SIAM J CONTROL AND OPTIMIZATION
Vol. 32 No 5 pp. 1195–1225 September 1994

001

STABILITY OF RECURSIVE STOCHASTIC TRACKING ALGORITHMS*

LEI GUO†

Abstract. First, the paper gives a stability study for the random linear equation $x_{n+1} = (I - A_n)x_n$. It is shown that for a quite general class of random matrices $\{A_n\}$ of interest, the stability of such a vector equation can be guaranteed by that of a corresponding scalar linear equation, for which various results are given without requiring stationary or mixing conditions Then, these results are applied to the main topic of the paper, i.e. to the estimation of time varying parameters in linear stochastic systems, giving a unified stability condition for various tracking algorithms including the standard Kalman filter least mean squares, and least squares with forgetting factor

Key words. stochastic systems, adaptive systems, parameter estimation tracking algorithms time varying, stability, excitation

AMS subject classifications. 93C40 93E12, 93E10

1. Introduction.

An important issue in system identification, signal processing, adaptive control and many other fields is whether the algorithms designed possess some tracking capabilities when the system parameters (or signals) to be estimated are changing with time. The basic time-varying model is that of a linear regression:

$$y_k = \varphi_k^\tau \theta_k + v_k, \qquad k \ge 0 \tag{1.1}$$

where y_k and v_k are the scalar observation and noise, respectively, and φ_k and θ_k are, respectively, the d-dimensional stochastic regressor and the unknown time-varying parameter. It is usually convenient to denote the parameter variation at instant k by Δ_k:

$$\Delta_k \triangleq \theta_k - \theta_{k-1}, \qquad k \ge 1. \tag{1.2}$$

It is well known that many problems from different application areas can be cast in the form (1.1) (see e.g., [1], [2]), and a variety of recursive algorithms have been derived for tracking the unknown parameters θ_k. These algorithms are basically of the following form:

$$\hat{\theta}_{k+1} = \hat{\theta}_k + L_k(y_k - \varphi_k^\tau \hat{\theta}_k) \tag{1.3}$$

where L_k is the adaptation gain that can be chosen in a number of ways (see e.g., [1]–[3]). In the present time-varying case, a common feature of the gain L_k is that it does not tend to zero as the time k goes to infinity. This is very natural from an intuitive point of view. When the system parameters are time-varying, the algorithm must be persistently alert to follow the parameter variations Here we illustrate three choices of L_k that correspond to three standard algorithms.

Kalman filtering (KF) algorithm.

$$L_k = \frac{P_k \varphi_k}{R + \varphi_k^\tau P_k \varphi_k} \tag{1.4}$$

$$P_{k+1} = P_k - \frac{P_k \varphi_k \varphi_k^\tau P_k}{R + \varphi_k^\tau P_k \varphi_k} + Q, \tag{1.5}$$

* Received by the editors February 3, 1992; accepted for publication February 25, 1993. This work was supported by the National Natural Science Foundation of China

† Institute of Systems Science Chinese Academy of Sciences, Beijing 100080. People's Republic of China

where $P_0 \geq 0, R > 0, Q > 0$ 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 Δ_k, respectively. Taking R and Q as constants is just for simplicity of discussion, and generalizations to time-varying cases are straightforward.

It is well known that (see e.g., [4, Chap. 13] and [5, Chap. 3]) if φ_k is $\mathcal{F}_{k-1}$ measurable, where $\mathcal{F}_k \triangleq \sigma\{y_i, i \leq k\}$, and if $\{\Delta_k, v_k\}$ is a Gaussian white noise process, then $\hat{\theta}_k$ generated by (1.3)–(1.5) is the minimum variance estimate for θ_k, and P_k is the estimation error covariance, i.e.,

$$\hat{\theta}_k = E[\theta_k | \mathcal{F}_{k-1}], \qquad P_k = E[\tilde{\theta}_k \tilde{\theta}_k^\tau | \mathcal{F}_{k-1}] \tag{1.6}$$

provided that $Q = E\Delta_k\Delta_k^\tau$, $R = Ev_k^2$, $\hat{\theta}_0 = E\theta_0$ and $P_0 = E[\tilde{\theta}_0\tilde{\theta}_0^\tau]$, where $\tilde{\theta}_k$ is the estimation error

$$\tilde{\theta}_k = \theta_k - \hat{\theta}_k, \tag{1.7}$$

which is of prime interest to us.

Least mean squares (LMS) algorithm.

$$L_k = \mu \frac{\varphi_k}{1 + \|\varphi_k\|^2}, \tag{1.8}$$

where $\mu \in (0, 1]$ is called the step size or adaptation rate. Such an algorithm is also referred to as a gradient algorithm because the increment of the algorithm (1.3) and (1.8) is opposite to the (stochastic) gradient of the mean square error

$$e_k(\theta) = E(y_k - \varphi_k^\tau \theta)^2.$$

Thus, it is a type of steepest descent algorithm that aims at minimizing $e_k(\theta)$ recursively.

Recursive least squares (RLS) algorithm.

$$L_k = \frac{P_k\varphi_k}{\alpha + \varphi_k^\tau P_k \varphi_k} \tag{1.9}$$

$$P_{k+1} = \frac{1}{\alpha}\left[P_k - \frac{P_k\varphi_k\varphi_k^\tau P_k}{\alpha + \varphi_k^\tau P_k\varphi_k}\right], \tag{1.10}$$

where $P_0 > 0$, and $\alpha \in (0, 1)$ is a forgetting factor. This algorithm is derived by minimizing the following criterion over $\theta \in \mathbb{R}^d$:

$$V_k(\theta) = \frac{1}{k}\sum_{i=0}^{k} \alpha^{k-i}(y_i - \theta^\tau\varphi_i)^2 \tag{1.11}$$

(see e.g., [1], pp. 57–58). Note that in (1.11) old measurements are exponentially discounted, and so the estimate is expected to be representative for the current properties of the system.

All of the above-mentioned algorithms are well known and widely used in applications. The KF algorithm is attractive due to the fact that it generates the conditional expectation of the unknown parameter given the past measurements in the ideal case (see (1.6)). The LMS has been used in many applications, mainly because of its simplicity for implementation. The advantage of the RLS algorithm over LMS is that it generates more accurate estimates in the transient phase (see e.g., [6]). In many cases, the RLS algorithm is optimal in the sense that it minimizes the criterion (1.11), while for the KF algorithm, it is not known if it is still optimal in some sense when the Gaussian assumption fails and the covariances of v_k and Δ_k are not available.

There is a vast literature on the analysis of algorithms of type (1.3). In the area of adaptive signal processing, the LMS algorithm has received a great deal of attention (see e.g., [7]–[12]). Most of the existing analysis require that the signals $\{y_k, \theta_k, \varphi_k\}$ possess some sort of stationarity, independence, or mixing properties. The KF algorithm has also attracted much research attention (e.g., [11], [13]–[15]). The first rigorous stability analysis for KF that allows $\{\varphi_k\}$ to be a large class of stochastic regressors seems to be that in [14]. Finally, for the RLS algorithm, we mention the preliminary works in [6], [16], [17], among many others.

In the related area of stochastic adaptive control, the Kalman filter was used by Meyn and Caines [31] to design the adaptive control law for a first-order stochastic system. By applying the Markov chain ergodic theory, they obtained the first concrete adaptive control result for systems with nontrivial (random) parameter variations. For high-order systems with randomly varying parameters, stability of an LMS-based adaptive minimum variance controller was demonstrated in [30]. Similar results were recently established in [32] for a KF-based model reference adaptive controller. However, the parameter tracking properties of the estimation algorithms are not studied in these papers.

In this paper, we first present a series of stability results on the vector random linear equation $x_{n+1} = (I - A_n)x_n$, where $\{A_n\}$ is a sequence of random matrices of the same dimension, which may not satisfy the usual stationary or mixing conditions. The key observation is that for a variety of $\{A_n\}$ of interest, the stability study of the vector linear equation may be reduced to that of a relatively simple scalar equation. Then we present a stability/excitation condition for recursive stochastic tracking algorithms and establish upper bounds for the tracking error.

The main contributions of the paper are as follows:

(i) The new stability condition is the weakest known and a unified one for the three standard algorithms mentioned above. This is important since establishing stability is known to be a crucial step for any further studies (see e.g., [18]).

(ii) For a large class of random models of interest in applications including time-varying autoregressive models, we can verify the present condition, whereas conditions introduced previously (see e.g., [14], [28]) cannot be verified;

(iii) For the commonly used ϕ-mixing process, we can prove that our stability condition is also a necessary one in some sense.

2. Stability of random equation $x_{n+1} = (I - A_n)x_n$.

2.1. Preliminaries. To begin, by substituting (1.1) into (1.3) and using the notations (1.2) and (1.7), we get the following error equation:

$$\tilde{\theta}_{k+1} = (I - L_k\varphi_k^\tau)\tilde{\theta}_k - L_k v_k + \Delta_{k+1}, \qquad k \geq 0. \tag{2.1}$$

Clearly, this equation falls into the following general form of linear equations:

$$x_{k+1} = (I - A_k)x_k + \xi_{k+1}, \qquad k \geq 0 \tag{2.2}$$

where $\{A_k\}$ is a sequence of $d \times d$ random matrices, and $\{\xi_{k+1}\}$ represents the disturbance. Usually, we are primarily interested in the following problem: does $\{x_k\}$ remain bounded in some sense when $\{\xi_k\}$ belongs to a certain class of random processes? To rigorously study this problem, we need to introduce some notations and definitions.

For any matrix X, its norm is defined as its maximum singular value, i.e. $\|X\| = \{\lambda_{\max}(XX^{\tau})\}^{1/2}$.

DEFINITION 2.1. A random matrix (or vector) sequence $\{A_k, k \geq 0\}$ defined on the basic probability space $(\Omega, \mathcal{F}, P)$ is called L_p-stable $(p > 0)$ if $\sup_{k\geq 0} E\|A_k\|^p < \infty$.

In the sequel, we will refer to $\|A_k\|_{L_p}$ defined by

$$\|A_k\|_{L_p} \triangleq \{E\|A_k\|^p\}^{1/p} \tag{2.3}$$

as the L_p-norm of A_k.

To motivate further discussions, let us consider the following propositions.

PROPOSITION 2.1. *Consider the random equation* (2.2) *with* $x_0 = 0$. *Suppose that* $\{A_k, k \geq 0\}$ *is an independent sequence and* $\det[E(I - A_k)(I - A_k)^{\tau}] \neq 0$. *Then for any* $\{\xi_k\} \in \mathcal{B}, \{x_k\}$ *is* L_2*-stable if and only if there exist two constants* $M > 0$ *and* $\lambda \in [0, 1)$ *such that*

$$\left\| \prod_{j=i+1}^{k} (I - A_j) \right\|_{L_2} \leq M\lambda^{k-i}, \qquad \forall k \geq i, \qquad \forall i \geq 0 \tag{2.4}$$

where $\mathcal{B}$ *is a set of random processes defined by*

$$\mathcal{B} = \{\xi = (\xi_k) : \xi \text{ is } L_2\text{-stable and independent of } \{A_k\}\} \tag{2.5}$$

and where by definition

$$\prod_{j=i+1}^{k} (I - A_j) = \begin{cases} (I - A_k) \cdots (I - A_{i+1}), & k > i; \\ I, & k \leq i. \end{cases} \tag{2.6}$$

The proof is in Appendix A. Obviously, the only nontrivial conclusion in this proposition is that (2.4) is a necessary condition for L_2-stability of $\{x_k\}$. Related results in the deterministic framework may be found in [19]. We remark that when the independence assumptions are removed, similar necessity results are also true. This is the content of Proposition 2.2.

PROPOSITION 2.2. *Consider the random equation* (2.2) *with* $x_0 = 0$. *Assume that* $(I - A_k)^{-1}$ *exists for any* $k \geq 0$. *Denote*

$$\mathcal{B}^0 = \{\xi : \sup_k \|\xi_k\|_{L_2} \leq 1\}; \tag{2.7}$$

then the following property also implies (2.4):

$$\sup_{\xi \in \mathcal{B}^0} \sup_k \|x_k\|_{L_2} < \infty. \tag{2.8}$$

The proof is also given in Appendix A. These two propositions indicate that (2.4) is in some sense the necessary (and also sufficient) condition for the stability of $\{x_k\}$ generated by (2.2). This prompts us to introduce the following definition.

DEFINITION 2.2. A sequence of $d \times d$ random matrices $A = \{A_k\}$ is called stably exciting of order p, $(p \geq 1)$ with parameter $\lambda \in [0, 1)$, if it belongs to the following set

$$\mathcal{S}_p(\lambda) = \left\{ A : \left\| \prod_{j=i+1}^{k} (I - A_j) \right\|_{L_p} \leq M\lambda^{k-i}, \forall k \geq i, \forall i \geq 0, \text{ for some } M > 0 \right\} \tag{2.9}$$

The investigation of products of random matrices has a long history (see e.g., [20]–[26] and the references therein), and almost all of the existing results rely on some stationary or mixing assumptions on the random coefficients. In particular, in [21] and [22] a time-invariant quadratic Lyapunov function was used to analyze the stability of a random linear differential equation under stationary and ergodic assumptions on the coefficients, while in [24] and [26] it was shown that under some mixing conditions, the stability of a random linear differential equation may be guaranteed by that of a corresponding "averaged" deterministic equation.

However, in general, stationary or mixing conditions cannot be directly imposed on the random coefficients in the study of tracking algorithms. Our treatment here is based on the observation that for a quite large class of matrix sequence $\{A_k\}$ of interest in applications, the study of its stably exciting property may be reduced to that of a certain class of scalar sequences. For convenience of discussion, we introduce the following subclass of $\mathcal{S}_1(\lambda)$ for scalar sequence $a = (a_k, k \geq 0)$:

$$\mathcal{S}^0(\lambda) = \left\{ a : a_k \in [0,1], E \prod_{j=i+1}^{k} (1 - a_j) \leq M\lambda^{k-i}, \forall k \geq i, \forall i \geq 0, \text{ for some } M > 0 \right\} \tag{2.10}$$

where $\lambda \in [0, 1)$ is a parameter reflecting the stability margin. Note that for λ given above, $\log \lambda$ is related to the familiar concept of Lyapunov exponent (cf. [25]), and its absolute value is proportional to the exciting extent of $\{a_k\}$

Clearly, for any constant $c \in (0,1], \{c\} \in \mathcal{S}^0(1-c)$, and if $0 \leq \alpha_k \leq \beta_k \leq 1$ and $\{\alpha_k\} \in \mathcal{S}^0(\lambda)$, then $\{\beta_k\} \in \mathcal{S}^0(\lambda)$.

LEMMA 2.1. *Let $\alpha = \{\alpha_k, \mathcal{F}_k\}$ and $a = \{a_k, \mathcal{F}_k\}$ be adapted processes, such that*

$$a_k \in [0,1], \qquad E[a_{k+1}|\mathcal{F}_k] \geq \alpha_k, \qquad k \geq 0.$$

Then $\alpha \in \mathcal{S}^0(\lambda)$ implies that $a \in \mathcal{S}^0(\sqrt{\lambda})$.

Proof. We first assume that $0 \leq \alpha_k < 1$. For any $n > m, k \in [m, n]$, set

$$A_k = \left\{ \prod_{i=m}^{k} (1 - \alpha_i) \right\}^{-1}, \qquad A_{m-1} = 1$$
$$x_{k+1} = (1 - a_{k+1})x_k, \qquad x_m = 1.$$

Then

$$x_{n+1} = \prod_{i=m}^{n} (1 - a_{i+1}).$$

Note that

$$\begin{aligned} EA_k x_{k+1} &= EA_k[1 - E(a_{k+1}|\mathcal{F}_k)]x_k \\ &\leq EA_k(1 - \alpha_k)x_k = EA_{k-1}x_k. \end{aligned}$$

Hence

$$EA_n x_{n+1} \leq EA_{n-1}x_n \leq \cdots \leq EA_{m-1}x_m = 1.$$

Consequently,

$$
\begin{aligned}
E\prod_{i=m}^{n}(1-a_{i+1}) &= Ex_{n+1} \le E\sqrt{x_{n+1}} \\
&= E\sqrt{x_{n+1}A_n}\sqrt{A_n^{-1}} \le \sqrt{E(x_{n+1}A_n)EA_n^{-1}} \le \sqrt{EA_n^{-1}} \\
&\le \left\{E\prod_{i=m}^{n}(1-\alpha_i)\right\}^{1/2} \le \sqrt{M}(\sqrt{\lambda})^{n-m+1}.
\end{aligned}
$$

Hence $a \in \mathcal{S}^0(\sqrt{\lambda})$.

Next, we consider the general case $\alpha_k \in [0, 1]$ By the monotonic convergence theorem, it is known that

$$\lim_{\varepsilon\to 1^-} E\prod_{k=m}^{n}(1-\varepsilon\alpha_k) \le M\lambda^{n-m+1}.$$

Hence there exists $0 < \varepsilon^* < 1$ such that for any $\varepsilon \in (\varepsilon^*, 1)$,

$$E\prod_{k=m}^{n}(1-\varepsilon\alpha_k) \le 2M\lambda^{n-m+1}.$$

Hence by $\varepsilon\alpha_k \in (0, 1)$ and the fact proved above we have

$$E\prod_{k=m}^{n}(1-\varepsilon a_{k+1}) \le \sqrt{2M}(\sqrt{\lambda})^{n-m+1}.$$

Thus, by noticing that $\varepsilon a_{k+1} \le a_{k+1}$, we have $a \in \mathcal{S}^0(\sqrt{\lambda})$. This completes the proof. □

LEMMA 2.2. *Let $\{\alpha_k, \mathcal{F}_k\}$ be an adapted process, $\alpha_k \in [0,1]$. If for some integer $h > 0$, $\{E[\alpha_{k+h}|\mathcal{F}_k]\} \in \mathcal{S}^0(\lambda)$, then $\{\alpha_k\} \in \mathcal{S}^0(\lambda^{2^{-h}})$.*

Proof. Set $a_k = E[\alpha_{k+h-1}|\mathcal{F}_k]$ Then since

$$E[a_{k+1}|\mathcal{F}_k] = E\{E[\alpha_{k+h}|\mathcal{F}_{k+1}]|\mathcal{F}_k\} = E\{\alpha_{k+h}|\mathcal{F}_k\},$$

we know by Lemma 2.1 that $a_k \in \mathcal{S}^0(\sqrt{\lambda})$ or

$$\{E[\alpha_{k+h-1}|\mathcal{F}_k]\} \in \mathcal{S}^0(\sqrt{\lambda}).$$

Continuing this procedure h times, we finally get $\{\alpha_k\} \in \mathcal{S}^0(\lambda^{2^{-h}})$. □

LEMMA 2.3. *Let $\{\alpha_k\} \in \mathcal{S}^0(\lambda)$, and $\alpha_k \le \alpha^* < 1$, where α^* is a constant. Then for any $0 < \varepsilon < 1$, $\{\varepsilon\alpha_k\} \in \mathcal{S}^0(\lambda^{(1-\alpha^*)\varepsilon})$.*

Proof. We will need the following inequality ([14, p. 145])

$$1 - x \le (1-tx)^{\frac{(1-\alpha)}{t}}, \qquad t > 1, \qquad 0 \le tx \le \alpha < 1, \tag{2.11}$$

which can be proven by using standard differentiation methods.

Let M and $\lambda \in (0, 1)$ be such that

$$E\prod_{k=m+1}^{n}(1-\alpha_k) \le M\lambda^{n-m}.$$

Then using the inequality (2.11) we have by taking $x = \varepsilon\alpha_k, t = 1/\varepsilon$,

$$E \prod_{k=m+1}^{n} (1-\varepsilon\alpha_k) \le E\left[\prod_{k=m+1}^{n} (1-\alpha_k)^{(1-\alpha^*)\varepsilon}\right]$$
$$\le \left\{E \prod_{k=m+1}^{n} (1-\alpha_k)\right\}^{(1-\alpha^*)\varepsilon} \le M^{(1-\alpha^*)\varepsilon}[\lambda^{(1-\alpha^*)\varepsilon}]^{n-m},$$

which implies the desired result □

We now give some examples to illustrate the class $S^0(\lambda)$.

Example 2.1. Nonzero strictly stationary processes do not necessarily belong to $S^0(\lambda)$ Consider the process $\alpha_k \equiv \alpha$, with α being uniformly distributed on [0, 1]. Obviously, $\{\alpha_k\}$ is a stationary process. For any $n > 0$, we have

$$E\prod_{k=1}^{n}(1-\alpha_k) = E(1-\alpha)^n = \int_0^1 (1-x)^n\, dx = \frac{1}{n+1}$$

This implies that $\{\alpha_k\} \notin S^0(\lambda)$ for any $\lambda \in [0,1)$, since the convergence rate of $E\prod_{k=1}^{n}(1-\alpha_k)$ is not exponentially fast.

Example 2.2 Let $\{\alpha_k, \mathcal{F}_k\}$ be any adapted process, $\alpha_k \in [0,1]$. If there exists some constant $\alpha > 0$ and an integer $h > 0$, such that $E[\alpha_{k+h}|\mathcal{F}_k] \ge \alpha$, then $\{\alpha_k\} \in S^0((1-\alpha)^{2^{-h}})$.

This fact can be easily proved by using Lemma 2.2. Example 2.2 contains many standard signals, for example, ϕ-mixing processes. To be precise, let ξ_k be a ϕ-mixing process, i.e., there exists a sequence $\phi(n) \xrightarrow[n\to\infty]{} 0$, such that

$$\sup_{A\in\mathcal{F}_{t+s}^{\infty}, B\in\mathcal{F}_0^t} |P(A|B) - P(A)| \le \phi(s), \quad \forall t, s,$$

where $\mathcal{F}_t^s \triangleq \sigma\{\xi(u), t \le u \le s\}$. Then for any $\mathcal{F}_t^\infty$-measurable f_t, with $|f_t| \le 1$, the following inequality holds (cf. [10], p. 82)

$$|E[f_{t+h}|\mathcal{F}_0^t] - Ef_{t+h}| \le 2\phi(h), \quad \forall t, h \tag{2.12}$$

Hence if we take $f_t = f(\xi(t))$ and assume that $Ef_t \ge \alpha > 0$, for all t, where $f(\ ) \in [0,1]$ is a measurable function, then there exists an integer $h > 0$, such that $E[f_{t+h}|\mathcal{F}_0^t] \ge \alpha/2 > 0$, for all t. This verifies the conditions of Example 2.2 for ϕ-mixing processes.

2.2. A_k nonnegative definite. We are now in a position to study the more general class $S_p(\lambda)$ defined by (2.9) We first study the stably exciting properties of nonnegative matrices $A_k, k \ge 1$, and see how the verification of $\{A_k\} \in S_p(\lambda)$ can be transferred to that of a certain scalar sequence in $S^0(\lambda)$.

THEOREM 2.1. *Let $\{A_i, \mathcal{F}_i\}$ be an adapted sequence of random matrices, $0 \le A_i \le I$. If there exists an integer $h > 0$, such that $\{\lambda_k\} \in S^0(\lambda)$, where λ_k is defined by*

$$\lambda_k \triangleq \lambda_{\min}\left\{E\left[\frac{1}{1+h}\sum_{i=kh+1}^{(k+1)h} A_i|\mathcal{F}_{kh}\right]\right\},$$

then $\{A_k\} \in S_2(\lambda^\alpha)$, with $\alpha = 1/[8h(1+h)^2]$.

Proof. Recursively define

$$\Phi(n+1,m) = (I - A_n)\Phi(n,m), \quad \Phi(m,m) = I, \quad n \ge m \ge 0 \tag{2.13}$$

Then it can be shown that (see Appendix B) for any $m \geq 1$,

$$\lambda_{\max}\{E[\Phi^\tau((m+1)h+1, mh+1)\Phi((m+1)h+1, mh+1)|\mathcal{F}_{mh}]\} \leq 1 - \frac{\lambda_m}{(1+h)}. \tag{2.14}$$

Now, for any $n > m + h$, let us define

$$k_0 = \min\{k : m \leq kh + 1 \leq n\}, \qquad k_1 = \max\{k : m \leq kh + 1 \leq n\}.$$

Then it is clear that

$$E\|\Phi(n, m)\|^2 \leq E\|\Phi(k_1 h + 1, k_0 h + 1)\|^2 \tag{2.15}$$

and

$$(k_1 + 1)h + 1 > n, \qquad (k_0 - 1)h + 1 < m. \tag{2.16}$$

Hence for $\{A_i\} \in S_2(\lambda^\alpha)$, it suffices to find a constant c which is free of k_1 and k_0 such that, for all $k_1 \geq k_0$,

$$E\|\Phi(k_1 h + 1, k_0 h + 1)\|^2 \leq c\lambda^{2\alpha h(k_1 - k_0 + 1)}. \tag{2.17}$$

To prove this, we consider the following equation:

$$x_k = \Phi(kh + 1, (k-1)h + 1)x_{k-1}, \qquad k \geq k_0 + 1 \tag{2.18}$$

where x_{k_0} is deterministic and $\|x_{k_0}\| = 1$. It is easily seen that $x_k \in \mathcal{F}_{kh}$, and $x_{k_1} = \Phi(k_1 h + 1, k_0 h + 1)x_{k_0}$. Therefore, for (2.17), we need only to prove that for any deterministic x_{k_0} with $\|x_{k_0}\| = 1$,

$$E\|x_{k_1}\|^2 \leq c\lambda^{2\alpha h(k_1 - k_0)} \tag{2.19}$$

where c is independent of k_0, k_1 and x_{k_0}.

Let us set for any $k \geq k_0 + 1$,

$$\alpha_k = \begin{cases} 1 - \dfrac{\|\Phi(kh+1, (k-1)h+1)x_{k-1}\|}{\|x_{k-1}\|}, & \text{if } \|x_{k-1}\| \neq 0; \\ 1, & \text{otherwise.} \end{cases} \tag{2.20}$$

Since $0 \leq A_i \leq I$, $i \geq 0$, implies $\|\Phi(n, m)\| \leq 1$, for all $n \geq m$, $m \geq 0$, it is clear that $\alpha_k \in [0, 1]$, $\alpha_k \in \mathcal{F}_{kh}$, and by (2.18) and (2.20),

$$\|x_k\| \leq (1 - \alpha_k)\|x_{k-1}\|$$

and

$$\|x_{k_1}\| \leq \prod_{k=k_0+1}^{k_1} (1 - \alpha_k). \tag{2.21}$$

We now show that

$$E[\alpha_{k+1}|\mathcal{F}_{kh}] \geq \frac{\lambda_k}{2(1+h)}. \tag{2.22}$$

Set $\Omega_k = \{\omega : \|x_k\| = 0\}$. Then $\Omega_k \in \mathcal{F}_{kh}$, and by (2.20)

$$I_{\Omega_k} E[\alpha_{k+1}|\mathcal{F}_{kh}] = E[I_{\Omega_k}\alpha_{k+1}|\mathcal{F}_{kh}] = I_{\Omega_k}.$$

Hence by noting $\lambda_k < 1$ we see that (2.22) is true on the set Ω_k.

To prove that (2.21) is also true on the set Ω_k^c, we first note that by (2.14), we have

$$\begin{aligned}
&E[\|\Phi((k+1)h+1, kh+1)x_k\||\mathcal{F}_{kh}] \\
&\le \{E[\|\Phi((k+1)h+1, kh+1)x_k\|^2|\mathcal{F}_{kh}]\}^{1/2} \\
&\le \{x_k^\tau E[\Phi^\tau((k+1)h+1, kh+1)\Phi((k+1)h+1, kh+1)|\mathcal{F}_{kh}]x_k\}^{1/2} \\
&\le \left\{x_k^\tau\left(1-\frac{\lambda_k}{1+h}\right)x_k\right\}^{1/2} \le \left(1-\frac{\lambda_k}{2(1+h)}\right)\|x_k\|.
\end{aligned}$$

Consequently, by (2.20) we have

$$I_{\Omega_k^c} E[\alpha_{k+1}|\mathcal{F}_{kh}] \ge I_{\Omega_k^c}\left(1-\left(1-\frac{\lambda_k}{2(1+h)}\right)\right) = \frac{\lambda_k}{2(1+h)} I_{\Omega_k^c} \tag{2.23}$$

Hence (2.22) is also true on Ω_k^c.

Since $\{\lambda_k\} \in \mathcal{S}^0(\lambda)$ and $\lambda_k \le h/(1+h)$, then by Lemma 2.3 we know that $\{\lambda_k/[2(1+h)]\} \in \mathcal{S}^0(\lambda^{4h\alpha})$. From this, (2.22) and Lemma 2.1 (together with its proof), we know that

$$\prod_{k=k_0+1}^{k_1}(1-\alpha_k) \le c\lambda^{2h\alpha(k_1-k_0)},$$

for some constant c independent of k_1, k_0, and x_{k_0}. Consequently, by (2.21) we see that (2.19) is true. Hence the proof of Theorem 2.1 is complete. □

COROLLARY 2.1. *Under the same conditions and notations as in Theorem* 2.1, *the following property holds*:

$$\{A_k\} \in \begin{cases} \mathcal{S}_p(\lambda^\alpha), & 1 \le p \le 2; \\ \mathcal{S}_p(\lambda^{2\alpha/p}), & p > 2. \end{cases} \tag{2.24}$$

Proof. For $1 \le p \le 2$, we use the monotonicity of the norm $\|\cdot\|_{L_p}$, while for $p > 2$ we apply the simple inequality $\|I - A_j\| \le 1$, and then derive

$$\left\|\prod_{j=i+1}^{k}(I-A_j)\right\|_{L_p} \le \begin{cases} \|\prod_{j=i+1}^{k}(I-A_j)\|_{L_2}, & 1 \le p \le 2; \\ \|\prod_{j=i+1}^{k}(I-A_j)\|_{L_2}^{2/p}, & p > 2. \end{cases}$$

Consequently (2.24) follows from this and Theorem 2.1. □

THEOREM 2.2. *Let $\{A_i, \mathcal{F}_i\}$ be an adapted sequence of random matrices*, $0 \le A_i \le I$. *If $\{A_i\} \in \mathcal{S}_1(\lambda)$ for some $\lambda \in [0,1)$, then there exists an integer $h > 0$ such that*

$$\inf_m \lambda_{\min}\left\{\sum_{i=mh+1}^{(m+1)h} EA_i\right\} > 0.$$

Proof. By the assumption we know that there exists a suitably large integer $h > 0$ such that

$$E\left\|\prod_{i=mh+1}^{(m+1)h}(I - A_i)\right\| \le M\lambda^h < \frac{1}{2}, \qquad \forall m. \tag{2.25}$$

Let ρ_m be the smallest eigenvalue of the matrix $E[\sum_{i=mh+1}^{(m+1)h} A_i]$, and x_m be its corresponding unit eigenvector. Then we have

$$\rho_m = E\left[\sum_{i=mh+1}^{(m+1)h} x_m^\tau A_i x_m\right].$$

Hence for any integers $i_j \in [mh+1, (m+1)h]$, $j = 1, \ldots, k, k \le h$,

$$\begin{aligned}
&Ex_m^\tau A_{i_1} \cdots A_{i_k} x_m \le E\|x_m^\tau A_{i_1}^{1/2}\|\|A_{i_1}^{1/2}A_{i_2} \cdots A_{i_k}^{1/2}\|\|A_{i_k}^{1/2}x_m\| \\
&\le E\|x_m^\tau A_{i_1}^{1/2}\| \, \|A_{i_k}^{1/2}x_m\| \le \{E\|x_m^\tau A_{i_1}^{1/2}\|^2 \cdot E\|A_{i_k}^{1/2}x_m\|^2\}^{1/2} \\
&= \{E(x_m^\tau A_{i_1}x_m)E(x_m^\tau A_{i_k}x_m)\}^{1/2} \le \max_{mh+1\le i\le(m+1)h} E(x_m^\tau A_i x_m) \le \rho_m
\end{aligned}$$

Consequently, by (2.25) we have

$$\begin{aligned}
\frac{1}{2} &> E\left\|\prod_{i=mh+1}^{(m+1)h}(I - A_i)\right\| \ge Ex_m^\tau \prod_{i=mh+1}^{(m+1)h}(I - A_i)x_m \\
&= 1 - \sum_{k=1}^{h} \sum_{mh+1\le i_1<\cdots<i_k\le(m+1)h} E(x_m^\tau A_{i_1} \cdots A_{i_k} x_m) \\
&\ge 1 - \sum_{k=1}^{h} \sum_{mh+1\le i_1<\cdots<i_k\le(m+1)h} \rho_m = 1 - \sum_{k=1}^{h}\binom{h}{k}\rho_m,
\end{aligned}$$

which implies that

$$\rho_m \ge \frac{1}{2\sum_{k=1}^{h}\binom{h}{k}}.$$

Hence Theorem 2.2 is true. □

We remark that the converse assertion of Theorem 2.2 is not true in general. This fact can be seen from Example 2.1. However, it will be true if we impose additional assumptions on $\{A_k\}$, for example, the ϕ-mixing properties. The following theorem provides necessary and sufficient conditions for such a matrix process to be in $\mathcal{S}_1(\lambda)$.

THEOREM 2.3. *If $\{A_k, k \ge 0\}$ is a ϕ-mixing matrix sequence with dimension $d \times d$, and $0 \le A_k \le I$, then the following three properties are equivalent:*

(i) $\{A_k\} \in \mathcal{S}_1(\lambda)$ *for some* $\lambda \in [0, 1)$;

(ii) *There is an integer* $h_0 > 0$ *such that*

$$\delta \triangleq \inf_m \lambda_{\min}\left\{\sum_{i=mh_0+1}^{(m+1)h_0} EA_i\right\} > 0;$$

(iii) *There exist some* $h > 0, \lambda \in (0,1)$, *such that* $\{\lambda_k\} \in \mathcal{S}^0(\lambda)$ *where* λ_k *is defined as in Theorem* 2.1 *with* $\mathcal{F}_k \triangleq \sigma\{A_i, i \leq k\}$.

Proof. By Theorems 2.1 and 2.2 we need only to prove that (ii) implies (iii).

Let the mixing rate of $\{A_k, k \geq 0\}$ be $\phi(k)$. Then applying the inequality (2.12), we are easily convinced of the following property:

$$\|E[A_{t+k}|\mathcal{F}_t] - EA_{t+k}\| \leq 2d\phi(k), \quad \forall t, k. \tag{2.26}$$

Since $\phi(n) \xrightarrow[n\to\infty]{} 0$, we can find a constant (integer) M such that

$$\phi(k) \leq \frac{\delta}{4(2h_0+1)d}, \quad \forall k \geq M, \tag{2.27}$$

where δ is defined in (ii).

Set $h = M + 2h_0 + 1$. Then by (ii) and the assumption $A_i \geq 0$, it is easy to convince oneself that

$$\lambda_{\min}\left\{\sum_{k=mh+1+M}^{(m+1)h} EA_k\right\} \geq \delta, \quad \forall m \geq 0 \tag{2.28}$$

Finally, combining (2.26)–(2.28) we conclude that for any $m \geq 0$,

$$\begin{aligned}(1+h)\lambda_m &= \lambda_{\min}\left\{E\left[\sum_{k=mh+1}^{(m+1)h} A_k|\mathcal{F}_{mh}\right]\right\} \\ &\geq \lambda_{\min}\left\{E\left[\sum_{k=mh+1+M}^{(m+1)h} A_k|\mathcal{F}_{mh}\right]\right\} \\ &\geq \lambda_{\min}\left\{E\left[\sum_{k=mh+1+M}^{(m+1)h} A_k\right]\right\} - \left\|E\left[\sum_{k=mh+1+M}^{(m+1)h} A_k|\mathcal{F}_{mh}\right] - E\left[\sum_{k=mh+1+M}^{(m+1)h} A_k\right]\right\| \\ &\geq \delta - (h-M)\frac{2d\delta}{4(2h_0+1)d} = \delta - \frac{\delta}{2} = \frac{\delta}{2} > 0.\end{aligned}$$

Hence for the h defined above, we have proved that $\{\lambda_k\} \in \mathcal{S}^0\{1 - \delta/[2(1+h)]\}$, i.e., (iii) holds. This completes the proof. □

2.3. A_k nonsymmetric. We now turn to the case where A_k is possibly nonsymmetrical and see how to transfer the study of $\{A_k\} \in \mathcal{S}_p(\lambda)$ to that of a scalar random sequence in $\mathcal{S}^0(\lambda)$.

Before pursuing this further, it is worth mentioning that in the continuous-time case, if $\{A(t)\}$ is a stationary ergodic matrix process and satisfies

$$E\lambda_{\max}\{A(0)^\tau + PA(0)P^{-1}\} < 0$$

for some positive definite matrix P, then the results of [21] and [22] state that the random differential equation $\dot{x}(t) = A(t)x(t)$ is almost surely asymptotically stable. This result may be generalized to the discrete-time case. However, this kind of results have the following limitations: (i) ergodicity is required; (ii) exponential stability can not be guaranteed, and (iii) applications to stochastic tracking algorithms are difficult.

Here, we will present a result that does not have the above-mentioned limitations. For this, we introduce the following recursive random Lyapunov equation:

$$P_{k+1} = (I - A_k)P_k(I - A_k)^\tau + Q_k, \qquad P_0 > 0, \quad k \geq 0, \tag{2.29}$$

where $\{Q_k\}$ is a sequence of nonnegative random matrices.

THEOREM 2.4. *Let $\{A_k\}$ be a sequence of $d \times d$ random matrices, and $\{Q_k\}$ be a sequence of positive definite random matrices. Then for $\{P_k\}$ recursively defined by (2.29) we have, for all $n > m$,*

$$\left\| \prod_{k=m}^{n-1} (I - A_k) \right\|^2 \leq \prod_{k=m}^{n-1} \left(1 - \frac{1}{1 + \|Q_k^{-1} P_{k+1}\|}\right) \|P_n\| \cdot \|P_m^{-1}\|. \tag{2.30}$$

Hence if $\{P_k\}$ satisfies the following two conditions,

$$\text{(i)} \left\{ \frac{1}{1 + \|Q_k^{-1} P_{k+1}\|} \right\} \in \mathcal{S}^0(\lambda), \quad \textit{for some } \lambda \in [0, 1);$$

$$\text{(ii)} \sup_{n \geq m \geq 0} \|(\|P_n\| \ \|P_m^{-1}\|)\|_{L_p} < \infty, \quad \textit{for some } p \geq 1,$$

then $\{A_k\} \in \mathcal{S}_p(\lambda^{1/2p})$.

Proof. Let us consider the following equation for $n > m$,

$$x_{k+1} = (I - A_k)x_k, \qquad k \in [m, n-1]$$

where x_m is taken to be deterministic and $\|x_m\| = 1$. Then

$$x_n = \prod_{i=m}^{n-1} (I - A_i) x_m. \tag{2.31}$$

Next we consider the following Lyapunov function $V_k = x_k^\tau P_k^{-1} x_k$. Then by denoting $B_k = I - A_k$, we have

$$V_{k+1} = x_{k+1}^\tau P_{k+1}^{-1} x_{k+1} = x_k^\tau B_k^\tau P_{k+1}^{-1} B_k x_k. \tag{2.32}$$

But, by (2.29) and the matrix inversion formula (see e.g., [27, p. 824]) we have

$$\begin{aligned}
B_k^\tau P_{k+1}^{-1} B_k &= B_k^\tau [B_k P_k B_k^\tau + Q_k]^{-1} B_k \\
&= P_k^{-1} - [P_k + P_k B_k^\tau Q_k^{-1} B_k P_k]^{-1} \\
&= P_k^{-1/2} \{I - [I + P_k^{1/2} B_k^\tau Q_k^{-1} B_k P_k^{1/2}]^{-1}\} P_k^{-1/2} \\
&\leq \{1 - [1 + \|Q_k^{-1} B_k P_k B_k^\tau\|]^{-1}\} P_k^{-1} \leq \left(1 - \frac{1}{1 + \|Q_k^{-1} P_{k+1}\|}\right) P_k^{-1},
\end{aligned}$$

which in conjunction with (2.32) yields

$$V_{k+1} \leq \left(1 - \frac{1}{1 + \|Q_k^{-1} P_{k+1}\|}\right) V_k$$

and so

$$V_n \leq \prod_{k=m}^{n-1} \left(1 - \frac{1}{1 + \|Q_k^{-1} P_{k+1}\|}\right) V_m.$$

Hence by this, (2.31) and the dependence of V_k on x_m we have

$$\begin{aligned}
\left\| \prod_{k=m}^{n-1} (I - A_i) \right\|^2 &= \max_{\|x_m\|=1} \|x_n\|^2 = \max_{\|x_m\|=1} \|x_n^\tau P_n^{-1/2} P_n^{1/2}\|^2 \\
&\leq \max_{\|x_m\|=1} \|x_n^\tau P_n^{-1/2}\|^2 \|P_n^{1/2}\|^2 = \max_{\|x_m\|=1} (V_n \|P_n\|) \\
&\leq \left\{ \prod_{k=m}^{n-1} \left(1 - \frac{1}{1 + \|Q_k^{-1} P_{k+1}\|}\right) \right\} \left\{ \|P_n\| \max_{\|x_m\|=1} V_m \right\} \\
&\leq \left\{ \prod_{k=m}^{n-1} \left(1 - \frac{1}{1 + \|Q_k^{-1} P_{k+1}\|}\right) \right\} \{ \|P_n\| \cdot \|P_m^{-1}\| \}.
\end{aligned}$$

Hence (2.30) holds. The second assertion $\{A_k\} \in \mathcal{S}_p(\lambda^{1/2p})$ follows directly from (2.30) and the Hölder inequality.

This theorem does not require that $A_i's$ are nonnegative definite matrices and means that the verification of $\{A_k\} \in \mathcal{S}_p(\lambda^{1/2p})$ can be reduced to two relatively simple tasks: (i) to verify that a certain scalar sequence is in $\mathcal{S}^0(\lambda)$, and (ii) to prove that a certain process is "L_p-stable." We remark that suitably choosing the sequence $\{Q_k\}$ is crucial in simplifying the tasks (i) and (ii). In §4, we will see that for the analysis of KF or RLS algorithms, the sequence $\{P_k\}$ may simply be taken as that defined by (1.5) or (1.10). □

3. **Stability/excitation condition.** For the basic time-varying model (1.1), we will need the following excitation condition for estimating $\{\theta_k\}$.

CONDITION 3.1 (Excitation condition). *The regressor $\{\varphi_k, \mathcal{F}_k\}$ is an adapted sequence of random vectors (i.e., φ_k is $\mathcal{F}_k$-measurable, for all k, where $\{\mathcal{F}_k\}$ is a sequence of nondecreasing σ-algebras), and there exists an integer $h > 0$ such that $\{\lambda_k\} \in \mathcal{S}^0(\lambda)$ for some $\lambda \in (0, 1)$, where λ_k is defined by*

$$\lambda_k \triangleq \lambda_{\min} \left\{ E \left[\frac{1}{1+h} \sum_{i=kh+1}^{(k+1)h} \frac{\varphi_i \varphi_i^\tau}{1 + \|\varphi_i\|^2} \,\middle|\, \mathcal{F}_{kh} \right] \right\} \tag{3.1}$$

In the next section, we will show that this condition guarantees the L_p-stability of all three standard algorithms described in §1. The main purpose of this section is to illustrate this condition by several propositions and examples of interest in application.

PROPOSITION 3.1. *Let $\{\varphi_k\}$ be a ϕ-mixing process; then the necessary and sufficient condition for Condition 3.1 to be satisfied is that there exists an integer $h > 0$ such that*

$$\inf_{k \geq 0} \lambda_{\min} \left\{ \sum_{i=kh+1}^{(k+1)h} E \left[\frac{\varphi_i \varphi_i^\tau}{1 + \|\varphi_i\|^2} \right] \right\} > 0 \tag{3.2}$$

This fact directly follows from the equivalence of the assertions (ii) and (iii) in Theorem 2.3, since $\{\varphi_i \varphi_i^\tau / (1 + \|\varphi_i\|^2)\}$ is also a ϕ-mixing process. The ϕ-mixing process is commonly used in the literature (e.g., [8], [9], [17], [18]). It includes a large class of important processes, for instance, deterministic processes, M-dependent processes and processes generated from bounded white noise filtered through a stable finite-dimensional linear filter. However, as is well known, ϕ-mixing is not perfect as a model in many applications, so next we show that Condition 3.1 is still satisfied by another important class of regressors that does not verify the ϕ-mixing condition.

In the sequel, for convenience of discussion we set $\mathcal{G}_k = \mathcal{F}_{kh}$ where h is defined in Condition 3.1. Note that λ_k is $\mathcal{G}_k$-measurable for any $k \geq 1$.

PROPOSITION 3.2. *If for some $h > 0$, $\{\lambda_k\}$ defined by* (3.1) *has the following time-varying lower bound:*

$$\lambda_k \geq \frac{1}{a_k}, \qquad \forall k \geq 1,$$

where $\{a_k, \mathcal{G}_k\}$ is an adapted sequence, $a_k \geq 1, Ea_0 < \infty$, and

$$E[a_k|\mathcal{G}_{k-1}] \leq \alpha a_{k-1} + \beta, \qquad 0 \leq \alpha < 1, 0 < \beta < \infty, \qquad \forall k \geq 1. \tag{3.3}$$

Then $\{\lambda_k\} \in \mathcal{S}^0(\lambda)$ for some $\lambda \in (0,1)$, i.e., Condition 3.1 *holds.*

Proof. By Lemma 4 in [14], we know that there exists a constant $\lambda \in (0,1)$ such that $\{1/a_k\} \in \mathcal{S}^0(\lambda)$. Hence Condition 3.1 follows immediately. □

Remark 3.1. Intuitively speaking, in order to guarantee $\{\lambda_k\} \in \mathcal{S}^0(\lambda)$, the lower bound $\{1/a_k\}$ should not "diminish" or equivalently, $\{a_k\}$ should not "grow unboundedly." Condition (3.3) effectively is a growth constraint on the random process $\{a_k\}$. If in (3.3) we take $\alpha = 0$ and $a_k = \beta$, then we get the excitation condition used in [14]. Moreover, if we assume that $\{a_k\}$ satisfies $a_k \in \mathcal{G}_k, a_k \geq 1$ and

$$a_k \leq \alpha a_{k-1} + \eta_k, \qquad \alpha \in [0,1), \qquad E[|\eta_k|^{1+\delta}|\mathcal{G}_{k-1}] \leq M, \quad \forall k \geq 1 \tag{3.4}$$

for some constants $\delta > 0$, and $M < \infty$, then we get the excitation condition proposed in [28], which obviously satisfies (3.3). Therefore, the condition of Proposition 3.2 (and hence Condition 3.1) is weaker than those proposed in [14] and [28]. Consequently, all examples presented in [14] and [28] satisfy the condition of Proposition 3.2. In particular, we have Example 3.1.

Example 3.1. Let the regressor $\{\varphi_k\}$ be generated by the following state space model:

$$\begin{aligned} x_k &= Ax_{k-1} + B\xi_k, \qquad && E\|x_0\|^4 < \infty \\ \varphi_k &= Cx_k + \zeta_k, && k \geq 0, \end{aligned}$$

where $A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times q}$ and $C \in \mathbb{R}^{d \times n}$ are deterministic matrices, A is stable, (A, B, C) is output controllable and $\{\xi_k, \zeta_k\}$ is an independent process with zero mean, and

$$E\xi_k\xi_k^\tau \geq \varepsilon I > 0, \qquad E[\|\xi_k\|^4 + \|\zeta_k\|^4] \leq M, \qquad \forall k \geq 0,$$

where ε and M are constants. Then the condition of Proposition 3.2 is satisfied.

The proof of this example is essentially the same as that for Example 2 in [28], but here the moment condition imposed on the driving signal $\{\xi_k, \zeta_k\}$ is weaker. It is also worth noting that to verify the condition in [14] we have to assume that $\{\xi_k, \zeta_k\}$ is uniformly bounded in the sample path (see [14, p. 142]).

We now turn to the main task of this section, i.e., to study the case where $\{\varphi_k\}$ is generated by a time-varying $AR(p)$ model. This model not only is a natural extension of the standard time-invariant $AR(p)$ models extensively studied in a variety of areas, but also is closely related to the closed-loop systems resulting from adaptive control (cf. [30]). We remark that in this case, the existing excitation conditions (e.g., in [14] and [28]) do not seem to be satisfied. The basic reason is that the "contraction" factor α in (3.4) is a random process rather than a constant.

Let the time-varying $AR(p)$ model be described by

$$\begin{aligned} y_k &= a_1(k)y_{k-1} + \cdots + a_p(k)y_{k-p} + v_k \\ &\triangleq \theta_k^\tau \varphi_k + v_k, \qquad k \geq 0 \end{aligned} \tag{3.5}$$

where θ_k and φ_k are p-dimensional vectors defined in a standard way, and where $\{v_k\}$ is an independent sequence that is independent of φ_0 and satisfies

$$Ev_k = 0, \qquad Ev_k^2 \geq \sigma_v^2 > 0, \qquad \sup_k E|v_k|^9 < \infty. \tag{3.6}$$

Obviously, the regressor satisfies the following state space equation:

$$\varphi_{k+1} = A_k \varphi_k + b v_k \tag{3.7}$$

where

$$A_k = \begin{bmatrix} a_1(k) & \cdots & \cdots & a_p(k) \\ 1 & \cdots & \cdots & 0 \\ \vdots & \ddots & & \\ 0 & \cdots & 1 & 0 \end{bmatrix}, \qquad b = [1, 0 \cdots 0]^\tau. \tag{3.8}$$

Example 3.2. Consider the $AR(p)$ model (3.5)–(3.6). Let $\{A_k\}$ defined by (3.8) be an independent sequence that is independent of $\{v_k\}$. If

$$\sup_k \|A_k\|_{L_q} < \infty, \qquad \left\| \prod_{i=kp}^{(k+1)p-1} A_i \right\|_{L_4} \leq \delta, \quad \forall k \geq 0, \tag{3.9}$$

where $q = \max\{4, 2(p-1)\}$ and $\delta \in (0,1)$, then the condition of Proposition 3.2 is satisfied.

The proof is given in Appendix C.

When the coefficient sequence $\{A_k\}$ is (strongly) dependent, the analysis becomes more complicated. We now consider a standard situation.

PROPOSITION 3.3. Consider the $AR(p)$ model (3.5)–(3.6). Let $\{A_k, \mathcal{F}_k'\}$ be an adapted sequence that can be decomposed as

$$A_k = A + \bar{A}_k \tag{3.10}$$

where A is a stable matrix and $\{\bar{A}_k, \mathcal{F}_k'\}$ is dominated by a nonnegative linear process:

$$\|\bar{A}_k\| \leq \beta_k, \qquad \beta_k = \beta\beta_{k-1} + e_k, \qquad 0 \leq \beta < 1, \tag{3.11}$$

where $e_k \geq 0$, $e_k \in \mathcal{F}_k'$ and e_{k+1} is independent of $\mathcal{F}_k'$. Assume that $\mathcal{F}_\infty' \triangleq \sigma\{\cup_i \mathcal{F}_i'\}$ is independent of $\{v_k\}$ and that for some constants $\varepsilon > 0$ and $b > 0$

$$\log\{E[\exp(be_k)]\} \leq \varepsilon, \qquad \forall k \geq 0. \tag{3.12}$$

Then Condition 3.1 is satisfied provided that ε and b are suitably small and large respectively.

The proof of this proposition is given in Appendix C.

Example 3.3. Let the parameter θ_k in (3.5) be the superposition of a "nominal" parameter θ and a "fluctuation" $\bar{\theta}_k$, i.e., $\theta_k = \theta + \bar{\theta}_k$. Moreover, let the time-invariant $AR(p)$ model obtained by replacing θ_k by θ in (3.5) be stable. If either $\|\bar{\theta}_k\|$ is small or $\bar{\theta}_k$ is generated by a stable *ARMA* model:

$$\bar{\theta}_k + F_1\bar{\theta}_{k-1} + \cdots + F_q\bar{\theta}_{k-q} = w_k + G_1 w_{k-1} + \cdots + G_r w_{k-r}$$

where $\{w_k\}$ is a Gaussian white noise sequence which is independent of $\{v_k\}$ with small variance. Then conditions (3.10)–(3.12) of Proposition 3.3 hold.

The proof of this example is straightforward and the details are omitted.

Remark 3.2. Conditions in Example 3.2, Proposition 3.3, and Example 3.3 are stronger than necessary as can be easily seen from the proof; they are used for simplicity of discussion. Certainly, various generalizations are possible, for example, a more general state space model (3.7) may be considered without requiring that A_k and b have the canonical form (3.8); in Example 3.2, the independence assumption of $\{A_k\}$ can be replaced by some weakly dependent conditions; and in Example 3.3, the Gaussian assumption on $\{w_k\}$ can be weakened by assuming that the distribution of $\{w_k\}$ has exponentially decaying tail (a condition similar to (3.12)).

The following result plays an essential role in the proof of Proposition 3.3 and will also be used in the next section.

LEMMA 3.1. *Let $\{x_k, \mathcal{F}_k\}$ be an adapted process, $x_k \geq 1$, and*

$$x_{k+1} \leq \alpha_{k+1}x_k + \xi_{k+1}, \qquad k \geq 0, \qquad Ex_0^2 < \infty \tag{3.13}$$

where $\{\alpha_k, \mathcal{F}_k\}$ and $\{\xi_k, \mathcal{F}_k\}$ are adapted nonnegative processes with properties:

$$\alpha_k \geq \varepsilon_0 > 0, \quad \forall k, \quad \left\| \prod_{k=m}^{n} E[\alpha_{k+1}^4 | \mathcal{F}_k] \right\|_{L_1} \leq M\gamma^{n-m+1}, \quad \forall n \geq m, \quad \forall m \tag{3.14}$$

and

$$E[\xi_{k+1}^2 | \mathcal{F}_k] \leq N < \infty, \qquad \forall k \tag{3.15}$$

where ε_0, M, N, and $\gamma \in (0,1)$ are constants. Then

(i) $\left\| \prod_{k=m}^{n} \alpha_k \right\|_{L_2} \leq M^{1/4}\gamma^{(1/4)(n-m+1)}, \qquad \forall n \geq m, \quad \forall m;$

(ii) $\sup_k E\|x_k\| < \infty;$

(iii) $\{1/x_k\} \in S^0(\lambda)$ *for some* $\lambda \in (0,1)$.

Proof. Denote $\beta_k = E[\alpha_{k+1}^4 | \mathcal{F}_k]$, and set $z_{k+1} = \left(\prod_{i=m}^{k} \beta_i\right)^{-1} \prod_{i=m}^{k} \alpha_{i+1}^4$. Then we have $z_{k+1} = z_k \beta_k^{-1} \alpha_{k+1}^4$, and so

$$Ez_{k+1} = E\{E[z_{k+1}|\mathcal{F}_k]\} = Ez_k = \cdots = Ez_{m+1} = 1, \qquad \forall k \geq m.$$

Consequently, for all $n \geq m$,

$$E\prod_{i=m}^{n} \alpha_{i+1}^2 = E\sqrt{z_{n+1}}\sqrt{\prod_{m}^{n}\beta_i}$$

$$\leq \sqrt{Ez_{n+1}}\sqrt{E\prod_{m}^{n}\beta_i} \leq \sqrt{M}\sqrt{\gamma^{n-m+1}}$$

so (i) holds, while (ii) follows immediately from (i), (3.15), and (3.13). We now proceed to prove the last assertion (iii).

We first consider the case where N defined by (3.15) is less than one. In this case, by (3.15) we have $E[\xi_{k+1}|\mathcal{F}_k] \le 1$.

For any $n > m$, set for $k \in [m, n]$

$$y_k = \left(1 - \frac{1}{x_k}\right) y_{k-1}, \qquad y_{m-1} = 1. \tag{3.16}$$

Then $y_k \in \mathcal{F}_k$ and by (3.13) we have

$$x_k y_k = (x_k - 1)y_{k-1} \le (\alpha_k x_{k-1} + \xi_k - 1)y_{k-1}$$

so with $\gamma_k \triangleq E[\alpha_{k+1}|\mathcal{F}_k]$ by noticing that $E[\xi_k|\mathcal{F}_{k-1}] \le 1$ we get

$$E[x_k y_k|\mathcal{F}_{k-1}] \le \gamma_{k-1}(x_{k-1}y_{k-1}), \qquad k \ge m. \tag{3.17}$$

Denote $z_k = \left(\prod_{i=m-1}^{k-1} \gamma_i\right)^{-1} x_k y_k, k \ge m-1$. Then by (3.17) we have for $k \ge m$,

$$E[z_k|\mathcal{F}_{k-1}] \le \left(\prod_{i=m-1}^{k-2} \gamma_i\right)^{-1} x_{k-1}y_{k-1} = z_{k-1}.$$

Consequently,

$$Ez_k \le Ez_{k-1} \le \cdots \le Ez_{m-1} = Ex_{m-1}. \tag{3.18}$$

Hence by (ii) we have for some constant $M_0 < \infty, \sup_{m\ge 0}\sup_{k\ge m} Ez_k \le M_0$. Thus by the Schwarz inequality and (3.14) we have

$$E\prod_{k=m}^{n}\left(1 - \frac{1}{x_k}\right) = Ey_n \le E\sqrt{x_n y_n} = E\sqrt{z_n \prod_{i=m-1}^{n-1}\beta_i}$$
$$\le \sqrt{Ez_n}\sqrt{E\prod_{i=m-1}^{n-1}\beta_i} \le \sqrt{M_0}M^{1/8}\gamma^{1/8(n-m+1)}$$

where for the last inequality (3.14) has been used. Hence (iii) holds.

Next, we consider the general case where N in (3.15) is an arbitrary constant. By (3.15) we may take a constant c large enough such that

$$E[\xi_{k+1}I(\xi_{k+1} \ge c)|\mathcal{F}_k] \le 1, \quad \text{and} \quad \delta \triangleq (1+\varepsilon_0)\frac{c}{1+c} > 1.$$

Then we have by (3.13),

$$x_{k+1} \le \alpha_{k+1}x_k + c + \xi_{k+1}I(\xi_{k+1} > c), \qquad k \ge 0 \tag{3.19}$$

Without loss of generality, we may assume that the equality in (3.19) holds for all k. Hence by setting $\bar{x}_k = x_k/(1+c)$ we get

$$\bar{x}_{k+1} = \alpha_{k+1}\bar{x}_k + \eta_{k+1} \tag{3.20}$$

where $\eta_{k+1} = [c + \xi_{k+1} I(\xi_{k+1} > c)]/(1+c)$. It is clear that $E[\eta_{k+1}|\mathcal{F}_k] \le 1$. Then by the fact we have just proved we know that $\{1/\bar{x}_k\} \in S^0(\gamma^{1/8})$, where γ is given in (3.14).

Note that by (3.14) and (3.20)

$$\bar{x}_{k+1} \ge \alpha_{k+1}\left(\frac{c}{1+c}\right) + \frac{c}{1+c} \ge \frac{c}{1+c}(1+\varepsilon_0) > 1, \qquad k \ge 1$$

Hence applying Lemma 2.3 with $\varepsilon = 1/(1+c)$ we know that $\{1/x_k\} \in S^0(\lambda)$, for some $\lambda \in (0,1)$. This completes the proof of Lemma 3.1. □

We remark that the condition $x_k \ge 1$ in Lemma 3.1 is by no means a restrictive condition in applications since if $x_k \ge 0$ satisfies (3.13), then the shifted process $x'_k \triangleq x_k + 1$ satisfies both $x'_k \ge 1$ and $x'_{k+1} \le \alpha_{k+1} x'_k + \xi'_{k+1}$ where $\xi'_{k+1} = \xi_{k+1} + 1$.

COROLLARY 3.1. *Let $\{x_k\}$ satisfy conditions in Lemma 3.1. If $\{y_n, \mathcal{F}_n\}$ is a nonnegative adapted process and satisfies*:

$$y_{k+1} \le \beta y_k + \eta_{k+1}, \qquad 0 \le \beta < 1, \qquad \forall k \tag{3.21}$$

where $E[\eta_k^{2q}|\mathcal{F}_{k-1}] \le M_1 < \infty$, M_1 is a positive constant and $q > \log\varepsilon_0/\log\beta$ is a positive integer and ε_0 is defined in (3.14), then $\{1/(x_k + y_k)\} \in S^0(\lambda)$ for some $\lambda \in (0,1)$.

Proof. Take ε so small such that $(1+\varepsilon)\beta^q \le \varepsilon_0$, and define $T_k = (1/q)y_k^q + (1/s)$, where $s = (1 - 1/q)^{-1}$. Note that for any $\varepsilon > 0$ and $q > 0$ there is a constant $M > 0$ depending on ε and q such that

$$(x+y)^q \le (1+\varepsilon)x^q + My^q, \qquad \forall x \ge 0, \quad \forall y \ge 0. \tag{3.22}$$

Then we have

$$\begin{aligned} T_k &\le \frac{1}{q}[\beta y_{k-1} + \eta_k]^q + \frac{1}{s} \\ &\le \frac{1}{q}[(1+\varepsilon)(\beta y_{k-1})^q + M\eta_k^q] + \frac{1}{s} \\ &\le \varepsilon_0\left[\frac{1}{q}y_{k-1}^q + \frac{1}{s}\right] + \frac{M}{q}\eta_k^q + \frac{1}{s} \le \varepsilon_0 T_{k-1} + \frac{M}{q}\eta_k^q + \frac{1}{s}. \end{aligned}$$

Hence

$$\begin{aligned} x_k + T_k &\le \alpha_k x_{k-1} + \xi_k + \varepsilon_0 T_{k-1} + \frac{M}{q}\eta_k^q + \frac{1}{s} \\ &\le \alpha_k(x_{k-1} + T_{k-1}) + \xi_k + \frac{M}{q}\eta_k^q + \frac{1}{s}. \end{aligned}$$

Applying Lemma 3.1 we know that $\{1/(x_k + T_k)\} \in S^0(\lambda)$, for some $\lambda \in (0,1)$. Finally note that $y_k \le T_k$; we conclude that $\{1/(x_k + y_k)\} \in S^0(\lambda)$. □

4. Tracking error bounds. In this section we establish tracking error bounds for the standard algorithms introduced in §1. We first present a lemma.

LEMMA 4.1. *Let $\{c_{nk}, n \ge k \ge 0\}$, $\{d_{nk}, n \ge k \ge 0\}$, and $\{\xi_k, k \ge 0\}$ be three nonnegative random processes satisfying*:

(i) *$c_{nk} \in [0,1]$, $Ec_{nk} \le M\lambda^{n-k}$, for all $n \ge k \ge 0$, for some $M > 0$ and $\lambda \in [0,1)$;*

(ii) *There exist some constants $\varepsilon > 0$ and $\alpha > 0$ such that*

$$\sup_{n \ge k \ge 0} E[\exp(\varepsilon d_{nk}^{1/\alpha})] < \infty;$$

(iii) $\sigma_p \triangleq \sup_k \|\xi_k \log^\beta(e+\xi_k)\|_{L_p} < \infty$, *for some* $p \geq 1, \beta > 0$.
Then

$$\sum_{k=0}^{n} \|c_{nk} d_{nk} \xi_k\|_{L_p} \leq c\sigma_p f(\sigma_p^{-1}), \qquad \forall n \geq 0, \tag{4.1}$$

where c is a constant independent of σ_p, *and*

$$f(\sigma_p^{-1}) = \begin{cases} \log^{1+(\beta/2)}(e+\sigma_p^{-1}), & \text{if } \beta > 2\max(1,\alpha); \\ \log^{\beta}(e+\sigma_p^{-1}), & \text{if } \{c_{nk}\} \text{ is deterministic and } \beta = \alpha; \\ \log(e+\sigma_p^{-1}), & \text{if } \{d_{nk}\} \text{ is deterministic and } \beta > 1. \end{cases} \tag{4.2}$$

The proof is given in Appendix D.

We now proceed to analyze the Kalman filter algorithm. To apply Theorem 2.4 we need to prove some boundedness properties of $\{P_k\}$ first.

LEMMA 4.2. *For* $\{P_k\}$ *generated by* (1.5), *if Condition* 3.1 *holds, then there exists a constant* $\varepsilon^* > 0$ *such that for any* $\varepsilon \in [0, \varepsilon^*)$,

$$\sup_{k \geq 0} E \exp(\varepsilon \|P_k\|) < \infty.$$

Proof. Denote

$$T_m = \sum_{k=(m-1)h+1}^{mh} tr(P_{k+1}), \qquad T_0 = 0. \tag{4.3}$$

Then $T_m \in \mathcal{G}_m \triangleq \mathcal{F}_{mh}$, and similar to Lemma 3 in [28] we have

$$T_{m+1} \leq (1 - a_{m+1})T_m + b \tag{4.4}$$

where

$$a_{m+1} = \frac{tr\left[(P_{mh+1} + hQ)^2 \sum_{k=mh+1}^{(m+1)h} \frac{\varphi_k \varphi_k^\tau}{1 + \|\varphi_k\|^2}\right]}{h(R+1)[1 + \lambda_{\max}(P_{mh+1} + hQ)]tr(P_{mh+1} + hQ)}, \qquad b = \frac{3}{2}h(h+1)trQ.$$

Similar to (39) and (40) in [28] we have $a_{m+1} \in [0, 1/(1+R)]$ and

$$E[a_{m+1}|\mathcal{G}_m] \geq \frac{(1+h)\|Q\|\lambda_m}{d(R+1)(1+h\|Q\|)} \tag{4.5}$$

where λ_m is defined by (3.1). By using Condition 3.1 and applying Lemmas 2.1 and 2.3, it is easy to see that $\{a_{k+1}\} \in S^0(\lambda)$ for some $\lambda \in [0, 1)$. Hence, the rest of the proof is completely the same as that for Lemma 4 in [28], because the key property (43) in [28] is still true. □

LEMMA 4.3. *Let* $\{P_k\}$ *be generated by* (1.5). *Then under Condition* 3.1, *for any* $\mu \in (0, 1]$ *there is a constant* $\lambda \in (0, 1)$ *such that* $\{\mu/(1 + \|Q^{-1}\| \cdot \|P_k\|)\} \in S^0(\lambda)$.

Proof. Denote $x_k = \mu^{-1}(h + \|Q^{-1}\|T_k)$, where T_k is defined by (4.3). Then it follows from (4.4) that

$$x_{k+1} \leq (1 - a_{k+1})x_k + \mu^{-1}(h + b\|Q^{-1}\|). \tag{4.6}$$

It is easy to see from (4.5), Condition 3.1, and Lemma 2.3 that Lemma 3.1 is applicable to (4.6); hence, we have $\{1/x_k\} \in S^0(\gamma)$, for some $\gamma \in (0,1)$. Note that $x_k = \sum_{i=(k-1)h+1}^{kh} \mu^{-1} \cdot [1 + \|Q^{-1}\| tr(P_{i+1})]$; hence, it is easy to conclude that $\{\mu/[1 + \|Q^{-1}\| tr(P_k)]\} \in S^0(\lambda)$ for some $\lambda \in (0,1)$ (see the proof of Lemma 5 in [14]), which ensures the desired result. □

THEOREM 4.1. *Consider the time-varying model* (1.1) *and the Kalman filter algorithm* (1.3)–(1.5). *Suppose that Condition* 3.1 *is satisfied and that for some* $p \geq 1$ *and* $\beta > 2$,

$$\sigma_p \triangleq \sup_k \|\xi_k \log^\beta (e + \xi_k)\|_{L_p} < \infty \tag{4.7}$$

and

$$\|\tilde{\theta}_0\|_{L_{2p}} < \infty \tag{4.8}$$

where $\xi_k = |v_k| + \|\Delta_{k+1}\|, \tilde{\theta}_0 = \theta_0 - \hat{\theta}_0$, *and* v_k *and* Δ_{k+1} *are given by* (1.1) *and* (1.2), *respectively. Then the tracking error* $\{\theta_k - \hat{\theta}_k, k \geq 0\}$ *is* L_p*-stable and*

$$\limsup_{k\to\infty} \|\theta_k - \hat{\theta}_k\|_{L_p} \leq c[\sigma_p \log^{1+\beta/2}(e + \sigma_p^{-1})], \tag{4.9}$$

where c *is a finite constant depending on* $\{\varphi_k\}, R, Q$ *and* p *only*; *its precise value may be found from the proof.*

Proof. By (1.4) we may rewrite (1.5) as

$$P_{k+1} = (I - L_k\varphi_k^\tau)P_k(I - L_k\varphi_k^\tau)^\tau + Q_k$$

where $Q_k = RL_kL_k^\tau + Q$. It is easy to see that $Q_k \geq Q$ and $P_{k+1} \geq Q$. Hence by applying Theorem 2.4 we have for all $n > m$,

$$\left\|\prod_{k=m}^{n-1}(I - L_k\varphi_k^\tau)\right\| \leq \prod_{k=m}^{n-1}\left(1 - \frac{1}{1 + \|Q^{-1}\| \cdot \|P_{k+1}\|}\right)^{1/2} \|P_n\|^{1/2}\|Q^{-1}\|^{1/2}. \tag{4.10}$$

Note also that $\|L_k\| \leq \|P_k\|^{1/2}/(2\sqrt{R})$, so by (2.1) we get

$$\begin{aligned}\|\tilde{\theta}_{k+1}\|_{L_p} \leq & \left\|\prod_{i=0}^{k}(I - L_i\varphi_i^\tau)\tilde{\theta}_0\right\|_{L_p} + \|Q^{-1}\|^{1/2}\sum_{i=0}^{k} \\ & \cdot \left\|\prod_{j=i+1}^{k}\left(1 - \frac{1}{2(1 + \|Q^{-1}\| \cdot \|P_{j+1}\|)}\right)\|P_{k+1}\|^{1/2}\left(1 + \frac{\|P_i\|^{1/2}}{2\sqrt{R}}\right)\xi_i\right\|_{L_p}.\end{aligned} \tag{4.11}$$

Note that by the Schwarz inequality and Lemma 4.2,

$$\sup_{k\geq i} E\exp(\varepsilon\|P_{k+1}\|^{1/2}\|P_i\|^{1/2}) \leq \sup_{k\geq i}[E\exp(\varepsilon\|P_{k+1}\|)]^{1/2}[E\exp(\varepsilon\|P_i\|)]^{1/2} < \infty.$$

So by noting Lemma 4.3 and applying Lemma 4.1 to the second term on the right-hand side of (4.11), we get the desired result. □

Next, we consider the LMS algorithm.

THEOREM 4.2. *Consider the time-varying model* (1.1) *and the LMS algorithm* (1.3) *and* (1.8). *Suppose that Condition* 3.1 *holds and that for some* $p \geq 1$ *and* $\beta > 1$, (4.7) *and* (4.8) *hold. Then* $\{\theta_k - \hat{\theta}_k, k \geq 0\}$ *is* L_p*-stable, and*

$$\limsup_{k\to\infty} \|\theta_k - \hat{\theta}_k\|_{L_p} \leq c[\sigma_p \log(e + \sigma_p^{-1})], \tag{4.12}$$

where σ_p is defined by (4.7) *and c is a constant.*

Proof. Let $c_{ki} = \| \prod_{j=i+1}^{k} \left(I - \mu \frac{\varphi_j \varphi_j^\tau}{1+\|\varphi_j\|^2} \right) \|$. Then by Condition 3.1, Lemma 2.3, and Theorem 2.1 we know that $\{c_{ki}\}$ satisfies conditions in Lemma 4.1. Note that $\|L_k\| \le \mu$, so by (2.1) we have

$$\|\tilde{\theta}_{k+1}\|_{L_p} \le \|c_{k,-1}\tilde{\theta}_0\|_{L_p} + \sum_{i=0}^{k} \|c_{ki}\xi_i\|_{L_p}$$

and the desired result (4.12) follows by applying Lemma 4.1. □

Remark 4.1. Combining Propositions 2.1 and 2.2 with Theorem 2.3, we see that Condition 3.1 is also a necessary one for the stability of the LMS algorithm *in some sense.*

Finally, we study the recursive least squares algorithm.

LEMMA 4.4. *Let $\{P_k\}$ be generated by* (1.10) *with forgetting factor $\alpha \in (0,1)$. If Condition* 3.1 *holds, then for any $p \ge 1$*

$$\sup_{k\ge 0} E\|P_k\|^p < \infty,$$

provided that α satisfies $\lambda^{[16hd(2h-1)p]^{-1}} < \alpha < 1$, where λ and h are given by Condition 3.1, *and d is the dimension of $\{\varphi_k\}$.*

Proof. The proof ideas are similar to those for Lemmas 1 and 2 in [14] for the Kalman filter algorithm. For any $m \ge 0$ by (1.10) we have

$$P_k \le \frac{1}{\alpha}P_{k-1} \le \cdots \le \left(\frac{1}{\alpha}\right)^{h-1} P_{mh+1}, \qquad k \in [mh+1,(m+1)h].$$

Then by the matrix inverse formula from (1.10) again we have for $k \in [mh+1,(m+1)h]$,

$$\begin{aligned} P_{k+1} &= [\alpha P_k^{-1} + \varphi_k\varphi_k^\tau]^{-1} \le [\alpha\alpha^{h-1}P_{mh+1}^{-1} + \varphi_k\varphi_k^\tau]^{-1} \\ &= \left(\frac{1}{\alpha}\right)^h \left[P_{mh+1} - \frac{P_{mh+1}\varphi_k\varphi_k^\tau P_{mh+1}}{\alpha^h + \varphi_k^\tau P_{mh+1}\varphi_k}\right] \\ &\le \left(\frac{1}{\alpha}\right)^h \left[P_{mh+1} - \frac{P_{mh+1}\varphi_k\varphi_k^\tau P_{mh+1}}{[\alpha^h + \lambda_{\max}(P_{mh+1})][1+\|\varphi_k\|^2]}\right]. \end{aligned} \tag{4.13}$$

Denote

$$T_m = \sum_{k=(m-1)h+1}^{mh} tr(P_{k+1}), \qquad a_{m+1} = \frac{tr\left[P_{mh+1}^2 \sum_{k=mh+1}^{(m+1)h} \frac{\varphi_k\varphi_k^\tau}{1+\|\varphi_k\|^2}\right]}{[\alpha^h + \lambda_{\max}(P_{mh+1})]htr(P_{mh+1})}. \tag{4.14}$$

Then summing up both sides of (4.13) we get

$$T_{m+1} \le \alpha^{-h}[1-a_{m+1}]htr(P_{mh+1}). \tag{4.15}$$

But by the inequality $P_{k+1} \le \alpha^{-1}P_k$ it follows that

$$\begin{aligned} htr(P_{mh+1}) &= \sum_{k=(m-1)h+1}^{mh} tr(P_{mh+1}) \\ &\le \sum_{k=(m-1)h+1}^{mh} \alpha^{k-mh}tr(P_{k+1}) = \alpha^{1-h}T_m. \end{aligned}$$

Hence by (4.15)

$$T_{m+1} \le \alpha^{1-2h}[1 - a_{m+1}]T_m \tag{4.16}$$

For any $p \ge 1$, denote

$$b_{m+1} = \alpha^{(1-2h)p}\left[1 - \frac{a_{m+1}}{2}\right] I(tr(P_{mh+1}) \ge 1)$$

Then by (4.15) and (4.16),

$$\begin{aligned} T_{m+1}^p &\le T_{m+1}^p[I(tr(P_{mh+1}) \ge 1) + I(tr(P_{mh+1} \le 1))] \\ &\le b_{m+1}T_m^p + (h\alpha^{-h})^p. \end{aligned} \tag{4.17}$$

By the definition of a_{m+1} in (4.14) and the fact that $tr(P_k^2) \ge d^{-1}(tr P_k)^2$,

$$\begin{aligned} E[a_{m+1}|\mathcal{F}_{mh}] &\ge \frac{(h+1)\lambda_m tr(P_{mh+1}^2)}{h(1+tr(P_{mh+1}))tr(P_{mh+1})} \\ &\ge \frac{(h+1)\lambda_m}{2hd}, \quad \text{on } \{tr(P_{mh+1}) \ge 1\}. \end{aligned}$$

Hence by the definition of b_{m+1},

$$E[b_{m+1}|\mathcal{F}_{mh}] \le \alpha^{(1-2h)p}\left(1 - \frac{(h+1)\lambda_m}{4hd}\right) I(tr(P_{mh+1}) \ge 1). \tag{4.18}$$

Denote

$$\alpha_{m+1} = \begin{cases} b_{m+1}, & \text{if } tr(P_{mh+1}) \ge 1; \\ \alpha^{(1-2h)p}\left(1 - \dfrac{(1+h)\lambda_m}{4hd}\right), & \text{otherwise} \end{cases} \tag{4.19}$$

Then we have by (4.17)

$$T_{m+1}^p \le \alpha_{m+1}T_m^p + (h\alpha^{-h})^p. \tag{4.20}$$

By Condition 3.1, $\lambda_m \in S^0(\lambda)$ for some $\lambda \in (0,1)$. Since $\lambda_m \le h/(1+h)$, by Lemma 2.3 we know that $\{[(1+h)/(4hd)]\lambda_m\} \in S^0(\lambda^{(4hd)^{-1}})$. Hence by (4.18) and (4.19) and the assumption that $\lambda^{[16hd(2h-1)p]^{-1}} < \alpha$, it is easy to see that Lemma 3.1 is applicable to (4.20) and thus we get $\sup_m ET_m^p < \infty$. So Lemma 4.4 holds. □

THEOREM 4.3. *Consider the time-varying model* (1.1) *together with the forgetting factor algorithm* (1.3), (1.9), *and* (1.10). *Suppose that the following conditions are satisfied:*

(i) *Conditions* 3.1 *holds, i.e.,* $\lambda_m \in S^0(\lambda)$ *for some* $\lambda \in (0,1)$ *and some integer* $h > 0$, *where* λ_m *is defined by* (3.1);

(ii) *For some* $p \ge 1$

$$\sup_k(\|v_k\|_{L_{3p}} + \|\Delta_k\|_{L_{3p}}) \le \sigma_{3p};$$

(iii) $\sup_k \|\varphi_k\|_{L_{6p}} < \infty$;

(iv) *The forgetting factor* α *satisfies* $\lambda^{[48hd(2h-1)p]^{-1}} < \alpha < 1$, *where d is the dimension of* $\{\varphi_k\}$.

Then there exists a constant c such that

$$\limsup_{k\to\infty} \|\theta_k - \hat{\theta}_k\|_{L_p} \le c\sigma_{3p}.$$

Proof. We may complete the proof by using Theorem 2.4 just as it has been used for Theorem 4.1. However, in the present case the following analysis appears to be more straightforward.

By the matrix inverse formula, it follows from (1.10) that

$$P_{k+1}^{-1} = \alpha P_k^{-1} + \varphi_k \varphi_k^\tau. \tag{4.21}$$

Multiplying P_k^{-1} from both sides of (1.10) and using (1.9) we get $[I - L_k\varphi_k^\tau] = \alpha P_{k+1}P_k^{-1}$, and so

$$\prod_{j=i+1}^{k} (I - L_j\varphi_j^\tau) = \alpha^{k-i} P_{k+1}P_{i+1}^{-1}. \tag{4.22}$$

On the other hand, multiplying φ_k from both sides of (1.10) we have $P_{k+1}^{-1}L_k = \varphi_k$. Hence by (2.1) and (4.22),

$$\|\theta_{k+1} - \hat{\theta}_{k+1}\|_{L_p} \le \alpha^k \|P_{k+1}P_0^{-1}\tilde{\theta}_0\|_{L_p} + \sum_{i=0}^{k} \alpha^{k-i}(\|P_{k+1}\varphi_i v_i\|_{L_p} + \|P_{k+1}P_{i+1}^{-1}\Delta_{i+1}\|_{L_p}).$$

By the Hölder inequality, Assumptions (i)–(iv), and Lemma 4.4 we know that the proof will be complete if we can show that $\sup_i \|P_{i+1}^{-1}\|_{L_{3p}} < \infty$. But, this can be easily seen from (4.21) and Assumption (iii), since

$$\|P_{k+1}^{-1}\|_{L_{3p}} \le \alpha\|P_k^{-1}\|_{L_{3p}} + \|\varphi_k\|_{L_{6p}}^2, \qquad \forall k \ge 0. \qquad \square$$

Remark 4.2. Under additional statistical assumptions on the processes $\{\varphi_k, v_k, \Delta_k\}$, a refined upper bound for the tracking error of the forgetting factor RLS can be derived (see [33]).

Conclusions. In this paper, stability and tracking error bounds are established for several standard estimation algorithms under a very general excitation condition. The various stability results presented in the paper are believed to be necessary preliminaries for further study of tracking properties, e.g., approximate expressions of the variance of the tracking errors (see e.g., [18]). Also, applications of the results to adaptive control systems as studied in e.g., [30] are possible. These issues will be discussed in detail elsewhere.

Appendix A.

Proof of Proposition 2.1. The solution of (2.2) may be expressed by

$$x_{n+1} = \sum_{i=0}^{n} \left[\prod_{j=i+1}^{n} (I - A_j) \right] \xi_{i+1}. \tag{A.1}$$

From this and the independence of $\{A_k\}$ and $\{\xi_k\}$ we know that the sufficiency of (2.4) is obvious.

To prove the necessity, we take $\{\xi_k\}$ to be an independently and identically distributed (i.i.d.) sequence with zero mean and unit variance. Then by denoting $B_k = I - A_k$, we have for some $c > 0$ and for any $n \geq k > 0$,

$$
\text{(A.2)}\qquad
\begin{aligned}
c \geq E\|x_{n+1}\|^2 &= tr \sum_{i=0}^{n} E\left[\prod_{j=i+1}^{n} B_j\right]\left[\prod_{j=i+1}^{n} B_j\right]^{\tau} \\
&\geq \sum_{i=k}^{n} tr\left\{E\left[\prod_{j=i+1}^{n} B_j\right]\left[\prod_{j=i+1}^{n} B_j\right]^{\tau}\right\}
\end{aligned}
$$

Denote

$$
a(n,i) = tr\left\{E\left[\prod_{j=i+1}^{n} B_j\right]\left[\prod_{j=i+1}^{n} B_j\right]^{\tau}\right\}
$$

It is easy to verify that $a(n,i) > 0$, for all $n \geq i$. Then by the independency of $\{A_j\}$ we have for any $n \geq i \geq k$,

$$
\begin{aligned}
a(n,k) &= trE[B_n \cdots B_{k+1}B_{k+1}^{\tau} \cdots B_n^{\tau}] \\
&= trE\{B_n \cdots B_{i+1}E[B_i \cdots B_{k+1}B_{k+1}^{\tau} \cdots B_i^{\tau}]B_{i+1}^{\tau} \cdots B_n^{\tau}\} \\
&\leq trE[B_n \cdots B_{i+1}B_{i+1}^{\tau} \cdots B_n^{\tau}]trE[B_i \cdots B_{k+1}B_{k+1}^{\tau} \cdots B_i^{\tau}] \\
&= a(n,i)a(i,k)
\end{aligned}
$$

Hence by (A.2) we have

$$
c \geq a(n,k)\sum_{i=k}^{n} a^{-1}(i,k)
$$

or

$$
\text{(A.3)}\qquad \sum_{i=k}^{n} a^{-1}(i,k) \leq ca^{-1}(n,k), \quad \forall n \geq k \geq 0.
$$

From this we have

$$
\begin{aligned}
\sum_{i=k}^{n} a^{-1}(i,k) &= a^{-1}(n,k) + \sum_{i=k}^{n-1} a^{-1}(i,k) \\
&\geq \left(1+\frac{1}{c}\right)\sum_{i=k}^{n-1} a^{-1}(i,k) \geq \cdots \\
&\geq \left(1+\frac{1}{c}\right)^{n-k} a^{-1}(k,k) = \left(1+\frac{1}{c}\right)^{n-k} d.
\end{aligned}
$$

Therefore, by (A.3)

$$
ca^{-1}(n,k) \geq \left(1+\frac{1}{c}\right)^{n-k} d
$$

or

$$a(n,k) \le \frac{c}{d}\left(\frac{c}{1+c}\right)^{n-k}, \qquad \forall n \ge k, \quad \forall k \ge 0.$$

So (2.4) holds with $\lambda = [c/(1+c)]^{1/2} < 1$. □

Proof of Proposition 2.2. Denote $\psi(i,k) = \prod_{j=k+1}^{i}(I - A_j)$, and set for any fixed $k \ge 0$,

$$\xi_{i+1} = \psi(i,k)[E\psi(i,k)^\tau \psi(i,k)]^{-1/2}\eta_{i+1}, \tag{A.4}$$

where $\{\eta_{i+1}\}$ is a d-dimensional i.i.d. sequence independent of $\{A_i\}$ with $E\eta_{i+1} = 0$, $E\eta_i\eta_i^\tau = (1/d)I$. It is easy to see that

$$\begin{aligned} E\|\xi_{i+1}\|^2 &= tr[E\xi_{i+1}\xi_{i+1}^\tau] \\ &= \frac{1}{d} tr E\{\psi(i,k)[E\psi(i,k)^\tau\psi(i,k)]^{-1}\psi(i,k)^\tau\} = 1. \end{aligned}$$

Hence for any $k \ge 0, \xi \in \mathcal{B}^0$. Substituting (A.4) into (A.1) and calculating the covariance, we get

$$Ex_{n+1}x_{n+1}^\tau = \frac{1}{d}E\sum_{i=0}^{n}\psi(n,k)[E\psi(i,k)^\tau\psi(i,k)]^{-1}\psi(n,k)^\tau, \qquad \forall k \ge 0,$$

and so

$$\begin{aligned} &[E\psi(n,k)^\tau\psi(n,k)]^{1/2}\sum_{i=0}^{n}[E\psi(i,k)^\tau\psi(i,k)]^{-1}[E\psi(n,k)^\tau\psi(n,k)]^{1/2} \\ &\le tr\, E\left\{\sum_{i=0}^{n}\psi(n,k)[E\psi(i,k)^\tau\psi(i,k)]^{-1}\psi(n,k)^\tau\right\} I \\ &= dE\|x_{n+1}\|^2 I \le cI, \quad \forall n \ge k, \quad \forall k, \end{aligned}$$

where for the last inequality we have used the assumption (2.8) and where c is a finite constant. This inequality implies that

$$\sum_{i=0}^{n}[E\psi(i,k)^\tau\psi(i,k)]^{-1} \le c[E\psi(n,k)^\tau\psi(n,k)]^{-1}$$

Hence by denoting $a^{-1}(i,k) \triangleq \lambda_{\min}\{[E\psi(i,k)^\tau\psi(i,k)]^{-1}\}$, we obtain

$$\sum_{i=k}^{n} a^{-1}(i,k) \le \sum_{i=0}^{n} a^{-1}(i,k) \le ca^{-1}(n,k), \qquad \forall n \ge k \ge 0$$

This inequality is exactly the same as (A.3). Hence by the same arguments as those in the proof of Proposition 2.1, we get

$$a(n,k) \le \frac{c}{d}\left(\frac{c}{1+c}\right)^{n-k}, \quad \forall n \ge k, \qquad \forall k \ge 0.$$

Finally, observing that $a(n,k) = \lambda_{\max}\{E[\psi(n,k)^\tau\psi(n,k)]\}$, we get the desired result (2.4). □

Appendix B.

Proof of (2.14). For simplicity of notations, set $k = mh + 1$. Following the ideas in the proofs of Theorem 4.5 and Lemma 10.7 in [5], we denote z_{k-1} as the unit eigenvector corresponding to the largest eigenvalue ρ_{k-1} of the matrix $E[\Phi^\tau(k+h,k)\Phi(k+h,k)|\mathcal{F}_{k-1}]$, and recursively define z_j by

$$z_j = (I - A_j)z_{j-1}, \qquad j \geq k. \tag{B.1}$$

It follows from (2.13) that $z_{k+h-1} = \Phi(k+h,k)z_{k-1}$. Hence we have

$$\begin{aligned} E(\|z_{k+h-1}\|^2|\mathcal{F}_{k-1}) &= z_{k-1}^\tau E[\Phi^\tau(k+h,k)\Phi(k+h,k)|\mathcal{F}_{k-1}]z_{k-1} \\ &= \rho_{k-1}\|z_{k-1}\|^2 = \rho_{k-1}. \end{aligned} \tag{B.2}$$

By (B.1) we have

$$z_j = z_{k-1} - \sum_{i=k}^{j} A_i z_{i-1}, \quad \forall j \in [k, k+h-1].$$

Hence by the Schwarz inequality

$$\begin{aligned} E[\|z_{j-1} - z_{k-1}\|^2|\mathcal{F}_{k-1}] &= E\left[\left\|\sum_{i=k}^{j-1} A_i z_{i-1}\right\|^2 |\mathcal{F}_{k-1}\right] \\ &\leq E\left[\left(\sum_{i=k}^{j-1}\|A_i^{1/2}z_{i-1}\|^2\right)\sum_{i=k}^{j-1}\|A_i^{1/2}\|^2|\mathcal{F}_{k-1}\right] \\ &\leq hE\left[\sum_{i=k}^{j-1} z_{i-1}^\tau A_i z_{i-1}|\mathcal{F}_{k-1}\right], \qquad j \in [k, k+h]. \end{aligned} \tag{B.3}$$

By the definition of λ_m and the Minkowski inequality we have

$$\begin{aligned} &\sqrt{(1+h)}\lambda_m^{1/2} \\ &\leq \left\{z_{k-1}^\tau E\left[\sum_{i=mh+1}^{(m+1)h} A_i|\mathcal{F}_{mh}\right]z_{k-1}\right\}^{1/2} = \left\{E\left[\sum_{i=k}^{k+h-1}\|A_i^{1/2}z_{k-1}\|^2|\mathcal{F}_{k-1}\right]\right\}^{1/2} \\ &\leq \left\{E\left[\sum_{i=k}^{k+h-1}\|A_i^{1/2}z_{i-1}\|^2|\mathcal{F}_{k-1}\right]\right\}^{1/2} + \left\{E\left[\sum_{i=k}^{k+h-1}\|z_{i-1} - z_{k-1}\|^2|\mathcal{F}_{k-1}\right]\right\}^{1/2}. \end{aligned}$$

From this and (B.3) it follows that

$$\sqrt{(1+h)}\lambda_m^{1/2} \leq (1+h)\left\{E\left[\sum_{i=k}^{k+h-1}\|A_i^{1/2}z_{i-1}\|^2|\mathcal{F}_{k-1}\right]\right\}^{1/2}$$

or

$$E\left[\sum_{i=k}^{k+h-1}\|A_i^{1/2}z_{i-1}\|^2|\mathcal{F}_{k-1}\right] \geq \frac{\lambda_m}{1+h}. \tag{B.4}$$

By (B.1) and the fact that $0 \le A_i \le I$ it is easily derived that

$$z_j^\tau z_j \le z_{j-1}^\tau z_{j-1} - z_{j-1}^\tau A_j z_{j-1},$$

from which we have

$$\begin{aligned}\|z_{k+h-1}\|^2 &\le \|z_{k-1}\| - \sum_{i=k}^{k+h-1} z_{i-1}^\tau A_i z_{i-1}\\ &= 1 - \sum_{i=k}^{k+h-1} z_{i-1}^\tau A_i z_{i-1}.\end{aligned}$$

Combining this with (B.2) and (B.4) we get

$$\begin{aligned}\rho_{k-1} &= E[\|z_{k+h-1}\|^2|\mathcal{F}_{k-1}]\\ &\le 1 - E\left[\sum_{i=k}^{k+h-1} z_{i-1}^\tau A_i z_{i-1}|\mathcal{F}_{k-1}\right] \le 1 - \frac{\lambda_m}{1+h},\end{aligned}$$

which is tantamount to (2.14). □

Appendix C. We first prove Proposition 3.3. The proof is divided into two steps.

Step 1. We first prove that

$$\text{(C.1)}\qquad \lambda_k \ge \frac{1}{P(\beta_{kp-1})[1 + \|\varphi_{kp}\|^4]}, \quad \forall k \ge 1$$

where λ_k is defined by (3.1) with $h = p$ and $\mathcal{F}_k = \sigma\{\mathcal{F}_i', v_i, i \le k-1\}$ and where $P(x)$ is a polynomial of x with nonnegative coefficients.

By (3.7) we have

$$\text{(C.2)}\qquad \varphi_{k+1} = \left(\prod_{j=s}^{k} A_j\right)\varphi_s + \sum_{i=s}^{k}\left(\prod_{j=i+1}^{k} A_j\right) b v_i, \quad \forall k \ge s, \quad \forall s \ge 0$$

By (3.1) with $h = p$ and the Schwarz inequality it is easy to show that (cf. [14] or [28], p. 168)

$$\text{(C.3)}\qquad \begin{aligned}\lambda_k &\ge \frac{1}{p+1}\lambda_{\min}\left\{E\left[\frac{\varphi_{(k+1)p}\varphi_{(k+1)p}^\tau}{1+\|\varphi_{(k+1)p}\|^2}\,\middle|\,\mathcal{F}_{kp}\right]\right\}\\ &\ge \frac{1}{1+p}\frac{\{\lambda_{\min}(E[\varphi_{(k+1)p}\varphi_{(k+1)p}^\tau|\mathcal{F}_{kp}])\}^2}{E[(\|\varphi_{(k+1)p}\|^4 + \|\varphi_{(k+1)p}\|^2)|\mathcal{F}_{kp}]}.\end{aligned}$$

We first analyze the numerator. Denote the controllability Gramian by H_{kp+1}:

$$\text{(C.4)}\qquad H_{kp+1} \triangleq \sum_{i=kp}^{(k+1)p-1}\left(\prod_{j=i+1}^{(k+1)p-1} A_j\right) bb^\tau \left(\prod_{j=i+1}^{(k+1)p-1} A_j\right)^\tau$$

Then by (C.2), (3.6), and the independence assumptions we have

$$\text{(C.5)}\qquad E[\varphi_{(k+1)p}\varphi_{(k+1)p}^\tau|\mathcal{F}_{kp}] \ge \sigma_v^2 E[H_{kp+1}|\mathcal{F}_{kp}].$$

By (3.8) and (C.4) it is easy to verify that $det[H_{kp+1}] = 1$, and hence by (C.5)

$$\lambda_{\min}\{E[\varphi_{(k+1)p}\varphi^\tau_{(k+1)p}|\mathcal{F}_{kp}]\} \geq \sigma_v^2 E[\lambda_{\min}(H_{kp+1})|\mathcal{F}_{kp}] \geq \sigma_v^2 E\left[\frac{det(H_{kp+1})}{\{\lambda_{\max}(H_{kp+1})\}^{p-1}}\,\middle|\,\mathcal{F}_{kp}\right] \geq \frac{\sigma_v^2}{E\{\|H_{kp+1}\|^{p-1}|\mathcal{F}_{kp}\}}. \tag{C.6}$$

Concerning the denominator in (C.3), we first note that

$$E[\|\varphi_{(k+1)p}\|^2|\mathcal{F}_{kp}] \leq \frac{1}{2} + \frac{1}{2}E[\|\varphi_{(k+1)p}\|^4|\mathcal{F}_{kp}] \tag{C.7}$$

and that by (C.2)

$$E[\|\varphi_{(k+1)p}\|^4|\mathcal{F}_{kp}] \leq 8E\left[\left\|\prod_{i=kp}^{(k+1)p-1} A_i\right\|^4 \middle|\, \mathcal{F}_{kp}\right]\|\varphi_{kp}\|^4 + 8p^3\|b\|^4 \sup_k Ev_k^4 \sum_{i=kp}^{(k+1)p-1} E\left[\left\|\prod_{j=i+1}^{(k+1)p-1} A_j\right\|^4 \middle|\, \mathcal{F}_{kp}\right].$$

Then, substituting this, (C.6), and (C.7) into (C.3) and using (3.10)–(3.11) together with the Markovian properties of β_k, it is not difficult to conclude (C.1). □

Step 2. We prove that

$$\left\{\frac{1}{P(\beta_{kp-1})[1+\|\varphi_{kp}\|^4]}\right\} \in S^0(\lambda), \quad \text{for some } \lambda \in (0,1). \tag{C.8}$$

Since A is a stable matrix, there is a norm $\|\cdot\|_\delta$ on $\mathbb{R}^p$ such that its induced norm on $\mathbb{R}^{p\times p}$ (also denoted by $\|\cdot\|_\delta$) satisfies $\|A\|_\delta \stackrel{\Delta}{=} \delta < 1$. Clearly, there is a constant $c > 1$ such that, for all $x \in \mathbb{R}^p$, $\|x\| \leq c\|x\|_\delta$. In order to apply Corollary 3.1 we denote

$$x_k = \|\varphi_{kp}\|_\delta^8 + \beta_{kp-1}^L + 1$$
$$y_k = \frac{c^8}{2}P^2(\beta_{kp-1}), \qquad \mathcal{G}_k = \sigma\{\mathcal{F}_i', v_i, i \leq kp-1\}$$

where L is a suitably large number defined later on. Then both $\{x_k, \mathcal{G}_k\}$ and $\{y_k, \mathcal{G}_k\}$ are adapted processes. Clearly, $P(\beta_{kp-1})[1+\|\varphi_{kp}\|^4] \leq x_k + y_k$. Hence, (C.8) will be proved if conditions in Corollary 3.1 can be verified.

By (3.11) and the convexity of the function $P^2(x), x \geq 0$, we have

$$\begin{aligned} y_{k+1} &\leq \frac{c^8}{2}P^2\left(\beta^p\beta_{kp-1} + \sum_{i=kp}^{(k+1)p-1} e_i\right) \\ &\leq \frac{c^8}{2}P^2\left(\beta\beta_{kp-1} + (1-\beta)\frac{1}{1-\beta}\sum_{i=kp}^{(k+1)p-1} e_i\right) \\ &\leq \beta y_k + \frac{c^8}{2}(1-\beta)P^2\left(\frac{1}{1-\beta}\sum_{i=kp}^{(k+1)p-1} e_i\right). \end{aligned}$$

Hence $\{y_k\}$ satisfies the required properties.

Now, it only remains to prove that $\{x_k\}$ satisfies conditions in Lemma 3.1. By (3.10)–(3.11) we have

$$\|A_k\|_\delta \leq \delta + \|\bar{A}_k\|_\delta \leq \delta + c_\delta \beta_k,$$

where $c_\delta > 0$ is a constant. This motivates us to set $\alpha_{k+1} \triangleq \prod_{i=kp}^{(k+1)p-1}(\delta + c_\delta\beta_i)$. Clearly, $\alpha_k \in \mathcal{G}_k$, and by a completely similar argument as that used in [29] we know that under condition (3.12) (with small ε and large b) there are constants $M > 0, \gamma \in (0,1)$ such that

$$\left\| \prod_{k=m}^{n} E[\alpha_{k+1}^4 | \mathcal{G}_k] \right\| \leq M\gamma^{n-m+1}, \quad \forall n \geq m, \quad \forall m \geq 0.$$

Let α be a positive number such that $(1+\alpha)^4\gamma < 1$, where γ is defined above. It is easy to see from (C.2) and the definition of α_{k+1} that there is a constant $M_\alpha > 0$ such that for any $\varepsilon_0 > 0$,

$$\text{(C.9)} \quad \|\varphi_{(k+1)p}\|_\delta^8 \leq (1+\alpha)\alpha_{k+1}\|\varphi_{kp}\|_\delta^8 + M_\alpha \left\| \sum_{i=kp}^{(k+1)p-1} \left(\prod_{j=i+1}^{(k+1)p-1} A_j \right) bv_i \right\|_\delta^8$$

$$\text{(C.10)} \quad \begin{aligned} &\leq (1+\alpha)\alpha_{k+1}\|\varphi_{kp}\|_\delta^8 + \frac{\varepsilon_0}{2}\beta_{kp-1}^L \\ &\quad + c_1 \left(\sum_{i=kp}^{(k+1)p-1} e_i + 1 \right)^L + c_2 \left(\sum_{i=kp}^{(k+1)p-1} |v_i| + 1 \right)^9 \end{aligned}$$

for some constants L, c_1, and c_2, where for the last inequality we have used the fact that $\|A_j\|_\delta \leq \delta + c_\delta\beta_j$ together with the Markovian property of $\{\beta_j\}$.

Without loss of generality we may assume that L in (C.10) is so large that

$$\text{(C.11)} \quad 4\beta^{pL} \leq (1+\alpha)\delta^{8p} \triangleq \varepsilon_0$$

By (C.11) and (3.11) it is easy to see that there is a constant $c_3 > 0$ such that

$$\text{(C.12)} \quad \beta_{kp-1}^L \leq \frac{\varepsilon_0}{2}\beta_{(k-1)p-1}^L + c_3 \left(\sum_{i=(k-1)p}^{kp-1} e_i \right)^L.$$

Combining (C.11) and (C.12), using the definition of x_k and the fact that $(1+\alpha)\alpha_{k+1} \geq \varepsilon_0$, we get for some constant $c_4 > 0$,

$$x_{k+1} \leq (1+\alpha)\alpha_{k+1}x_k + c_4 \left[1 + \left(\sum_{i=kp}^{(k+1)p-1} e_i + 1 \right)^L + \left(\sum_{i=kp}^{(k+1)p-1} |v_i| + 1 \right)^9 \right].$$

Hence both $\{x_k\}$ and $\{y_k\}$ satisfy conditions of Corollary 3.1, and so $\{1/(x_k+y_k)\} \in S^0(\lambda)$ for some $\lambda \in (0,1)$. This proves (C.8). Finally, combining (C.1) and (C.8) we know that Proposition 3.3 is true. □

Proof of Example 3.2. Set $\mathcal{F}_k = \sigma\{A_i, v_i, i \le k-1\}$. Since $\{A_i\}$ is an independent sequence, similar to the proof of (C.1) we have for some constant $c_5 \ge 1$,

$$\lambda_k \ge \frac{1}{a_k}, \qquad a_k = c_5(1 + \|\varphi_{kp}\|^4), \qquad k \ge 0. \tag{C.13}$$

So we need only to prove that $\{a_k\}$ verifies (3.3). Let $\alpha > 0$ be such that $(1+\alpha)\delta^4 < 1$ where δ is given by (3.9). By (C.2), (3.9) and the independence assumptions we know that

$$E[\|\varphi_{(k+1)p}\|^4 | \mathcal{F}_{kp}] \le (1+\alpha)\delta^4 \|\varphi_{kp}\|^4 + c_6$$

for some constant $c_6 > 0$. Consequently, we have

$$E[a_{k+1} | \mathcal{F}_{kp}] \le (1+\alpha)\delta^4 \alpha_k + c_5(1+c_6)$$

Hence (3.3) is true. □

Appendix D.

Proof of Lemma 4.1. We first consider the case where $\beta > 2\max(1,\alpha)$. Let $\delta_p > 0$ be such that

$$\sup_{n \ge k \ge 0} \|(d_{nk}\xi_k)\log^{\beta/2}(e + d_{nk}\xi_k)\|_{L_p} \le \delta_p.$$

Then exactly the same argument as that used for Lemma 8 in [28] yields

$$\sum_{k=0}^{n} \|c_{nk} d_{nk} \xi_k\|_{L_p} \le c[\delta_p \log(e + \delta_p^{-1})] \tag{D.1}$$

So we need only to find a relationship between δ and σ. By inequality (52) in [28] we know that

$$xy \le \sigma \exp(\varepsilon x^{1/\alpha}) + c_1 y[\log^\alpha(e + \sigma^{-1}) + \log^\alpha(e+y)] \tag{D.2}$$

holds for all $\sigma > 0, \varepsilon > 0$, and $\alpha > 0$, where c_1 is a constant depending only on ε and α. Applying (D.2) with $x = d_{nk}^p \log^{p\beta/2}(e + d_{nk}), y = \xi_k^p \log^{p\beta/2}(e+\xi_k), \sigma = \sigma_p^p, \alpha = p\beta/2$ we have

$$\begin{aligned} &E d_{nk}^p \xi_k^p \log^{p\beta/2}(e + d_{nk}\xi_k) \le 2^{p\beta/2} Exy \\ &\le 2^{p\beta/2} E\{\sigma_p^p \exp(\varepsilon x^{2/(p\beta)}) + c_1 y[\log^{p\beta/2}(e + \sigma_p^{-p}) + \log^{p\beta/2}(e+y)]\} \\ &\le c\sigma_p^p \log^{p\beta/2}(e + \sigma_p^{-1}), \quad \text{for some constant } c. \end{aligned} \tag{D.3}$$

Hence we may take $\delta_p = c\sigma_p \log^{\beta/2}(e + \sigma_p^{-1})$. Substituting this into (D.1) we know that the first case in (4.2) is true, while the second case can be proved in a similar way. Finally, the last case can be derived from (D.1) by noting $\delta_p \le c_1 \sigma_p$ for some $c_1 > 0$. □

REFERENCES

[1] L. LJUNG AND T. SÖDERSTRÖM, *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, MA, 1983.

[2] B. WIDROW AND S. STEARNS, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1985.

[3] G. C. GOODWIN AND K. S. SIN, *Adaptive Filtering, Prediction and Control*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

[4] R. S. LIPSTER AND A. N. SHIRYAYEV, *Statistics of Random Processes II*, Springer-Verlag, New York, 1977.

[5] H. F. CHEN AND L. GUO, *Identification and Stochastic Adaptive Control*, Birkhäuser, Boston, 1991
[6] O. MACCHI AND E. EWEDA, *Compared speed and accuracy of RLS and LMS algorithms with constant forgetting factors*, Traitement Signal, 22 (1988), pp 255–267.
[7] B. WIDROW, J. M. MCCOOL, M. G. LARIMORE, AND C. R. JOHNSON, JR, *Stationary and nonstationary learning characteristics of the LMS adaptive filter*, Proc. IEEE, 64 (1976), pp. 1151–1162
[8] B. BITMEAD, *Convergence properties of LMS adaptive estimators with unbounded dependent inputs* IEEE Trans Automat. Control, 29 (1984), pp. 477–479.
[9] O. MACCHI, *Optimization of adaptive identification for time-varying filters*, IEEE Trans. Automat. Control, 31 (1986), pp 283–287
[10] H. J. KUSHNER, *Approximation and Weak Convergence Methods for Random Processes With Applications to Stochastic Systems Theory*, MIT Press, Cambridge MA, 1984
[11] A. BENVENISTE, *Design of adaptive algorithms for the tracking of time-varying systems*, Internat. J. Adapt Control Signal Process., 1 (1987), pp. 1–29
[12] V. SOLO, *The limit behavior of LMS*, IEEE Trans. Acoustics, Speech, Signal Process 37 (1989), pp 1909–1922.
[13] L. GUO, L. G. XIA, AND J. B. MOORE, *Tracking randomly varying parameters: analysis of a standard algorithm* in Proc. 27th IEEE CDC, Austin, TX, 1988, pp 1514–1519
[14] L. GUO, *Estimating time-varying parameters by Kalman filter based algorithm: stability and convergence*, IEEE Trans. Automat Control, 35 (1990) pp. 141–147.
[15] G. KITAGAWA AND W. GERSCH, *A smoothness priors time varying AR coefficient modelling of nonstationary covariance time series*, IEEE Trans Automat Control, 30 (1985), pp. 48–56.
[16] M. NIEDŹWIECKI AND L. GUO, *Nonasymptotic results for finite-memory WLS filters* IEEE Trans Automat Control, 36 (1991), pp 198–206.
[17] S. BITTANTI AND M. CAMPI, *Adaptive RLS algorithms under stochastic excitation—L^2 convergence analysis*, IEEE Trans Automat. Control, 36 (1991), pp 963–967
[18] L. LJUNG AND P. PRIOURET, *A result on the mean square error obtained using general tracking algorithms*, Internat J Adapt Control Signal Process., 5 (1991) pp. 231–250.
[19] B.D.O. ANDERSON, *Internal and external stability of linear time-varying systems*, SIAM J. Control Optim, 20 (1982), pp. 408–413.
[20] H. FURSTENBERG AND H. KESTEN, *Products of random matrices*, Ann Math. Statist., 31 (1960) pp. 457–469.
[21] E. F. INFANTE, *On the stability of some linear nonautonomous random systems*, J Appl. Mech 35 (1968) pp 7–12
[22] G. BLANKENSHIP, *Stability of linear differential equations with random coefficients*, IEEE Trans Automat Control, 22 (1987), pp. 834–838.
[23] R. Z. HASMINSKII, *Stochastic Stability of Differential Equations*. Sijthoff and Nordhoff., Alphen aan den Rijn The Netherlands, 1980
[24] S. GEMAN, *Some averaging and stability results for random differential equations* SIAM J Appl Math, 36 (1979), pp. 87–107.
[25] L. ARNOLD AND V. WIHSTUTZ, EDS., *Lyapunov exponents*, Lecture Notes in Math 1186 Springer-Verlag, New York, Berlin, 1987.
[26] L. GERENCSER, *Almost sure exponential stability of random linear differential equations* Stochastics Stochastics Rep., 36 (1991), pp 91–107.
[27] P. E. CAINES, *Linear Stochastic Systems*, Wiley, New York 1988
[28] J. F. ZHANG, L. GUO, AND H. F. CHEN, *L_p-stability of estimation errors of Kalman filter for tracking time-varying parameters* Internat J Adapt. Control Signal Process., 5 (1991), pp. 155–174.
[29] S. P. MEYN AND L. GUO. *Geometric ergodicity of a bilinear time series model*. J Time Ser. Anal 14 (1993), pp 93–108.
[30] L. GUO, *On adaptive stabilization of time-varying stochastic systems*. SIAM J Control Optim 28 (1990) pp. 1432–1451.
[31] S. P. MEYN AND P. E. CAINES, *A new approach to stochastic adaptive control*, IEEE Trans Automat. Control AC-32 (1987) pp 220–226.
[32] S. P. MEYN AND L. BROWN, *Model Reference Adaptive Control of Time-Varying and Stochastic Systems*, Tech. Rep., Coordinated Science Laboratory, University of Illinois Urbana, IL, 1992.
[33] L. GUO, L. LJUNG, AND P. PRIOURET, *Tracking performance analysis of the forgetting factor RLS algorithm* in Proc 31st IEEE CDC, Tucson, AZ, 1992