i^{3\gamma/2}} \,\Big|\, \mathcal{F}'_{i-1}\right] < \infty.$$

Hence $\sum_{i=1}^{\infty} [v_i v_i^\tau - (1/\log^\varepsilon i)I]/i^\gamma$ is convergent by the martingale convergence theorem. Then from the Kronecker lemma it follows that

$$\lim_{n\to\infty} \frac{1}{n^\gamma} \sum_{i=1}^{n} \left(v_i v_i^\tau - \frac{1}{\log^\varepsilon i} I\right) = 0 \quad \forall \gamma \in (\tfrac{2}{3}, 1). \tag{A.27}$$

It is clear that

$$\int_2^{n+1} \frac{dx}{\log^\varepsilon x} \leqq \sum_{i=2}^{n} \frac{1}{\log^\varepsilon i} \leqq \frac{1}{\log^\varepsilon 2} + \int_2^n \frac{dx}{\log^\varepsilon x}$$

and the l'Hôpital rule shows

$$\frac{\log^\varepsilon n}{n} \sum_{i=2}^{n} \frac{1}{\log^\varepsilon i} \xrightarrow[n\to\infty]{} 1; \tag{A.28}$$

hence by (A.27)

$$\frac{\log^\varepsilon n}{n} \sum_{i=1}^{n} v_i v_i^\tau \xrightarrow[n\to\infty]{} I \quad \text{a.s.} \tag{A.29}$$

From (42), (A.26) and (A.29) we see

$$\frac{1}{n} \sum_{i=1}^{n} \|u_i\|^2 = O(\log^\delta n). \tag{A.30}$$

Then by condition (a)

$$\frac{1}{n} \sum_{i=1}^{n} \|y_i\|^2 = O(\log^\delta n); \tag{A.31}$$

hence

$$r_n^0 = O(n \log^\delta n), \tag{A.32}$$

which means

$$\lambda_{\max}\left(\sum_{i=1}^{n} \varphi_i^0 \varphi_i^{0\tau} + \frac{1}{d} I\right) = O(n \log^\delta n). \tag{A.33}$$

Again by (41), (42), (A.26), (A.29), and noting that $q \geqq 1$, we have for all sufficiently large n

$$r_n^0 \geqq \sum_{i=1}^{n} \|u_i\|^2 \geqq \tfrac{1}{2} \sum_{i=1}^{n} \|v_i\|^2 \geqq \frac{l}{4} \frac{n}{\log^\varepsilon n}. \tag{A.34}$$

Then $r_n^0 \xrightarrow[n\to\infty]{} \infty$ a.s. and

$$\frac{r_{n+1}^0}{r_n^0} = O\left(\frac{(n+1)\log^\delta(n+1)}{n/\log^\varepsilon n}\right) = O(\log^{\delta+\varepsilon} n) = O((\log r_n^0)^{\delta+\varepsilon}).$$

Comparing with conditions in Theorem 2 we find that $a=\delta+\varepsilon$ and by (A.33) and (A.34) for (40) to hold we only need to verify

$$\lim_{n\to\infty}\frac{(\log n)^{1/4-2\delta-\varepsilon}}{n}\lambda_{\min}\left(\sum_{i=1}^{n}\varphi_i^0\varphi_i^{0\tau}+\frac{1}{d}I\right)\neq 0. \tag{A.35}$$

By condition (a) it is easy to see that

$$\begin{aligned}y_{n-i}&=A^{-1}(z)B(z)z^iu_n+A^{-1}(z)C(z)z^iw_n\\&=z^iA^{-1}(z)[B(z),\,C(z)]\cdot\begin{bmatrix}u_n\\w_n\end{bmatrix}.\end{aligned}$$

Then φ_n^0 can be written as

$$\varphi_n^0=\begin{bmatrix}F_{n1}(z)\\F_{n2}(z)\\F_{n3}(z)\end{bmatrix}\cdot\begin{bmatrix}u_n\\w_n\end{bmatrix}, \tag{A.36}$$

where by definition

$$F_{n1}(z)=\begin{bmatrix}A^{-1}(z)[B(z),\,C(z)]\\zA^{-1}(z)[B(z),\,C(z)]\\\vdots\\z^{p-1}A^{-1}(z)[B(z),\,C(z)]\end{bmatrix},\qquad F_{n2}(z)=\begin{bmatrix}[I_l,0]\\z[I_l,0]\\\vdots\\z^{q-1}[I_l,0]\end{bmatrix},$$

$$F_{n3}(z)=\begin{bmatrix}[0,\,I_m]\\z[0,\,I_m]\\\vdots\\z^{r-1}[0,\,I_m]\end{bmatrix},$$

where I_x denotes the identity matrix of dimension x.

Set

$$\psi_n=[\det A(z)]\varphi_n^0 \tag{A.37}$$

and notice that A_p is of full rank, then $\deg A(z)=p$, $\deg[\det A(z)]=mp$, and $\deg[\operatorname{Adj} A(z)]=mp-p$, since $A(z)[\operatorname{Adj} A(z)]=[\det A(z)]\cdot I$.

Let

$$\det A(z)=a_0+a_1z+\cdots+a_{mp}z^{mp}.$$

Since $\varphi_i^0=0$ for $i<0$ we have

$$\begin{aligned}\lambda_{\min}\left(\sum_{i=1}^{n}\psi_i\psi_i^\tau\right)&=\inf_{\|x\|=1}\sum_{i=1}^{n}(x^\tau\psi_i)^2\\&=\inf_{\|x\|=1}\sum_{i=1}^{n}\left(\sum_{j=0}^{mp}a_jx^\tau\varphi_{i-j}^0\right)^2\\&\leqq\sum_{j=0}^{mp}a_j^2\inf_{\|x\|=1}\sum_{i=1}^{n}\sum_{j=0}^{mp}(x^\tau\varphi_{i-j}^0)^2\\&\leqq(mp+1)\sum_{j=0}^{mp}a_j^2\inf_{\|x\|=1}\sum_{i=1}^{n}(x^\tau\varphi_i^0)^2\\&=(mp+1)\sum_{j=0}^{mp}a_j^2\lambda_{\min}\left(\sum_{i=1}^{n}\varphi_i^0\varphi_i^{0\tau}\right).\end{aligned}$$

Hence for (A.35) it is sufficient to prove

$$\varliminf_{n\to\infty} \frac{(\log n)^\lambda}{n} \lambda_{\min}\left(\sum_{i=1}^{n} \psi_i\psi_i^\tau\right) \neq 0 \quad \text{a.s.,} \tag{A.38}$$

where for simplicity we set $\lambda = \frac{1}{4} - 2\delta - \varepsilon$.

Let D be the set on which (A.38) is not satisfied. Suppose that $P(D) > 0$. Then for any $\omega \in D$ there exist vectors

$$\eta_{n_k} = (\alpha_{n_k}^{0\tau} \cdots \alpha_{n_k}^{(p-1)\tau} \beta_{n_k}^{0\tau} \cdots \beta_{n_k}^{(q-1)\tau} \gamma_{n_k}^{0\tau} \cdots \gamma_{n_k}^{(\gamma-1)\tau})^\tau \in R^d,$$

where $\|\eta_{n_k}\| = 1$ such that

$$\frac{(\log n_k)^\lambda}{n_k} \sum_{i=1}^{n_k} (\eta_{n_k}^\tau \psi_i)^2 \xrightarrow[k\to\infty]{} 0. \tag{A.39}$$

Set

$$\begin{aligned} H_{n_k}(z) &\triangleq \sum_{i=0}^{p-1} \alpha_{n_k}^{i\tau} z^i (\operatorname{Adj} A(z))[B(z), C(z)] + \sum_{i=0}^{q-1} \beta_{n_k}^{i\tau} z^i [\det A(z) I_l, 0] \\ &\quad + \sum_{i=0}^{r-1} \gamma_{n_k}^{i\tau} z^i [0, \det A(z) I_m] \end{aligned} \tag{A.40}$$

$$\triangleq \sum_{i=0}^{t} [h_{n_k}^{i\tau}, g_{n_k}^{i\tau}] z^i, \tag{A.41}$$

where $t = mp + s - 1$, and $h_{n_k}^i$ and $g_{n_k}^i$ are l- and m-dimensional vectors, respectively.

Since $\|\alpha_{n_k}^i\| \leqq 1$, $\|\beta_{n_k}^j\| \leqq 1$, $\|\gamma_{n_k}^\nu\| \leqq 1$, for any $k \geqq 1$, $i = 0 \cdots p-1$, $j = 0 \cdots q-1$, and $\nu = 0 \cdots r-1$, there exists a constant $c_1 > 0$ independent of k and i such that

$$\|h_{n_k}^i\| \leqq c_1, \quad \|g_{n_k}^i\| \leqq c_1 \quad \forall k \geqq 1, \quad i = 0, \cdots, t. \tag{A.42}$$

By (A.36), (A.37) and (A.41) we can rewrite (A.39) as

$$\frac{(\log n_k)^\lambda}{n_k} \sum_{i=1}^{n_k} (h_{n_k}^{0\tau} u_i + \cdots + h_{n_k}^{t\tau} u_{i-t} + g_{n_k}^{0\tau} w_i + \cdots + g_{n_k}^{t\tau} w_{i-t})^2 \xrightarrow[k\to\infty]{} 0, \tag{A.43}$$

or equivalently,

$$\begin{aligned} \frac{(\log n_k)^\lambda}{n_k} \Bigg\{ &\sum_{i=1}^{n_k} [(h_{n_k}^{0\tau} v_i)^2 + (h_{n_k}^{0\tau} u_i^s + h_{n_k}^{1\tau} u_{i-1} + \cdots + h_{n_k}^{t\tau} u_{i-t} + g_{n_k}^{0\tau} w_i + \cdots + g_{n_k}^{t\tau} w_{i-t})^2] \\ &+ 2h_{n_k}^{0\tau}\left(\sum_{i=1}^{n_k} u_i^s v_i^\tau\right) h_{n_k}^0 + 2\sum_{j=1}^{n_k} h_{n_k}^{j\tau}\left(\sum_{i=1}^{n_k} u_{i-j} v_i^\tau\right) h_{n_k}^0 \\ &+ 2\sum_{j=0}^{t} g_{n_k}^{j\tau}\left(\sum_{i=1}^{n_k} w_{i-j} v_i^\tau\right) h_{n_k}^0 \Bigg\} \xrightarrow[k\to\infty]{} 0. \end{aligned} \tag{A.44}$$

We now show that (A.44) implies

$$\|h_{n_k}^i\| \xrightarrow[k\to\infty]{} 0, \qquad \|g_{n_k}^i\| \xrightarrow[k\to\infty]{} 0 \quad \forall i: 0 \leqq i \leqq t. \tag{A.45}$$

Applying Lemma 2 to $\sum_{i=1}^{n_k} w_{i-j} v_i^\tau$ and noticing (7), (A.42), we find that

$$\varlimsup_{n\to\infty} \frac{\log^\lambda n_k}{n_k} \sum_{j=0}^{t} g_{n_k}^{j\tau}\left(\sum_{i=1}^{n_k} w_{i-j} v_i^\tau\right) h_{n_k}^0 \leqq (1+t)c_1^2 \varlimsup_{k\to\infty} \frac{\log^\lambda n_k}{n_k} O(n_k^{1/2} \log^{1/2+\eta}(n_k + e)) = 0$$

for any $\omega \in D$ with a possible exception set of probability zero. In the following discussion such a possible exception is always assumed. We note that no measurability of $h_{n_k}^i$ and $g_{n_k}^i$ is required.

Similarly, by applying Lemma 2 to $\sum_{i=1}^{n_k} u_i^s v_i^\tau$ and $\sum_{i=1}^{n_k} u_{i-j} v_i^\tau$ $(j \geqq 1)$ and by use of (41), (42) and (A.42) we conclude that for $\omega \in D$

$$\overline{\lim_{k\to\infty}} \frac{\log^\lambda n_k}{n_k}\left[h_{n_k}^{0\tau}\left(\sum_{i=1}^{n_k} u_i^s v_i^\tau\right)h_{n_k}^0 + \sum_{j=1}^{t} h_{n_k}^{j\tau}\left(\sum_{i=1}^{n_k} u_{i-j} v_i^\tau\right)h_{n_k}^0\right] = 0.$$

Hence from (A.44) we have

$$\frac{(\log n_k)^\lambda}{n_k}\sum_{i=1}^{n_k}(h_{n_k}^{0\tau}u_i^s + h_{n_k}^{1\tau}u_{i-1} + \cdots + h_{n_k}^{t\tau}u_{i-t} + g_{n_k}^{0\tau}w_i + \cdots + g_{n_k}^{t\tau}w_{i-t})^2 \underset{k\to\infty}{\longrightarrow} 0, \tag{A.46}$$

and

$$\frac{(\log n_k)^\lambda}{n_k}\sum_{i=1}^{n_k}(h_{n_k}^{0\tau}v_i)^2 \underset{k\to\infty}{\longrightarrow} 0 \quad \text{for } \omega \in D. \tag{A.47}$$

By (A.29) and (A.47) it is clear that

$$\|h_{n_k}^0\|^2 = o((\log n_k)^{-\lambda+\varepsilon}), \qquad \omega \in D; \tag{A.48}$$

hence by (42)

$$\frac{(\log n_k)^{\lambda-(\varepsilon+\delta)}}{n_k}\sum_{i=1}^{n_k}(h_{n_k}^{0\tau}u_i^s)^2 = o(1), \qquad \omega \in D.$$

Then from here and (A.46) we have for $\omega \in D$

$$\frac{(\log n_k)^{\lambda-(\varepsilon+\delta)}}{n_k}\sum_{i=1}^{n_k}(h_{n_k}^{1\tau}u_{i-1} + \cdots + h_{n_k}^{t\tau}u_{i-t} + g_{n_k}^{0\tau}w_i + \cdots + g_{n_k}^{t\tau}w_{i-t})^2 \underset{k\to\infty}{\longrightarrow} 0. \tag{A.49}$$

Comparing (A.49) with (A.43) we see that in (A.49) we have deleted u_i by changing the order of $\log n_k$ from λ to $\lambda - (\varepsilon + \delta)$.

Generally, using the same treatment as described above we conclude that

$$\|h_{n_k}^i\|^2 = o((\log n_k)^{-\lambda+i(\varepsilon+\delta)+\varepsilon}), \qquad 0 \leqq i \leqq t, \quad \omega \in D \tag{A.50}$$

and

$$\frac{(\log n_k)^{\lambda-(t+1)(\varepsilon+\delta)}}{n_k}\sum_{i=1}^{n_k}(g_{n_k}^{0\tau}w_i + \cdots + g_{n_k}^{t\tau}w_{i-t})^2 \underset{k\to\infty}{\longrightarrow} 0, \qquad \omega \in D. \tag{A.51}$$

The same argument applied to (A.51) by using (7) and (A.42) leads to

$$\|g_{n_k}^i\|^2 = o((\log n_k)^{-\lambda+(t+1)(\varepsilon+\delta)}) \quad \forall i: 0 \leqq i \leqq t. \tag{A.52}$$

Since $t = mp + s - 1$ and $\lambda = \frac{1}{4} - 2\delta - \varepsilon$, then by (29), (A.50) and (A.52) imply (A.45); hence we have

$$H_{n_k}(z) \underset{k\to\infty}{\longrightarrow} 0, \qquad \omega \in D. \tag{A.53}$$

Let $\{\eta_{m_k}\}$ be a convergent subsequence of $\{\eta_{n_k}\}$: $\eta_{m_k} \underset{k\to\infty}{\longrightarrow} \eta$ with

$$\|\eta\| = 1, \qquad \omega \in D, \\ \eta = (\alpha^{0\tau} \cdots \alpha^{(p-1)\tau}, \beta^{0\tau} \cdots \beta^{(q-1)\tau}, \gamma^{0\tau} \cdots \gamma^{(\gamma-1)\tau})^\tau. \tag{A.54}$$

Then by (A.40) and (A.53) we have

$$\begin{aligned}\sum_{i=0}^{p-1} \alpha^{i\tau} z^i (\operatorname{Adj} A(z))[B(z), C(z)] \\ = -\sum_{i=0}^{q-1} \beta^{i\tau} z^i [\det A(z) I_l, 0] - \sum_{i=0}^{r-1} \gamma^{i\tau} z^i [0, \det A(z) I_m].\end{aligned} \tag{A.55}$$

Since $A(z)$, $B(z)$ and $C(z)$ have no common left factor, there are matrix polynomials $M(z)$, $N(z)$ and $L(z)$ such that

$$A(z)M(z) + B(z)N(z) + C(z)L(z) = I.$$

Then by (A.55) we see

$$\begin{aligned}&\sum_{i=0}^{p-1} \alpha^{i\tau} z^i \operatorname{Adj} A(z) \\ &= \sum_{i=0}^{p-1} \alpha^{i\tau} z^i \operatorname{Adj} A(z)\left(A(z)M(z) + [B(z), C(z)]\begin{bmatrix} N(z) \\ L(z) \end{bmatrix}\right) \\ &= \det A(z)\left[\sum_{i=0}^{p-1} \alpha^{i\tau} z^i M(z) - \sum_{i=0}^{q-1} \beta^{i\tau} z^i N(z) - \sum_{i=0}^{r-1} \gamma^{i\tau} z^i L(z)\right], \qquad \omega \in D.\end{aligned} \tag{A.56}$$

But

$$\begin{aligned}\deg\left(\sum_{i=0}^{p-1} \alpha^{i\tau} z^i \operatorname{Adj} A(z)\right) &\leqq p - 1 + \deg(\operatorname{Adj} A(z)) \\ &= p - 1 + mp - p < mp = \deg(\det A(z)),\end{aligned}$$

so (A.56) implies

$$\sum_{i=0}^{p-1} \alpha^{i\tau} z^i \operatorname{Adj} A(z) = 0, \qquad \omega \in D.$$

Hence $\alpha^i = 0$, $i = 0, \ldots, p-1$, and by (A.55) $\beta^i = 0$, $i = 0 \cdots q-1$, and $\gamma^j = 0$, $j = 1 \cdots r-1$ for $\omega \in D$. This conclusion contradicts with $\|\eta\| = 1$; therefore, $P(D) = 0$ and (A.38) is verified.

REFERENCES

B. D. O. ANDERSON AND J. B. MOORE (1971), *Linear Optimal Control*, Prentice-Hall, Englewood Cliffs, NJ.

A. BECKER, P. R. KUMAR AND C. Z. WEI (1985), *Adaptive control with the stochastic approximation algorithm, Geometry and convergence*, IEEE Trans. Automat. Control, AC-30, pp. 330–338.

P. E. CAINES AND S. LAFORTUNE (1984), *Adaptive control with recursive identification for stochastic linear systems*, IEEE Trans. Automat. Control, AC-29, pp. 312–321.

P. E. CAINES AND H. F. CHEN (1985), *Optimal adaptive LQG control for systems with finite state process parameter*, IEEE Trans. Automat. Control, AC-30, pp. 185–189.

H. F. CHEN (1981), *Quasi-least squares identification and its strong consistency*, Internat. J. Control, 34, pp. 921–936.

——— (1982), *Strong consistency and convergence rate of least squares identification*, Sci. Sinica. Ser. A, 25, pp. 771–784.

——— (1984), *Recursive system identification and adaptive control by use of the modified least squares algorithm*, this Journal, 22, pp. 759–776.

——— (1985), *Recursive Estimation and Control for Stochastic Systems*, John Wiley, New York.

H. F. CHEN AND P. E. CAINES (1984), *Adaptive linear quadratic control for stochastic discrete-time systems*, Preprints of the 9th World Congress of the International Federation of Automatic Control, Vol. 12, Budapest, pp. 150–154.

H. F. CHEN AND P. E. CAINES (1985), *Strong consistency of the stochastic gradient algorithm of adaptive control*, IEEE Trans. Automat. Control, AC-30, pp. 189-192.

H. F. CHEN AND L. GUO (1985a), *Adaptive control with recursive identification for stochastic linear systems*, in Advances in Control and Dynamic Systems, C. T. Leondes, ed., Vol. 24, Academic Press, New York.

——— (1985b), *Strong consistency of parameter estimates for discrete-time stochastic systems*, J. Systems Sci. Math. Sci., 5, pp. 81-93.

——— (1986), *Optimal stochastic adaptive control with quadratic index*, Internat. J. Control, 43, pp. 869-881.

——— (1987), *Asymptotically optimal adaptive control with consistent parameter estimates*, this Journal, 25, pp. 558-575.

Y. S. CHOW (1965), *Local convergence of martingales and the low of large numbers*, Ann. Math. Stat., 36, pp. 552-558.

G. C. GOODWIN, P. J. RAMADGE AND P. E. CAINES (1981), *Discrete-time stochastic adaptive control*, this Journal, 19, pp. 829-853.

G. C. GOODWIN, D. J. HILL AND M. PALANISWAMI (1984), *A perspective on convergence of adaptive control algorithm*, Automatica J. IFAC, 20, pp. 519-531.

O. B. HIJAB (1983), *The adaptive* LQG *problem, Part* 1, IEEE Trans. Automat. Control, AC-28, pp. 171-178.

T. KAILATH (1980), *Linear Systems*, Prentice-Hall, Englewood Cliffs, NJ.

P. R. KUMAR (1983), *Optimal adaptive control of linear-quadratic-Gaussian systems*, this Journal, 21, pp. 163-178.

T. L. LAI AND C. Z. WEI (1982), *Least squares estimation in stochastic regression models with application to identification and control of dynamic systems*, Ann. Statist., 10, pp. 154-166.

R. S. LIPSTER AND A. N. SHIRYAYEV (1978), *Statistics of Random Processes*, Springer, New York.

L. LJUNG (1977), *Analysis of recursive stochastic algorithm*, IEEE Trans. Automat. Control, AC-22, pp. 551-575.

C. SAMSON (1983), *Stability analysis of adaptively controlled systems subject to bounded disturbances*, Automatica—J IFAC, 19, pp. 81-86.

K. S. SIN AND G. C. GOODWIN (1982), *Stochastic adaptive control using a modified least squares algorithm*, Automatica—J. IFAC, 18, pp. 315-321.

V. SOLO (1979), *The convergence of* AML, IEEE Trans. Automat. Control, AC-24, pp. 958-962.