INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING, VOL. 5, 155–174 (1991)

L_p-STABILITY OF ESTIMATION ERRORS OF KALMAN FILTER FOR TRACKING TIME-VARYING PARAMETERS

J. F. ZHANG, L. GUO AND H. F. CHEN

Institute of Systems Science, Academia Sinica, Beijing 100080, People's Republic of China

SUMMARY

The Kalman filtering algorithm, owing to its optimality in some sense, is widely used in systems and control, signal processing and many other fields. This paper presents a detailed analysis for the L_p-stability of tracking errors when the Kalman filter is used for tracking unknown time-varying parameters. The results of this paper differ from the previous ones in that the regression vector (in a linear regression model) or the output matrix (in state space terminology) is random rather than deterministic. The context is kept general so that, in particular, the time-varying parameter is allowed to be unbounded, and no assumption of stationarity or independence for signals is made.

KEY WORDS Stochastic systems Estimation Time-varying parameter Kalman filter

1. INTRODUCTION

Consider the time-varying linear model

$$y_k = \varphi_k^{\mathrm{T}} \theta_k + v_k. \quad \forall k \geqslant 0 \tag{1}$$

where y_k and v_k are the scalar output and noise respectively and φ_k and θ_k are the r-dimensional stochastic regression vector and the unknown time-varying parameter respectively. For simplicity of notation, denote the parameter variation at time instant k by w_k:

$$w_k \triangleq \theta_k - \theta_{k-1}, \quad \forall k \geqslant 1 \tag{2}$$

In the special case where v_k is a moving average process and φ_n consists of input-output data, i.e.

$$\varphi_k^{\mathrm{T}} = [y_{k-1} \ldots y_{k-s} \; u_{k-1} \ldots u_{k-t}]$$

with u_k being the input signal, then the linear model (1) is reduced to the ARMAX model with time-varying coefficients.

Tracking or estimating a system or a signal whose properties vary with time is a fundamental problem in system identification and signal processing. This problem has received considerable attention in the field of signal processing (see e.g. References 1–7), where most of the works are concerned with the study of the so-called least mean squares (LMS) algorithm or the normalized gradient algorithm, and usually some sort of stationarity and/or independence is required. In contrast to this, few precise studies have been done on the time-varying parameter-

This paper was recommended for publication by editor A. Benveniste

0890–6327/91/030155–20$10.00

Received 31 January 1990
Revised September 1990

tracking problem in the area of system identification, and most of the works are concentrated on the constant-parameter case, i.e. $w_k \equiv 0$ in (2) (see e.g. References 8–10).

Note that if we regard (1) and (2) as a state space model with state θ_k, then it is natural to use the Kalman filter to estimate the time-varying parameter θ_k (see e.g. References 6 and 11–14). The Kalman filter takes the form

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \frac{P_k\varphi_k}{R + \varphi_k^{\mathrm{T}} P_k \varphi_k}(y_k - \varphi_k^{\mathrm{T}}\hat{\theta}_k) \tag{3}$$

$$P_{k+1} = P_k - \frac{P_k\varphi_k\varphi_k^{\mathrm{T}} P_k}{R + \varphi_k^{\mathrm{T}} P_k \varphi_k} + Q \tag{4}$$

where $P_0 \geqslant 0$, $R > 0$, $Q > 0$ and $\hat{\theta}_0$ are deterministic and can be arbitrarily chosen (here R and Q may be regarded as the *a priori* estimates for the variances of v_k and w_k respectively).

It is well known that if φ_k is $\mathscr{F}_{k-1}$-measurable, where $\mathscr{F}_{k-1} = \sigma\{y_i, i \leqslant k-1\}$, and $\{w_k, v_k\}$ is a Gaussian white noise process, then θ_k generated by (3) and (4) is the minimum variance estimate for θ_k and P_k is the estimation error covariance, i.e.

$$\hat{\theta}_k = \mathrm{E}(\theta_k \mid \mathscr{F}_{k-1}), \qquad P_k = \mathrm{E}(\tilde{\theta}_k\tilde{\theta}_k^{\mathrm{T}} \mid \mathscr{F}_{k-1}) \qquad (\tilde{\theta}_k = \theta_k - \hat{\theta}_k) \tag{5}$$

provided that $Q = \mathrm{E}(w_k w_k^{\mathrm{T}})$, $R = \mathrm{E}v_k^2$, $\hat{\theta}_0 = \mathrm{E}\theta_0$ and $P_0 = \mathrm{E}(\tilde{\theta}_0\tilde{\theta}_0^{\mathrm{T}})$ (see e.g. References 15 and (16).

In studying asymptotic properties of the above algorithm, the primary issue is to establish boundedness (in some sense) of the tracking error $\tilde{\theta}_k$. This problem is obviously related to the stability theory of the Kalman filter, and the standard condition for such a stability (boundedness) is that the regression vector φ_k is deterministic and satisfies

$$\alpha I \leqslant \sum_{k=m+1}^{m+h} \varphi_k\varphi_k^{\mathrm{T}} \leqslant \beta I, \quad \forall m \tag{6}$$

for some positive constants α, β and integer h (see e.g. Reference 17).

As pointed out by Guo,[18] condition (6) is mainly a deterministic hypothesis; it excludes standard stochastic signals, including the Gaussian signal and white noise signals, and hence (6) is unsuitable for the stability study of the Kalman filter when φ_k is a random process. To the best of our knowledge, the first result which guarantees the stability of (3) and (4) and allows $\{\varphi_k\}$ to be a large class of stochastic processes appears in the recent work of Guo,[18] where it is assumed that $\{\varphi_k, \mathscr{F}_k\}$ is an adapted process ($\mathscr{F}_k$ is any family of non-decreasing σ-algebras) satisfying

$$\mathrm{E}\left(\sum_{k=m+1}^{m+h} \frac{\varphi_k\varphi_k^{\mathrm{T}}}{1 + \|\varphi_k\|^2} \,\middle|\, \mathscr{F}_m\right) \geqslant \alpha I, \quad \text{a.s.}, \quad \forall m \geqslant 0 \tag{7}$$

for some constant $\alpha > 0$ and integer h.

In this paper we will weaken condition (7) and generalize results in Reference 18, and at the same time provide new results and insights. The paper is organized as follows. In Section 2 we present our main results. Section 3 provides some lemmas and establishes the properties of $\{P_k\}$. The proof of the theorems is put into Section 4. Finally, in the Appendix the proof of auxiliary results which are used in the text is provided.

2. MAIN RESULTS

In the sequel the norm $\|X\|$ for a matrix X is defined as $\|X\| = \{\lambda_{\max}(XX^{\mathrm{T}})\}^{1/2}$, and $\lambda_{\max}(X)$ ($\lambda_{\min}(X)$) denotes the largest (smallest) eigenvalue of X. Let us first give a definition.

Definition 1

A random vector sequence $\{x_k, k \geqslant 0\}$ defined on the basic probability space $(\Omega, \mathscr{F}, P)$ is called L_p-stable $(p > 0)$ in the sample average sense if

$$\sup_{k \geqslant 0} \mathrm{E} \| x_k \|^p < \infty$$

and in the time average sense if

$$\sup_{k > 0} \frac{1}{k} \sum_{i=0}^{k} \| x_i \|^p < \infty, \quad p > 0, \quad \text{a.s.}$$

We now give the main condition that will be used in the paper.

Condition 1

$\{\varphi_k, \mathscr{F}_k\}$ is an adapted sequence (i.e. φ_k is $\mathscr{F}_k$-measurable for any k, where $\{\mathscr{F}_k\}$ is a family of non-decreasing σ-algebras) and there exists an integer $h > 0$ such that

$$\mathrm{E}\left(\sum_{k=m+1}^{m+h} \frac{\varphi_k \varphi_k^{\mathrm{T}}}{1 + \| \varphi_k \|^2} \middle| \mathscr{F}_m \right) \geqslant \frac{1}{\alpha_m} I, \quad \text{a.s.,} \quad \forall m \geqslant 0 \tag{8}$$

Here $\{\alpha_m, \mathscr{F}_m\}$ is an adapted non-negative sequence satisfying

$$\alpha_{m+1} \leqslant a \alpha_m + \eta_{m+1}, \quad \forall m \geqslant 0, \qquad M_0 \triangleq \mathrm{E} \alpha_0^{1+\delta} < \infty \tag{9}$$

where $\{\eta_m, \mathscr{F}_m\}$ is an adapted non-negative sequence,

$$\sup_{m \geqslant 0} \mathrm{E}(\eta_{m+1}^{1+\delta} \mid \mathscr{F}_m) \leqslant M, \quad \text{a.s.} \tag{10}$$

and where $a \in [0, 1)$, $0 < \delta < \infty$ and $0 \leqslant M < \infty$ are constants.

Remark 1

At first glance, Condition 1 looks rather complicated; however, it does have a clear meaning and is satisfied by a large class of stochastic signals. Note first that if in (9) we take $a = 0$ and $\eta_m \equiv \alpha^{-1}$ for some $\alpha > 0$, then (8) reduces to (7), which is weaker than (6), and hence Condition 1 includes (6) and (7). Next note that the sequence $\{\alpha_k\}$ required in (8) may not be bounded in the sample path and hence the matrix on the LHS of (8) may not be uniformly positive definite, so (8) is really weaker than (7).

Let us further illustrate Condition 1 by the following examples.

Example 1

Let $\{\varphi_k\}$ be an r-dimensional ϕ-mixing process; that is, there is a deterministic sequence $\{\phi(h), h \geqslant 0\}$ such that

(i) $$\phi(h) \to 0 \quad \text{as } h \to \infty$$

(ii) $$\sup_{\substack{A \in \mathscr{F}_{s+h}^{\infty} \\ B \in \mathscr{F}_0^{s}}} | P(A \mid B) - P(A) | \leqslant \phi(h), \quad \forall s \geqslant 0, \quad \forall h \geqslant 0$$

where, for any non-negative integers $s \geqslant 0$ and $h \geqslant 0$, $\mathscr{F}_0^s \triangleq \sigma\{\varphi_k, 0 \leqslant k \leqslant s\}$ and $\mathscr{F}_{s+h}^{\infty} \triangleq \sigma\{\varphi_k, s+h \leqslant k < \infty\}$.

Suppose further that

$$\inf_k \lambda_{\min} \mathrm{E}(\varphi_k \varphi_k^{\mathrm{T}}) > 0 \quad \text{and} \quad \sup_k \mathrm{E} \| \varphi_k \|^4 < \infty \tag{11}$$

Then Condition 1 holds with $\mathscr{F}_m = \mathscr{F}_0^m$.

The proof of this example is given in the Appendix. We remark that any h-dependent random process (including moving average processes of order h) is ϕ-mixing. This kind of condition has been previously used, for example, by Eweda and Macchi,[3] Macchi[4] and Kushner[19] in their study of the LMS algorithm.

Example 2

Let $\{\varphi_k\}$ be the output of the linear stochastic model

$$x_k = A x_{k-1} + B \xi_k, \quad \forall k \geqslant 1, \qquad \mathrm{E} \| x_0 \|^5 < \infty \tag{12}$$

$$\varphi_k = C x_k + \zeta_k, \quad \forall k \geqslant 0 \tag{13}$$

where $A \in R^{n \times n}$, $B \in R^{n \times q}$ and $C \in R^{r \times n}$ are deterministic matrices, A is stable and (A, B, C) is output-controllable in the sense of Fortmann and Hitz.[20]

Suppose that $\{\xi_k\}$ and $\{\zeta_k\}$ are independent processes which are also mutually independent and satisfy

$$\mathrm{E}\xi_k = 0, \qquad \mathrm{E}\zeta_k = 0 \tag{14}$$

$$\mathrm{E}(\xi_k \xi_k^{\mathrm{T}}) \geqslant \varepsilon I, \quad \forall k \geqslant 0 \tag{15}$$

$$\mathrm{E}(\| \xi_k \|^{4(1+\mu)} + \| \zeta_k \|^4) \leqslant M < \infty, \quad \forall k \geqslant 0 \tag{16}$$

for some constants $\varepsilon > 0$, $\mu > 0$ and $M > 0$. Then Condition 1 is fulfilled.

The proof is given in the Appendix. It is worth noting that it is generally hard to show that $\{\varphi_k\}$ defined in Example 2 satisfies condition (7) unless the noise process $\{\xi_k, \zeta_k\}$ is assumed to be bounded (see Reference 18, Example 2). We now present the main results of the paper.

Theorem 1

Consider the time-varying model (1) and (2). Suppose that $\{v_k, w_k\}$ is a stochastic sequence which satisfies for some $p > 0$ and $\beta > 1$

$$\sigma_p \triangleq \sup_{k \geqslant 0} \mathrm{E}\{ Z_k^p [\log(e + Z_k)]^{\beta + 3p/2} \} < \infty \tag{17}$$

$$\mathrm{E}\{ \| \tilde{\theta}_0 \|^p [\log(e + \| \tilde{\theta}_0 \|)]^{p/2} \} < \infty \tag{18}$$

where $Z_k = \| v_k \| + \| w_{k+1} \|$, $\tilde{\theta}_0 = \theta_0 - \hat{\theta}_0$ and v_k, w_k, θ_0 and $\hat{\theta}_0$ are given by (1)–(4) respectively. Then under Condition 1 the estimation error $\{\theta_k - \hat{\theta}_k, k \geqslant 0\}$ generated by (3) and (4) is L_p-stable in the sample average sense and

$$\limsup_{k \to \infty} \mathrm{E} \| \theta_k - \hat{\theta}_k \|^p \leqslant A [\sigma_p \log^{1 + 3p/2}(e + \sigma_p^{-1})] \tag{19}$$

where A is a finite constant dependent on h, a, M, M_0 and δ only.

Moreover, if $v_k \equiv 0$ and $w_k \equiv 0$ (i.e. $\theta_k \equiv \theta_0$), then

$$E\|\theta_k - \hat{\theta}_k\|^p \to 0 \quad \text{as } k \to \infty \tag{20}$$

$$E\|\theta_k - \hat{\theta}_k\|^q \to 0, \quad \text{exponentially fast} \tag{21}$$

for any $q \in (0, p)$.

The proof is given in Section 4.

Remark 2

If in Theorem 1 $\{\varphi_k\}$ and $\{v_k, w_k\}$ are assumed to be mutually independent, then for the L_p-stability of $\{\hat{\theta}_k - \theta_k\}$, condition (17) can be replace by a weaker one:

$$\sup_{k \geq 0} E Z_k^p < \infty \tag{22}$$

This condition is obviously a natural one for the desired L_p-stability. What condition (17) effectively means is that if the independence between $\{\varphi_k\}$ and $\{v_k, w_k\}$ is removed, then the L_p-stability of $\{\hat{\theta}_k - \theta_k\}$ is still preserved provided that the moment condition (22) is slightly strengthened.

Next we present a result on the time average of the estimation error $\{\hat{\theta}_k - \theta_k\}$.

Theorem 2

Consider the time-varying model (1) and (2). Suppose that $\{v_k, w_k\}$ is a stochastic sequence and for some $p > 0$

$$\varepsilon_p \triangleq \limsup_{k \to \infty} \frac{1}{k} \sum_{i=0}^{k-1} (\|v_i\|^p + \|w_{i+1}\|^p) < \infty, \quad \text{a.s.} \tag{23}$$

Then under Condition 1 $\{\hat{\theta}_k - \theta_k, k \geq 0\}$ is L_q-stable in the time average sense for any $q \in (0, p)$, and

$$\limsup_{k \to \infty} \frac{1}{k} \sum_{i=0}^{k} \|\hat{\theta}_i - \theta_i\|^q \leq B(\varepsilon_p)^{q/p} \tag{24}$$

where B is a finite *deterministic* constant depending on q, h, a, M, M_0 and δ only. Furthermore, if $v_k \equiv 0$ and $\theta_k \equiv \theta_0$, then

$$\hat{\theta}_k \to \theta_0, \quad \text{a.s., exponentially fast} \tag{25}$$

The proof of this theorem is also given in Section 4.

We remark that both Theorems 1 and 2 are significant extensions of those in a recent work of Guo.[18] For example, Guo[18] has only studied the L_2-stability of $\{\hat{\theta}_k - \theta_k\}$, for which he has assumed that

$$\sup_{k \geq 0} E Z_k^{3+\beta} < \infty$$

while here for such a stability we only need

$$\sup_{k \geq 0} E Z_k^2 [\log(e + Z_k)]^{3+\beta} < \infty$$

where $\beta > 1$ is a constant.

3. PROPERTIES OF $\{P_k\}$

In this section we establish some properties of $\{P_k\}$, which is a basic step in the stability study of $\{\theta_k - \theta_k\}$. To this end we first give some lemmas.

Lemma 1

Let $\{a_m, \mathscr{F}_m\}$ be an adapted random process satisfying

$$a_m \in [0, 1], \qquad E(a_{m+1} \mid \mathscr{F}_m) \geqslant 1/\alpha_m, \quad \forall m \geqslant 0 \tag{26}$$

where $\{\alpha_m\}$ satisfies properties (9) and (10). Then there exist two constants $C > 0$ and $\gamma \in (0, 1)$ such that

$$E \prod_{k=m}^{n} (1 - a_{k+1}) \leqslant C\gamma^{n-m+1}, \quad \forall n \geqslant m \geqslant 0 \tag{27}$$

where C and γ depend on a, M, M_0 and δ only.

The proof of this lemma is given in the Appendix.

Lemma 2

Let $\{x_k, \mathscr{F}_k\}$ be an adapted non-negative random process satisfying

$$x_{k+1} \leqslant (1 - a_{k+1})\alpha x_k + C, \quad \forall k \geqslant 0, \qquad x_0 = 1 \tag{28}$$

where $\{a_k\}$ is defined as in Lemma 1 and α and C are finite, positive constants. Then there exists a constant $\alpha^* > 1$ such that when $\alpha \in [1, \alpha^*)$

$$\limsup_{k \to \infty} \frac{1}{k} \sum_{i=0}^{k} x_i \leqslant C', \quad \text{a.s.} \tag{29}$$

where C' is a constant.

The proof is also presented in the Appendix.

Lemma 3

Let $\{P_k\}$ be generated by (4). Then

$$T_{m+1} \leqslant (1 - a_{m+1})T_m + d \tag{30}$$

where

$$T_m = \sum_{k=(m-1)h}^{mh-1} \operatorname{tr} P_{k+1}, \qquad T_0 = 0 \tag{31}$$

$$a_{m+1} = \frac{\operatorname{tr}\left((P_{mh} + hQ)^2 \sum_{k=mh}^{(m+1)h-1} \frac{\varphi_k \varphi_k^{\mathrm{T}}}{1 + \| \varphi_k \|^2} \right)}{h(R+1)[1 + \lambda_{\max}(P_{mh} + hQ)] \operatorname{tr}(P_{mh} + hQ)} \tag{32}$$

$$d = \tfrac{3}{2}\, h(h+1) \operatorname{tr} Q \tag{33}$$

and where h is the constant appearing in Condition 1.

Proof. Note that by (4)

$$P_k \leqslant P_{k-1} + Q \leqslant \ldots \leqslant P_{mh} + hQ \tag{34}$$

holds for any $k \in [mh, (m+1)h]$. Hence by the matrix inverse formula it follows that for any $k \in [mh, (m+1)h]$

$$\begin{aligned}
P_{k+1} &= (P_k^{-1} + R^{-1}\varphi_k\varphi_k^{\mathrm{T}})^{-1} + Q \\
&\leqslant [(P_{mh} + hQ)^{-1} + R^{-1}\varphi_k\varphi_k^{\mathrm{T}}]^{-1} + Q \\
&= P_{mh} - \frac{(P_{mh} + hQ)\varphi_k\varphi_k^{\mathrm{T}}(P_{mh} + hQ)}{R + \varphi_k^{\mathrm{T}}(P_{mh} + hQ)\varphi_k} + (h+1)Q \\
&\leqslant P_{mh} - \frac{(P_{mh} + hQ)\dfrac{\varphi_k\varphi_k^{\mathrm{T}}}{1 + \|\varphi_k\|^2}(P_{mh} + hQ)}{(R+1)[1 + \lambda_{\max}(P_{mh} + hQ)]} + (h+1)Q \\
&\leqslant P_{mh} - \frac{(P_{mh} + hQ)\dfrac{\varphi_k\varphi_k^{\mathrm{T}}}{1 + \|\varphi_k\|^2}(P_{mh} + hQ)}{h(R+1)[1 + \lambda_{\max}(P_{mh} + hQ)]\operatorname{tr}(P_{mh} + hQ)} h\operatorname{tr}P_{mh} + (h+1)Q
\end{aligned} \tag{35}$$

Summing both sides of (35) and taking into account (31) and (32), we obtain

$$T_{m+1} \leqslant h\operatorname{tr}P_{mh} - a_{m+1}h\operatorname{tr}P_{mh} + h(h+1)\operatorname{tr}Q \tag{36}$$

Again, by (31) and (34)

$$\begin{aligned}
h\operatorname{tr}P_{mh} &= \sum_{k=(m-1)h}^{mh-1} \operatorname{tr}P_{mh} \\
&\leqslant \sum_{k=(m-1)h}^{mh-1} \operatorname{tr}[P_{k+l} + (mh-k)Q] \\
&\leqslant T_m + \tfrac{1}{2}h(h+1)\operatorname{tr}Q
\end{aligned}$$

Substituting this into (36), we get the desired relation (30). □

We are now in a position to prove the following main result of this section.

Lemma 4

For $\{P_k\}$ generated by (4), if Condition 1 holds, then there exists a constant $\varepsilon^* > 0$ such that for any $\varepsilon \in [0, \varepsilon^*)$

$$\sup_{k \geqslant 0} \mathrm{E} \exp(\varepsilon \| P_k \|) \leqslant C \tag{37}$$

$$\limsup_{k \to \infty} \frac{1}{k} \sum_{i=0}^{k} \exp(\varepsilon \| P_i \|) \leqslant C'. \quad \text{a.s.} \tag{38}$$

where C and C' are constants.

Proof. Denote $\mathscr{Y}_m = \mathscr{F}_{mh-1}$, where $\{\mathscr{F}_m\}$ is the same as in Condition 1. Then it is clear that for T_m and a_m defined by (31) and (32), T_m and a_m are $\mathscr{Y}_m$-measurable, and moreover, by (32) and Condition 1

$$a_{m+1} \in \left[0, \frac{1}{R+1}\right] \tag{39}$$

$$\begin{aligned} \mathrm{E}[a_{m+1} \mid \mathscr{Y}_m] &\geqslant \frac{\operatorname{tr}(P_{mh}+hQ)^2}{\alpha_{mh-1}h(R+1)[1+\lambda_{\max}(P_{mh}+hQ)]\operatorname{tr}(P_{mh}+hQ)} \\ &\geqslant \frac{r^{-1}[\operatorname{tr}(P_{mh}+hQ)]^2}{\alpha_{mh-1}h(R+1)[1+\lambda_{\max}(P_{mh}+hQ)]\operatorname{tr}(P_{mh}+hQ)} \\ &\geqslant \frac{\|Q\|}{r(R+1)(1+h\|Q\|)}\frac{1}{\alpha_{mh-1}} \end{aligned}$$

Set

$$\beta_m = r\|Q\|^{-1}(R+1)(1+h\|Q\|)\alpha_{mh-1}$$

Then we have

$$\mathrm{E}(a_{m+1} \mid \mathscr{Y}_m) \geqslant 1/\beta_m \tag{40}$$

It is easy to verify that $\{\beta_m, \mathscr{Y}_m\}$ is an adapted sequence which satisfies via (9) and (10)

$$\beta_{m+1} \leqslant \bar{a}\beta_m + \bar{\eta}_{m+1}, \quad m \geqslant 0, \qquad \mathrm{E}\beta_0 < \infty \tag{41}$$

where $\bar{a} = a^h$ and $\{\bar{\eta}_m, \mathscr{Y}_m\}$ is an adapted sequence satisfying

$$\sup_{m\geqslant 0} \mathrm{E}(\bar{\eta}_{m+1}^{1+\delta} \mid \mathscr{Y}_m) \leqslant \bar{M} \tag{42}$$

for some constant $\bar{M} > 0$.

Consequently, by applying Lemma 1, we obtain

$$\mathrm{E}\prod_{k=m}^{n}(1-a_{k+1}) \leqslant C\gamma^{n-m+1}, \quad \forall n \geqslant m \geqslant 0 \tag{43}$$

for some constants $C > 0$ and $\gamma \in (0,1)$.

Next, from Lemma 3 it follows that for any $\varepsilon > 0$

$$\exp(\varepsilon T_{m+1}) \leqslant \exp[(1-a_{m+1})\varepsilon T_m]\, e^{d\varepsilon} \tag{44}$$

Consequently, noticing the obvious inequality

$$\exp(\alpha x) - 1 \leqslant \alpha \exp(x), \quad 0 < \alpha < 1, \quad x > 0$$

we get

$$\exp(\varepsilon T_{m+1}) \leqslant e^{d\varepsilon}[(1-a_{m+1})\exp(\varepsilon T_m) + 1] \tag{45}$$

Hence from this and (43) it is easy to convince oneself that if $\varepsilon^* > 0$ is taken small enough such that $e^{d\varepsilon^*}\gamma < 1$, then

$$\sup_{m\geqslant 0} \mathrm{E}\exp(\varepsilon T_m) < \infty, \quad \forall \varepsilon \in (0, \varepsilon^*)$$

This proves the first assertion (37) of the lemma, while the second assertion (38) follows immediately by applying Lemma 2 with $x_k = \exp(\varepsilon T_k)$ and $\alpha = e^{d\varepsilon}$ to (45).

Hence the proof is completed. □

The following result is a direct consequence of Lemma 4.

Corollary 1

For $\{P_k\}$ generated by (4), if Condition 1 holds, then for any $m > 0$

$$\sup_{k \geqslant 0} \mathrm{E} \| P_k \|^m < \infty$$

$$\limsup_{k \to \infty} \frac{1}{k} \sum_{i=0}^{k} \| P_i \|^m \leqslant C, \quad \text{a.s.}$$

where C is a constant.

4. PROOF OF THEOREMS

In this section we will give the proof for Theorems 1 and 2. Let us first prove some lemmas.

Denoting

$$\tilde{\theta}_k = \theta_k - \hat{\theta}_k \quad \text{and} \quad V_k = \tilde{\theta}_k^{\mathrm{T}} P_k^{-1} \tilde{\theta}_k, \quad \forall k \geqslant 0 \tag{46}$$

we then have the following result.

Lemma 5

For any $q > 0$ there exist two constants $\mu \in (0, 1)$ and $C \geqslant 1$ such that

$$V_{k+1}^q \leqslant \left(1 - \frac{\mu}{1 + \operatorname{tr} P_k}\right) V_k^q + C(1 + \operatorname{tr} P_k)^{2q} Z_k^{2q} \tag{47}$$

where $Z_k = \| v_k \| + \| w_{k+1} \|$.

Proof. From Lemma 6 of Reference 18 we know that

$$V_{k+1} \leqslant \left(1 - \frac{\mu_1}{1 + \operatorname{tr} P_k}\right) V_k + C_1(1 + \operatorname{tr} P_k) Z_k^2, \quad \forall k \geqslant 0 \tag{48}$$

holds for some constants $\mu_1 \in (0, 1)$ and $C_1 < \infty$.

We first consider the case where $q > 1$. Note that in this case x^q, $x > 0$, is a convex function. Hence it follows from (48) that

$$V_{k+1}^q \leqslant \left(1 - \frac{\mu_1}{1 + \operatorname{tr} P_k}\right) V_k^q + \frac{\mu_1}{1 + \operatorname{tr} P_k} \left(\frac{C_1(1 + \operatorname{tr} P_k)^2}{\mu_1} Z_k^2\right)^q$$

which implies the desired result.

We now assume $q \in [0, 1]$. In this case the desired result (47) can also be easily derived from (48) by applying the following elementary inequalities:

$$(x + y)^q \leqslant x^q + y^q, \quad \forall x \geqslant 0, \quad y \geqslant 0, \quad q \in [0, 1]$$
$$(1 - x)^q \leqslant 1 - qx, \quad 0 < x < 1, \quad q \in [0, 1]$$

This completes the proof. □

Lemma 6

Let $\{x_k, \mathscr{F}_k\}$ be a non-negative adapted process, $x_k \geqslant 1$, which satisfies

$$x_{k+1} \leqslant (1 - a_{k+1}) x_k + C, \quad \forall k \geqslant 0, \qquad \mathrm{E} x_0^2 \leqslant \infty \tag{49}$$

where $C>0$ is a constant, $\{a_k\}$ is defined as in Lemma 1 and $a_k \in [0, \bar{a}]$, $\bar{a}<1$. Then there exist constants $N>0$ and $\lambda \in (0,1)$ such that

$$\mathrm{E} \prod_{k=m}^{n} \left(1-\frac{1}{x_k}\right) \leqslant N\lambda^{n-m+1}, \quad \forall n \geqslant m \geqslant 0 \tag{50}$$

This lemma can be proved along the lines of arguments of Lemma 4 in Reference 18. For a detail derivation see the Appendix.

Lemma 7

Let $\{P_k\}$ be defined by (4) with $\{\varphi_k\}$ satisfying Condition 1. Then for any $\mu \in (0,1)$ there exist constants $N>0$ and $\lambda \in (0,1)$ such that

$$\mathrm{E} \prod_{k=m}^{n} \left(1-\frac{\mu}{1+\operatorname{tr} P_k}\right) \leqslant N\lambda^{n-m+1}, \quad \forall n \geqslant m \geqslant 0 \tag{51}$$

Proof. Let us denote

$$x_m = \frac{1}{\mu}(h+T_m)$$

where T_m is defined by (31); then it follows from Lemma 3 that

$$x_{m+1} \leqslant (1-a_{m+1})x_m + \frac{h+d}{\mu}$$

Hence by noting (39)–(42), we see that Lemma 6 is applicable. So there are $N_0>0$ and $\lambda_0 \in (0,1)$ such that

$$\mathrm{E} \prod_{k=m}^{n} \left(1-\frac{1}{x_{k+1}}\right) \leqslant N_0\lambda_0^{n-m+1}, \quad \forall n \geqslant m \geqslant 0$$

From this it is easy to conclude (51) (see the proof of Lemma 5 in Reference 18). □

Lemma 8

Let $C_{nk} \in [0,1]$, $n \geqslant k \geqslant 0$, be a double-indexed stochastic sequence satisfying

$$\mathrm{E}C_{nk} \leqslant N\lambda^{n-k}, \quad \forall n \geqslant k \geqslant 0$$

for some constants $N>0$ and $\lambda \in (0,1)$. If $\{x_k\}$ is a non-negative stochastic sequence and

$$\sigma \triangleq \sup_{k \geqslant 0} \mathrm{E}x_k \log^{\beta}(e+x_k) < \infty$$

for some $\beta>1$, then

$$\sum_{k=0}^{n} \mathrm{E}(C_{nk}x_k) \leqslant C[\sigma \log(e+\sigma^{-1})], \quad n \geqslant 0$$

where C is a constant depending on β, N and λ.

The proof is given in the Appendix.

We are now in a position to prove our main theorems.

Proof of Theorem 1

We will repeatedly use the following inequality

$$x^{\alpha}y \leqslant \sigma_p \exp(\varepsilon x) + C_1 y[\log^{\alpha}(e+\sigma_p^{-1}) + \log^{\alpha}(e+y)] \tag{52}$$

$$\leqslant \sigma_p \exp(\varepsilon x) + 2C_1[\log(e+\sigma_p^{-1})]^{\alpha} y \log^{\alpha}(e+y) \tag{53}$$

where x, y, ε and α are any non-negative numbers, σ_p is as defined by (17) and

$$C_1 = \left(\frac{1+\alpha}{\varepsilon}\right)^{\alpha} 4^{1+\alpha} \log^{\alpha}\left[e + \left(\frac{1+\alpha}{\varepsilon}\right)^{\alpha}\right]$$

Inequality (53) follows directly from (52), while (52) may be derived from the well-known Young's inequality (see the Appendix).

Denote

$$f(x) = x \log^{p/2}(e+x), \quad p>0, \quad x \geqslant 0 \tag{54}$$

Then from (52) (with $\alpha = p/2$) and the definition of V_k in (46) it follows that for any $p>0$

$$\|\tilde{\theta}_n\|^p \leqslant \|P_n\|^{p/2} V_n^{p/2} \leqslant \sigma_p \exp(\varepsilon\|P_n\|) + C'[\log^{p/2}(e+\sigma_p^{-1})V_n^{p/2} + f(V_n^{p/2})] \tag{55}$$

where and hereafter C' denotes some constant, which may be different from place to place.

The first term on the RHS of (55) is easy to deal with, since by Lemma 4

$$\sup_n \mathrm{E} \exp(\varepsilon\|P_n\|) < \infty, \quad \forall \varepsilon \in (0, \varepsilon^*) \tag{56}$$

We now proceed to estimate $\mathrm{E}V_n^{p/2}$ and $\mathrm{E}f(V_n^{p/2})$. By Lemma 5 we know that

$$V_{k+1}^{p/2} \leqslant \left(1 - \frac{\mu}{1+\operatorname{tr}P_k}\right) V_k^{p/2} + C(1+\operatorname{tr}P_k)^p Z_k^p \tag{57}$$

Note that $f(x)$ defined by (54) is convex. It follows from (57) that

$$f(V_{k+1}^{p/2}) \leqslant \left(1 - \frac{\mu}{1+\operatorname{tr}P_k}\right) f(V_k^{p/2}) + \xi_k \tag{58}$$

where

$$\xi_k = \frac{\mu}{1+\operatorname{tr}P_k} f\left(\frac{C(1+\operatorname{tr}P_k)^{p+1}}{\mu} Z_k^p\right)$$

It is easy to see that

$$\begin{aligned}
\xi_k &= C(1+\operatorname{tr}P_k)^p Z_k^p \log^{p/2}\left(e + \frac{C(1+\operatorname{tr}P_k)^{p+1}}{\mu} Z_k^p\right) \\
&\leqslant C(1+\operatorname{tr}P_k)^p Z_k^p \log^{p/2}\left[(e+Z_k^p)\left(e + \frac{C(1+\operatorname{tr}P_k)^{p+1}}{\mu}\right)\right] \\
&\leqslant 2^{p/2} C(1+\operatorname{tr}P_k)^p \left[f(Z_k^p) + Z_k^p \log^{p/2}\left(e + \frac{C(1+\operatorname{tr}P_k)^{p+1}}{\mu}\right)\right] \\
&\leqslant C'[(1+\operatorname{tr}P_k)^p f(Z_k^p) + (1+\operatorname{tr}P_k)^{3p/2} Z_k^p]
\end{aligned} \tag{59}$$

By (53) we see that

$$(1+\operatorname{tr}P_k)^p f(Z_k^p) \leqslant \sigma_p \exp[\varepsilon(1+\operatorname{tr}P_k)] + C'[\log(e+\sigma_p^{-1})]^p f(Z_k^p) \log^p[e + f(Z_k^p)]$$

$$(1+\operatorname{tr} P_k)^{3p/2} Z_k^p \leqslant \sigma_p \exp[\varepsilon(1+\operatorname{tr} P_k)] + C'[\log(e+\sigma_p^{-1})]^{3p/2} Z_k^p \log^{3p/2}(e+Z_k^p)$$

From this and (59) we arrive at

$$\xi_k \leqslant C'\sigma_p \exp[\varepsilon(1+\operatorname{tr}P_k)] + C'[\log(e+\sigma_p^{-1})]^{3p/2} Z_k^p \log^{3p/2}(e+Z_k^p) \tag{60}$$

Now let us define $\Phi(n,k)$ as follows:

$$\Phi(n+1,k) = \left[1-\frac{\mu}{1+\operatorname{tr}P_n}\right]\Phi(n,k), \qquad \Phi(k,k)=I. \tag{61}$$

Then by Lemma 7 we known that $\{\Phi(n+1,k)\}$ satisfies the same condition as that of C_{nk} in Lemma 8.

By (60) and (61) it follows from (58) that ($\varepsilon \in (0, \varepsilon^*/r)$)

$$f(V_{n+1}^{p/2}) \leqslant \Phi(n+1,0) f(V_0^{p/2}) + C'\sigma_p \sum_{k=0}^{n+1} \Phi(n+1,k)\exp[\varepsilon(1+\operatorname{tr}P_k)]$$

$$+ C' \log^{3p/2}(e+\sigma_p^{-1}) \sum_{k=0}^{n+1} \Phi(n+1,k) Z_k^p \log^{3p/2}(e+Z_k^p) \tag{62}$$

By Lemma 7 it is seen that

$$E\Phi(n+1,k) \leqslant N\lambda^{n-k}, \quad \forall n \geqslant k \geqslant 0 \tag{63}$$

for some constants $n > 0$ and $\lambda \in (0,1)$. Hence we have

$$E[\lambda_1^n \Phi(n+1,0)] \leqslant N(\lambda\lambda_1)^n, \quad \forall \lambda_1 \in (1, 1/\lambda)$$

$$E\left(\sum_{n=0}^{\infty} \lambda_1^n \Phi(n+1,0)\right) = \sum_{n=0}^{\infty} E[\lambda_1^n \Phi(n+1,0)] < \infty$$

which implies

$$\lambda_1^n \Phi(n+1,0) \to 0, \quad \text{a.s.} \tag{64}$$

Hence the expectation of the first term on the RHS of (62) converges to zero by condition (18) and the dominated convergence theorem, while the other two terms can be estimated by applying Lemma 8. Therefore we have

$$\limsup_{n\to\infty} Ef(V_{n+1}^{p/2}) \leqslant C'\sigma_p \log^{3p/2+1}(e+\sigma_p^{-1}) \tag{65}$$

Next we estimate $EV_n^{p/2}$. By (53) it is seen that

$$(1+\operatorname{tr}P_k)^p Z_k^p \leqslant \sigma_p \exp[\varepsilon(1+\operatorname{tr}P_k)] + C'[\log(e+\sigma_p^{-1})]^p Z_k^p \log^p(e+Z_k^p)$$

Hence, in exactly the same way as the proof of (65), from (57) we conclude that

$$\limsup_{n\to\infty} EV_{n+1}^{p/2} \leqslant C'\sigma_p \log^{p+1}(e+\sigma_p^{-1}) \tag{66}$$

Finally, substituting (65) and (66) in (55) and taking account of (56), we obtain

$$\limsup_{n\to\infty} E\|\tilde{\theta}_n\|^p \leqslant C'\sigma_p \log^{3p/2+1}(e+\sigma_p^{-1})$$

This proves the first assertion (19).

We now prove (20) and (21). In this case $Z_k \equiv 0$, so by (57) and (58)

$$V_{n+1}^{p/2} \leqslant \Phi(n+1,0) V_0^{p/2} \tag{67}$$

$$f(V_{n+1}^{p/2}) \leqslant \Phi(n+1,0) f(V_0^{p/2}) \tag{68}$$

Note that (55) is true for any $\sigma_p > 0$ (not necessarily the one defined by (17)). Consequently, by (67), (68) and the dominated convergence theorem we see from (55) that

$$\limsup_{n\to\infty} \mathrm{E}\|\tilde{\theta}_n\|^p \leqslant C'\sigma_p, \quad \sigma_p > 0$$

Hence (20) follows by letting $\sigma_p \to 0$.

For the proof of (21) we take $s > 1$ and $t > 1$ so that $1/s + 1/t + q/p = 1$; then by the Hölder inequality we have

$$\begin{aligned}\mathrm{E}\|\tilde{\theta}_n\|^q &\leqslant \mathrm{E}\|P_n\|^{q/2}V_n^{q/2}\\ &\leqslant \|P_0^{-1}\|^{q/2}\mathrm{E}\|P_n\|^{q/2}\Phi(n,0)\|\tilde{\theta}_0\|^q\\ &\leqslant \|P_0^{-1}\|^{q/2}(\mathrm{E}\|P_n\|^{qs/2})^{1/s}[\mathrm{E}\Phi(n,0)]^{1/t}(\mathrm{E}\|\tilde{\theta}_0\|^p)^{q/p}\end{aligned}$$

Hence by virtue of Corollary 1, (63) and (18), $\mathrm{E}\|\tilde{\theta}_n\|^q$ tends to zero exponentially fast. This completes the proof of Theorem 1. □

Proof of Theorem 2

Let us take $s \in (q, p)$. Then by Corollary 1 and the Hölder inequality we have $(q/s + 1/t = 1)$

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n}\|\tilde{\theta}_k\|^q &\leqslant \limsup_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n}\|P_k\|^{q/2}V_k^{q/2}\\ &\leqslant \limsup_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n}\left(\frac{V_k^{s/2}}{1+\operatorname{tr}P_k}\right)^{q/s}(1+\operatorname{tr}P_k)^{(1/2+1/s)q}\\ &\leqslant C'\left(\limsup_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n}\frac{V_k^{s/2}}{1+\operatorname{tr}P_k}\right)^{q/s}\end{aligned}$$

where C' is a constant.

Summing both sides of (47) leads to

$$\sum_{k=0}^{n}\frac{V_k^{s/2}}{1+\operatorname{tr}P_k} \leqslant \mu^{-1}V_0^{s/2} + \mu^{-1}C\sum_{k=0}^{n}(1+\operatorname{tr}P_k)^s Z_k^s$$

Then, applying the Hölder inequality $(1/u + s/p = 1)$, we see that

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n}\frac{V_k^{s/2}}{1+\operatorname{tr}P_k} \leqslant C'\left(\limsup_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n}Z_k^p\right)^{s/p}$$

Hence from this, (69) and condition (23) it follows that

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n}\|\tilde{\theta}_k\|^q \leqslant C'(\varepsilon_p)^{q/p}, \quad \text{a.s.}$$

This proves (24).

To prove (25), we first note that by (38) of Lemma 4

$$\limsup_{n\to\infty}\frac{1}{n}\exp(\varepsilon\|P_n\|) < \infty, \quad \text{a.s.}$$

and hence

$$\limsup_{n\to\infty}\frac{\|P_n\|}{\log n} < \infty, \quad \text{a.s.} \tag{70}$$

Therefore from this and (64) we see that

$$\| \tilde{\theta}_n \|^2 \leqslant V_n \| P_n \| \leqslant \Phi(n,0) V_0 \| P_n \|$$

tends to zero exponentially fast. Hence the proof of Theorem 2 is complete. □

ACKNOWLEDGEMENTS

Particular thanks are due to the referee for his detailed and valuable comments.

This work was supported by the National Natural Science Foundation of China.

APPENDIX

Proof of Examples 1 and 2

Proof of Example 1. From Reference 19 it is shown that

$$\left\| \mathrm{E}\left(\frac{\varphi_{m+h}\varphi_{m+h}^{\mathrm{T}}}{1+\|\varphi_{m+h}\|^2} \,\middle|\, \mathscr{F}_m \right) - \mathrm{E}\, \frac{\varphi_{m+h}\varphi_{m+h}^{\mathrm{T}}}{1+\|\varphi_{m+h}\|^2} \right\| \leqslant 2r\phi(h), \quad \forall m \geqslant 0 \tag{71}$$

We will need the following inequality:[18]

$$\lambda_{\min}\left(\mathrm{E}\, \frac{\varphi_{m+h}\varphi_{m+h}^{\mathrm{T}}}{1+\|\varphi_{m+h}\|^2} \right) \geqslant \frac{\{\lambda_{\min}[\mathrm{E}(\varphi_{m+h}\varphi_{m+h}^{\mathrm{T}})]\}^2}{\mathrm{E}(\|\varphi_{m+h}\|^2 + \|\varphi_{m+h}\|^4)} \tag{72}$$

For the proof of this let us denote x as the unit eigenvalue corresponding to

$$\lambda_{\min}\left(\mathrm{E}\, \frac{\varphi_{m+h}\varphi_{m+h}^{\mathrm{T}}}{1+\|\varphi_{m+h}\|^2} \right)$$

Then by the Schwarz inequality we see that

$$\begin{aligned} \{\lambda_{\min}[\mathrm{E}(\varphi_{m+h}\varphi_{m+h}^{\mathrm{T}})]\}^2 &\leqslant \left(\mathrm{E}\, \frac{|x^{\mathrm{T}}\varphi_{m+h}|}{(1+\|\varphi_{m+h}\|^2)^{1/2}} \|\varphi_{m+h}\| (1+\|\varphi_{m+h}\|^2)^{1/2} \right)^2 \\ &\leqslant \mathrm{E}\, \frac{x^{\mathrm{T}}\varphi_{m+h}\varphi_{m+h}^{\mathrm{T}}x}{1+\|\varphi_{m+h}\|^2}\, \mathrm{E}[\|\varphi_{m+h}\|^2(1+\|\varphi_{m+h}\|^2)]. \end{aligned}$$

Hence (72) is true. Therefore by (i), (11) and (71) we see that there is a constant $\alpha > 0$ such that

$$\mathrm{E}\left(\frac{\varphi_{m+h}\varphi_{m+h}^{\mathrm{T}}}{1+\|\varphi_{m+h}\|^2} \,\middle|\, \mathscr{F}_m \right) \geqslant \alpha I$$

holds for all $m \geqslant 0$ and large h. This shows that (7) and hence Condition 1 are true.

Proof of Example 2. Let us denote $\mathscr{F}_m = \sigma\{x_0, \xi_i, \zeta_i, i \leqslant n\}$. Similar to (72), it is not difficult to see that

$$\lambda_{\min}\left[\mathrm{E}\left(\frac{\varphi_{m+h}\varphi_{m+h}^{\mathrm{T}}}{1+\|\varphi_{m+h}\|^2} \,\middle|\, \mathscr{F}_m \right) \right] \geqslant \frac{\{\lambda_{\min}[\mathrm{E}(\varphi_{m+h}\varphi_{m+h}^{\mathrm{T}} \mid \mathscr{F}_m)]\}^2}{\mathrm{E}[(\|\varphi_{m+h}\|^2 + \|\varphi_{m+h}\|^4) \mid \mathscr{F}_m]} \tag{73}$$

We now proceed to estimate respectively the numerator and denominator on the RHS of (73). From (12) and (13) it follows that

$$\varphi_{m+h} = CA^h x_m + \sum_{i=m+1}^{m+h} CA^{m+h-i}B\xi_i + \zeta_{m+h} \tag{74}$$

From this, (14) and (15) we then have

$$\mathrm{E}(\varphi_{m+h}\varphi_{m+h}^{\mathrm{T}} \mid \mathscr{F}_m) \geqslant \sum_{i=m+1}^{m+h} CA^{m+h-i}B\mathrm{E}(\xi_i\xi_i^{\mathrm{T}})(CA^{m+h-i}B)^{\mathrm{T}}$$

$$\geqslant \varepsilon \sum_{i=0}^{h-1} CA^iB(CA^iB)^{\mathrm{T}} \geqslant \alpha I > 0, \quad \forall m \geqslant 0, \quad \forall h \geqslant n \tag{75}$$

where the last inequality is derived by the output controllability of (A, B, C) and α is a constant.

Since A is a stable matrix, there exists a norm $\|\cdot\|_1$ on R^n such that its induced matrix norm (also denoted by $\|\cdot\|_1$) satisfies $\|A\|_1 \triangleq a < 1$. Consequently, by (12) we have

$$\|x_{m+1}\|_1 \leqslant a\|x_m\|_1 + \|B\xi_{m+1}\|_1$$

$$\|x_{m+1}\|_1^i \leqslant a\|x_m\|_1^i + (1-a)\left\|\frac{B\xi_{m+1}}{1-a}\right\|_1^i$$

since x^i is convex for $x \geqslant 0$ and $i \geqslant 1$. From this it is easy to find a constant $b \geqslant 0$ such that

$$\|x_{m+1}\|_1^2 + \|x_{m+1}\|_1^4 \leqslant a(\|x_m\|_1^2 + \|x_m\|_1^4) + b(\|\xi_{m+1}\|_1^2 + \|\xi_{m+1}\|_1^4) \tag{76}$$

Again from (74) and (14)–(16) it is easy to see that there exists a constant $d > 0$ such that

$$\mathrm{E}(\|\varphi_{m+1}\|^2 + \|\varphi_{m+1}\|^4 \mid \mathscr{F}_m) \leqslant d(\|x_m\|^2 + \|x_m\|^4) + d$$

Consequently, by the equivalency of the norms $\|\cdot\|$ and $\|\cdot\|_1$ we know that there is $d_1 > 0$ such that

$$\mathrm{E}(\|\varphi_{m+1}\|^2 + \|\varphi_{m+1}\|^4 \mid \mathscr{F}_m) \leqslant d_1(\|x_m\|_1^2 + \|x_m\|_1^4) + d \tag{77}$$

Set

$$\alpha_m = \alpha^{-1}d_1(\|x_m\|_1^2 + \|x_m\|_1^4) + \alpha^{-1}d$$

Then from (73), (75) and (77) we see that (8) holds, while (9) and (10) can be seen from (76) and (16). This completes the proof. □

Proof of Lemmas 1 and 2

Let us first construct an adapted process $\{\beta_m, \mathscr{F}_m\}$ such that $\beta_m \geqslant 1$,

$$\mathrm{E}(a_{m+1} \mid \mathscr{F}_m) \geqslant 1/\beta_m, \quad \beta_m \geqslant 1 \tag{78}$$

$$\beta_{m+1} = b\beta_m + \xi_{m+1}, \quad 0 < b < 1, \qquad \mathrm{E}\beta_0^{1+\delta} < \infty \tag{79}$$

and $\{\xi_k, \mathscr{F}_k\}$ is an adapted sequence satisfying

$$\xi_k \geqslant 0, \qquad \sup_{k\geqslant 0} \mathrm{E}\xi_{k+1}^{1+\delta} < \infty \tag{80}$$

$$N \triangleq \sup_{k \geqslant 0} \mathrm{E}(\xi_{k+1} \mid \mathscr{F}_k) \leqslant b \tag{81}$$

The property (81) characterizes the key difference between $\{\eta_k\}$ and $\{\xi_k\}$.

We proceed as follows.

Define a constant L by

$$L = \max[1, (2M)^{1/\delta}] \tag{82}$$

where δ and M are the same as in (10). Then we take b as

$$b = \frac{L + (1-a)/2}{L + (1-a)} \tag{83}$$

where a is the same as in (9). Obviously, $b \in (0, 1)$ and $b > a$.

Next let us introduce two processes $\{\alpha_k^{(1)}\}$ and $\{\alpha_k^{(2)}\}$ as follows:

$$\begin{aligned} \alpha_{k+1}^{(1)} &= a\alpha_k^{(1)} + \eta_{k+1}I(\eta_{k+1} < L), \qquad & \alpha_0^{(1)} &= 0 \\ \alpha_{k+1}^{(2)} &= b\alpha_k^{(2)} + \eta_{k+1}I(\eta_{k+1} \geqslant L), \qquad & \alpha_0^{(2)} &= \alpha_0 \end{aligned} \tag{84}$$

Obviously, $\alpha_k^{(1)} \leqslant L/(1-a)$, $\forall k \geqslant 0$, and by (9) and the fact that $b \geqslant a$ we have

$$\alpha_k \leqslant \alpha_k^{(1)} + \alpha_k^{(2)} \leqslant \frac{L}{1-a} + \alpha_k^{(2)}$$

Now define $\beta_k = L/(1-a) + \alpha_k^{(2)}$. Then $\alpha_k \leqslant \beta_k$, $\beta_k \geqslant 1$ and hence (78) holds. Furthermore, from (84) it follows that

$$\beta_{k+1} = b\beta_k + \xi_{k+1}$$

where $\xi_{k+1} = (1-b)L/(1-a) + \eta_{k+1}I(\eta_{k+1} \geqslant L)$; hence (79) is true and we need only to verify that $\{\xi_k\}$ satisfies (80) and (81).

By (10) and (82) we know that

$$\mathrm{E}[\eta_{k+1}I(\eta_{k+1} \geqslant L) \mid \mathscr{F}_k] \leqslant \mathrm{E}\left(\frac{\eta_{k+1}^{1+\delta}}{L^\delta} \,\middle|\, \mathscr{F}_k\right) \leqslant \frac{M}{L^\delta} < \frac{1}{2}$$

and by (83)

$$b = \frac{1}{2} + \frac{(1-b)L}{1-a}$$

and hence (81) holds, while (80) is obvious. This proves (78)–(81).

We are now in a position to prove Lemmas 1 and 2.

Proof of Lemma 1. Let us define a sequence $\{x_k\}$, $k \in [m, n]$, by

$$x_{k+1} = (1 - a_{k+1})x_k, \qquad x_m = 1 \tag{85}$$

Then by (78)–(81) and (85) it follows that

$$\begin{aligned}
\mathrm{E}(\beta_{k+1}x_{k+1} \mid \mathscr{F}_k) &\leqslant \mathrm{E}[(b\beta_k + \xi_{k+1})(1-a_{k+1})x_k \mid \mathscr{F}_k] \\
&\leqslant b\beta_k[1 - \mathrm{E}(a_{k+1} \mid \mathscr{F}_k)]x_k + x_k\mathrm{E}(\xi_{k+1} \mid \mathscr{F}_k) \\
&\leqslant b\beta_k(1 - 1/\beta_k)x_k + Nx_k \\
&= b\beta_k x_k + (N-b)x_k \leqslant b\beta_k x_k
\end{aligned}$$

Consequently, we have

$$\mathrm{E}(\beta_{n+1}x_{n+1}) \leqslant b\mathrm{E}(\beta_n x_n) \leqslant \cdots \leqslant b^{n-m+1}\mathrm{E}(\beta_m x_m) = b^{n-m+1}\mathrm{E}\beta_m \leqslant b^{n-m+1}\left(\mathrm{E}\beta_0 + \frac{b}{1-b}\right) \tag{86}$$

where for the last inequality we have (79) and (81).

Finally, by (85) and (86) and noting that $\beta_n \geqslant 1$, we obtain

$$\mathrm{E}\prod_{k=m}^{n}(1-a_{k+1}) = \mathrm{E}x_{n+1} \leqslant \mathrm{E}(\beta_{n+1}x_{n+l}) \leqslant \left(\mathrm{E}\beta_0 + \frac{b}{1-b}\right)b^{n-m+1}$$

Hence Lemma 1 holds. □

Proof of Lemma 2. We will need the following fact: for any martingale difference sequence $\{f_n, \mathscr{F}_n\}$, if

$$\sup_n \mathrm{E}|f_n|^{1+\varepsilon} < \infty, \quad \text{for some } \varepsilon > 0 \tag{87}$$

then as $n \to \infty$

$$\frac{1}{n}\sum_{k=0}^{n} f_k \to 0, \quad \text{a.s.} \tag{88}$$

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

$$\mathrm{E}\left[\sum_{k=1}^{\infty}\mathrm{E}\left(\left|\frac{f_k}{k}\right|^{1+\varepsilon} \,\middle|\, \mathscr{F}_{k-1}\right)\right] = \sum_{k=1}^{\infty}\frac{\mathrm{E}|f_k|^{1+\varepsilon}}{k^{1+\varepsilon}} < \infty$$

so we have

$$\sum_{k=1}^{\infty}\mathrm{E}\left(\left|\frac{f_k}{k}\right|^{1+\varepsilon} \,\middle|\, \mathscr{F}_{k-l}\right) < \infty, \quad \text{a.s.}$$

Consequently, by Corollary 5 of Reference 21 we know that the series $\sum_{k=1}^{\infty} f_k/k$ converges almost surely. From this and the Kronecker lemma we see that (88) is true.

Now let us denote

$$\tilde{a}_k = a_k - \mathrm{E}(a_k \mid \mathscr{F}_{k-1}), \qquad \tilde{\xi}_k = \xi_k - \mathrm{E}(\xi_k \mid \mathscr{F}_{k-1})$$

where $\{\xi_k\}$ is the process appearing in (79)–(81). By (78) and (81) it is easy to see that

$$a_{k+1} \geqslant \tilde{a}_{k+1} + 1/\beta_k, \qquad \xi_{k+1} \leqslant \tilde{\xi}_{k+1} + b \tag{89}$$

By this, (28) and (79) we have

$$\begin{aligned}\beta_{k+1}x_{k+1} &\leqslant (b\beta_k + \xi_{k+1})(1 - a_{k+1})\alpha x_k + C\beta_{k+1} \\ &\leqslant b\beta_k(1 - a_{k+1})\alpha x_k + \xi_{k+1}\alpha x_k + C\beta_{k+1} \\ &\leqslant b\beta_k(1 - \tilde{a}_{k+1} - 1/\beta_k)\alpha x_k + (\tilde{\xi}_{k+1} + b)\alpha x_k + C\beta_{k+1} \\ &= \alpha b x_k\beta_k - \alpha b\tilde{a}_{k+1}x_k\beta_k + \tilde{\xi}_{k+1}\alpha x_k + C\beta_{k+l}\end{aligned} \tag{90}$$

We now proceed to estimate the last three terms on the RHS of (90). By (79) and (80) it is easy to see that

$$\sup_k \mathrm{E}\beta_k^{1+\delta} < \infty \tag{91}$$

By Lemma 1 and (28) it is easy to verify that

$$\sup_k \mathrm{E}(x_k)^q < \infty, \quad \forall q > 0 \tag{92}$$

provided that $\alpha < \min[1/\gamma, (1/\gamma)^{1/q}]$, where $\gamma \in (0, 1)$ is the constant appearing in (27).

Next let us take $\varepsilon \in (0, \delta)$, $q = (1+\delta)(1+\varepsilon)/(\delta - \varepsilon)$ and define $\alpha^* = \min[1/b, 1/\gamma, (1/\gamma)^{1/q}]$. By (91), (92) and the Hölder inequality we get ($|\tilde{a}_k| \leqslant 1$)

$$\sup_k \mathrm{E}\,|\tilde{a}_{k+1}x_k\beta_k|^{1+\varepsilon} \leqslant \sup_k \mathrm{E}(x_k\beta_k)^{1+\varepsilon} \leqslant \sup_k(\mathrm{E}\beta_k^{1+\delta})^{(1+\varepsilon)/(1+\delta)}(\mathrm{E}x_k^q)^{(\delta-\varepsilon)/(1+\delta)} < \infty$$

Similarly, we have

$$\sup_k \mathrm{E}\,|\tilde{\xi}_{k+1}x_k|^{1+\varepsilon} < \infty$$

Hence, applying (87), we obtain

$$\frac{1}{n}\sum_{k=1}^{n}(-\alpha b\tilde{a}_{k+1}x_k\beta_k + \alpha\tilde{\xi}_{k+1}x_k) \to 0, \quad \text{a.s.,} \quad \text{as } n \to \infty \tag{93}$$

Summing both sides of (79) and taking account of (89), we see that

$$\sum_{k=1}^{n}\beta_k \leqslant b\beta_0 + b\sum_{k=1}^{n}\beta_k + \sum_{k=1}^{n}\tilde{\xi}_k + bn \tag{94}$$

But by (80) and (88) we know that

$$\frac{1}{n}\sum_{k=1}^{n}\tilde{\xi}_k \to 0, \quad \text{a.s.,} \quad \text{as } n \to \infty$$

Hence by (94)

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\beta_k \leqslant \frac{b}{1-b}, \quad \text{a.s.} \tag{95}$$

Finally, summing both sides of (90), we get

$$\sum_{k=1}^{n}x_k\beta_k \leqslant \alpha b\sum_{k=1}^{n}x_k\beta_k + \alpha b x_0\beta_0 + \sum_{k=0}^{n-1}(-\alpha b\tilde{a}_{k+1}x_k\beta_k + \alpha\tilde{\xi}_{k+1}x_k) + C\sum_{k=1}^{n}\beta_k$$

Consequently, taking account of (93) and (95), we have

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}x_k\beta_k \leqslant \frac{C}{(1-b)(1-\alpha b)}, \quad \text{a.s.,} \quad \forall\alpha \in \left(1, \frac{1}{b}\right)$$

which implies the desired result (29) with $\alpha^* = 1/b$ and $C' = C/(1-b)(1-\alpha b)$ since $\beta_k \geqslant 1$, $\forall k$. □

Proof of Lemmas 6 and 8

Proof of Lemma 6. Without loss of generality we assume that

$$x_{k+1} = (1 - a_{k+1})x_k + C, \quad \forall k \geqslant 0, \qquad \mathrm{E}x_0^2 < \infty \tag{96}$$

We first prove (50) for the case $C \leqslant 1$.
Define a sequence $\{y_k, k \in [m, n]\}$ by

$$y_k = \left(1 - \frac{1}{x_k}\right) y_{k-1}, \qquad y_{m-1} = 1 \tag{97}$$

Then we have

$$x_k y_k = (x_k - 1) y_{k-1} \leqslant (1 - a_k) x_{k-1} y_{k-1}$$

and then

$$y_n \leqslant x_n y_n \leqslant \prod_{k=m}^{n} (1 - a_k) x_{m-1} \tag{98}$$

Note that by Lemma 1 and (96)

$$\sup_k \mathrm{E}x_k^2 < \infty$$

Hence by this, Lemma 1 and the Schwarz inequality we obtain

$$\mathrm{E} \prod_{k=m}^{n} \left(1 - \frac{1}{x_k}\right) = \mathrm{E}y_n \leqslant C'\gamma^{(n-m+1)/2}, \quad \forall n \geqslant m, \quad C' > 0, \quad \gamma \in (0,1)$$

which shows that (50) is valid.
Next we consider the case where $C > 1$.
Let us take $\varepsilon = C^{-1}$; then $\varepsilon \in (0,1)$ and

$$\varepsilon x_{k+1} = (1 - a_{k+1})\varepsilon x_k + 1, \quad \varepsilon x_k \geqslant 1, \quad k \geqslant 1 \tag{99}$$

Hence by the argument above we know that

$$\mathrm{E} \prod_{k=m}^{n} \left(1 - \frac{1}{\varepsilon x_k}\right) \leqslant C'\gamma^{(n-m+1)/2}, \quad \forall n \geqslant m \tag{100}$$

where $C' > 0$ and $\gamma \in (0,1)$.
Since $a_k \in [0, \bar{a}], \bar{a} < 1$, we see from (99) that

$$\varepsilon x_{k+1} \geqslant (1 - \bar{a}) + 1, \quad \forall k \geqslant 1$$

We need the following inequality:[18]

$$1 - x \leqslant (1 - dx)^{(1-t)/d}, \quad 0 \leqslant dx \leqslant t < 1, \quad d > 1$$

To see this, let $f(x) = x + (1 - dx)^{(1-t)/d} - 1$. Then $f(0) = 0$ and the derivative of $f(x)$ satisfies

$$f'(x) = 1 - (1-t)(1-dx)^{(1-t)/d-1} \geqslant 1 - (1-t)(1-t)^{(1-t)/d-1} = 1 - (1-t)^{(1-t)/d} \geqslant 0 \tag{101}$$

Hence (101) holds. Now, substituting $x = 1/x_k$, $d = 1/\varepsilon$ and $t = 1/(2 - \bar{a})$ in (101), we have

$$1 - \frac{1}{x_k} \leqslant \left(1 - \frac{1}{\varepsilon x_k}\right)^{(1-\bar{a})\varepsilon/(2-\bar{a})}, \quad \forall k \geqslant 2$$

Consequently, by the Hölder inequality and (100) we get

$$\begin{aligned} \mathrm{E} \prod_{k=m}^{n} \left(1 - \frac{1}{x_k}\right) &\leqslant \left[\mathrm{E} \prod_{k=m}^{n} \left(1 - \frac{1}{\varepsilon x_k}\right)\right]^{(1-\bar{a})\varepsilon/(2-\bar{a})} \\ &\leqslant C'\lambda^{n-m+1}, \quad \text{for some } C' > 0, \quad \gamma \in (0,1), \quad \forall n \geqslant m \geqslant 2 \end{aligned}$$

By suitably adjusting C' we know that the above inequality holds for all $n \geqslant m \geqslant 0$. Hence the proof of Lemma 6 is complete. □

Proof of Lemma 8. Let us take a constant $d > 1$ such that $d\lambda < 1$. Then we have

$$\begin{aligned} \mathrm{E}(C_{nk}x_k) &= \mathrm{E}(C_{nk}x_k)I(x_k \leqslant \sigma d^{n-k}) + \mathrm{E}(C_{nk}x_k)I(x_k > \sigma d^{n-k}) \\ &\leqslant \sigma N(d\lambda)^{n-k} + \mathrm{E}\,\frac{x_k \log^{\beta}(e+x_k)}{\log^{\beta}(e+\sigma d^{n-k})} \\ &\leqslant \sigma N(d\lambda)^{n-k} + \frac{\sigma}{\log^{\beta}(e+\sigma d^{n-k})} \end{aligned}$$

If $\sigma \geqslant 1$, then it is easy to see that

$$\sum_{k=0}^{n} \frac{1}{\log^{\beta}(e+\sigma d^{n-k})} \leqslant C', \quad \forall n \geqslant 0 \tag{103}$$

where C' is a constant independent of σ.

If $\sigma < 1$, define k_0 as the smallest integer such that $\sigma d^{k_0} \geqslant 1$; then it is easy to see that $k_0 \leqslant (\log d)^{-1} \log(e+\sigma^{-1}) + 1$. Thus we have

$$\sum_{k=0}^{n} \frac{1}{\log^{\beta}(e+\sigma d^{n-k})} = \sum_{k=0}^{n-k_0+1} \frac{1}{\log^{\beta}(e+\sigma d^{n-k})} + \sum_{k=n-k_0}^{n} \frac{1}{\log^{\beta}(e+\sigma d^{n-k})} \leqslant C''[1+\log(e+\sigma^{-1})] \tag{104}$$

where C'' is some constant independent of σ.

Combining (102)–(104), we finally get

$$\sum_{k=0}^{n} \mathrm{E}(C_{nk}x_k) \leqslant \frac{N}{1-d\lambda}\sigma + \sigma(C'+C'') + C''\sigma\log(e+\sigma^{-1}) \leqslant C\sigma\log(e+\sigma^{-1})$$

for some constant C. This completes the proof of Lemma 8. □

Proof of equation (52)

Take $\varphi(t) = \exp(t^{1/\alpha}) - 1$, $\psi(t) = \log^{\alpha}(t+1)$ in the following Young's inequality (see e.g. Reference 22):

$$xy \leqslant \int_0^x \varphi(t)\,\mathrm{d}t + \int_0^y \psi(t)\,\mathrm{d}t, \quad x \geqslant 0, \quad y \geqslant 0, \qquad \varphi(\psi(t)) = t$$

We get

$$xy \leqslant x\exp(x^{1/\alpha}) + y\log^{\alpha}(1+y)$$

Note that

$$x \leqslant \exp(\alpha x^{1/\alpha}), \quad \forall x \geqslant 0, \quad \alpha > 0$$

So we have

$$xy \leqslant \exp[(\alpha+1)x^{1/\alpha}] + y\log^{\alpha}(1+y)$$

Replacing x and y in the above inequality by

$$\left(\frac{\varepsilon x}{\alpha+1}\right)^{\alpha} \quad \text{and} \quad \left(\frac{\alpha+1}{\varepsilon}\right)^{\alpha}\frac{y}{\sigma_p}$$

respectively, we see that

$$\begin{aligned} x^{\alpha}y &= \sigma_p\left[\left(\frac{\varepsilon x}{\alpha+1}\right)^{\alpha}\left(\frac{\alpha+1}{\varepsilon}\right)^{\alpha}\frac{y}{\sigma_p}\right] \\ &\leqslant \sigma_p\exp(\varepsilon x) + \left(\frac{\alpha+1}{\varepsilon}\right)^{\alpha} y\log^{\alpha}\left[e + \left(\frac{\alpha+1}{\varepsilon}\right)^{\alpha}\frac{y}{\sigma_p}\right] \end{aligned}$$

$$\leqslant \sigma_p \exp(\varepsilon x) + \left(\frac{\alpha+1}{\varepsilon}\right)^{\alpha} y \log^{\alpha}\left\{\left[e + \left(\frac{\alpha+1}{\varepsilon}\right)^{\alpha}\right]\left(e + \frac{y}{\sigma_p}\right)\right\}$$

$$\leqslant \sigma_p \exp(\varepsilon x) + \left(\frac{\alpha+1}{\varepsilon}\right)^{\alpha} y 2^{\alpha}\left\{\log^{\alpha}\left[e + \left(\frac{\alpha+1}{\varepsilon}\right)^{\alpha}\right] + \log^{\alpha}\left(e + \frac{y}{\sigma_p}\right)\right\}$$

$$\leqslant \sigma_p \exp(\varepsilon x) + \left(\frac{\alpha+1}{\varepsilon}\right)^{\alpha} 2^{\alpha+1} y \log^{\alpha}\left[e + \left(\frac{\alpha+1}{\varepsilon}\right)^{\alpha}\right]\log^{\alpha}\left(e + \frac{y}{\sigma_p}\right)$$

$$\leqslant \sigma_p \exp(\varepsilon x) + \left(\frac{\alpha+1}{\varepsilon}\right)^{\alpha} 4^{\alpha+1} y \log^{\alpha}\left[e + \left(\frac{\alpha+1}{\varepsilon}\right)^{\alpha}\right][\log^{\alpha}(e + \sigma_p^{-1}) + \log^{\alpha}(e + y)]$$

Thus completing the proof. □

REFERENCES

1. Widrow, B., and S. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1985.
2. Benveniste, A., M. Metivier, and P. Priouret, *Algorithmes Adaptifs et Approximations Stochastiques*, Masson, Paris, 1987.
3. Eweda, E., and O. Macchi, 'Tracking error bounds of adaptive nonstationary filtering', *Automatica*, 293–302 (1985).
4. Macchi, O., 'Optimization of adaptive identification for time-varying filters', *IEEE Trans. Automatic Control*, **AC-31**, 283–287 (1986).
5. Benveniste, A., 'Design of adaptive algorithms for the tracking of time-varying systems', *Int. j. adaptive control signal processing*, **1**, 3–29 (1987).
6. Gunarsson, S., and L. Ljung, 'Frequency domain tracking characteristics adaptive algorithms', *IEEE Trans. Acoustics, Speech, Signal Processing,* **ASSP-37**, 1072–1089 (1989).
7. Solo, V., 'The limit behavior of LMS', *IEEE Trans, Acoustics, Speech, Signal Processing*, **37**, 1909–1922 (1989).
8. Ljung, L., and T. Söderström, *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, MA, 1983.
9. Caines, P. E., *Linear Stochastic Systems*, Wiley, New York, 1988.
10. Chen, H. F., *Recursive Estimation and Control for Stochastic Systems*, Wiley, New York, 1985.
11. Kitagava, G., and W. Gersh, 'A smoothness priors time varying AR coefficient modelling of nonstationary covariance time series', *IEEE Trans. Automatic Control*, **AC-30**, 48–56 (1985).
12. Benveniste, A., and G. Ruget, 'A measure of the tracking capability of recursive stochastic algorithms with constant gains', *IEEE Trans. Automatic Control*, **AC-27**, 639–649 (1982).
13. Moser, A. T., and D. Graupe, 'Identification of nonstationary models with application to myoelectric signals for controlling electrical stimulation of paraplegics', *IEEE Trans. Acoustics, Speech, Signal Processing*, **ASSP-37**, 713–719 (1989).
14. Guo, L., L. G. Xia, and J. B. Moore, 'Tracking randomly varying parameters: analysis of a standard algorithm', *Proc. 27th IEEE CDC*, Austin, TX, December 1988, pp. 1514–1519.
15. Lipster, R. S., and A. N. Shiryayev, *Statistics of Random Processes*, Springer, New York, 1977, Chap. 11.
16. Chen, H. F., P. R. Kumar, and J. H. Van Schuppen, 'On Kalman filtering for conditionally Gaussian systems with random matrices', *Syst. Control Lett.*, **13**. 397–404 (1989).
17. Jazwinski, A. H., *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.
18. Guo, L., 'Estimating time-varying parameters by Kalman filter based algorithm: stability and convergence', *IEEE Trans. Automatic Control*, **AC-35**, 141–147 (1990).
19. Kushner, H. J., *Approximation and Weak Convergence Methods for Random Process with Applications to Stochastic Systems Theory*, MIT Press, Cambridge, MA, 1984, pp. 82–83.
20. Fortmann, T. E., and K. L. Hitz, *An Introduction to Linear Control Systems*, Marcel Dekker, New York and Basel, 1977, pp. 416–417.
21. Chow, Y. C., 'Local convergence of martingales and the law of large numbers', *Ann. Math. Stat.*, **36**, 552–558 (1965).
22. Mitrinović, D. S., *Analytic Inequalities*, Springer, New York, 1970.