Exponential Stability of General Tracking Algorithms

Lei Guo and Lennart Ljung, *Fellow, IEEE*

***Abstract*— Tracking and adaptation algorithms are, from a formal point of view, nonlinear systems which depend on stochastic variables in a fairly complicated way. The analysis of such algorithms is thus quite complicated. A first step is to establish the exponential stability of these systems. This is of interest in its own right and a prerequisite for the practical use of the algorithm. It is also a necessary starting point to analyze the performance in terms of tracking and adaptation because that is how close the estimated parameters are to the time-varying true ones.**

In this contribution we establish some general conditions for the exponential stability of a wide and common class of tracking algorithms. This includes least mean squares, recursive least squares, and Kalman filter based adaptation algorithms. We show how stability of an averaged (linear and deterministic) equation and stability of the actual algorithm are linked to each other under weak conditions on the involved stochastic processes. We also give explicit conditions for exponential stability of the most common algorithms. The tracking performance of the algorithms is studied in a companion paper.

I. Introduction

To track parameters that describe time-varying properties in signals and systems is of fundamental importance in many applications. There is a quite substantial literature on this topic, and much effort has been spent on the analysis of these algorithm. See among many references, e.g., the books and surveys [26], [29], [28], and [24] as well as [2], [9], [11], [23], [6], and [5].

The analysis problem is surprisingly difficult, and no general, comprehensive theory is really available.

It is the purpose of this contribution, together with the companion paper [17], to provide some quite general results for the analysis of tracking algorithms. In particular, we are concerned with results that apply to more general algorithms than the familiar and simple least mean squares (LMS) method.

As remarked, e.g., in [24], the analysis of tracking algorithms naturally splits into two different, but related, problems: one is to establish that the algorithm is exponentially stable, and the other is to obtain useful expressions for its performance. It is also desirable to treat a whole family of algorithms under similar conditions.

To be more specific, consider the following standard linear regression

$$y_k = \varphi_k^\tau \theta_k + v_k \tag{1}$$

Manuscript received February 18, 1994. Recommended by Past Associate Editor, B. Pasik-Duncan. This work was supported in part by the National Natural Science Foundation of China and the Swedish Research Council for Engineering Sciences (TFR).

L. Guo is with the Institute of Systems Science, Chinese Academy of Sciences, Beijing, 100080, China.

L. Ljung is with the Department of Electrical Engineering, Linköping University, S-581 83 Linköping, Sweden.

IEEE Log Number 9412438.

where (y_k, φ_k) are measured signals, v_k is a disturbance sequence, and θ_k is an unknown time-varying (vector) parameter process whose variation at time k is denoted by Δ_k, i.e.,

$$\theta_k = \theta_{k-1} + \Delta_k. \tag{2}$$

Adaptive tracking and estimation algorithms for the unknown parameter θ_k usually have the following form (e.g., [26], [29])

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \mu L_k (y_k - \varphi_k^\tau \hat{\theta}_k), \qquad \mu \in (0,1) \tag{3}$$

where L_k is a random process, called the adaptation gain vector, which reflects the "direction" of adaptation. The constant μ is called the step size or speed of adaptation, which is usually chosen to be small for slowly varying parameters.

Denote

$$\tilde{\theta}_k \triangleq \theta_k - \hat{\theta}_k, \qquad F_k \triangleq L_k \varphi_k^\tau. \tag{4}$$

Then (1)–(4) give the following error equation

$$\tilde{\theta}_{k+1} = (I - \mu F_k)\tilde{\theta}_k - \mu L_k v_k + \Delta_{k+1}. \tag{5}$$

By taking different gain sequences L_k, we obtain the following three well-known basic algorithms:

a) *Least Mean Squares (LMS)*

$$F_k = \varphi_k \varphi_k^\tau. \tag{6}$$

b) *Recursive Least Squares (RLS)*

$$F_k = P_k \varphi_k \varphi_k^\tau \tag{7}$$

$$P_k = \frac{1}{1-\mu}\left\{P_{k-1} - \mu \frac{P_{k-1}\varphi_k \varphi_k^\tau P_{k-1}}{1 - \mu + \mu \varphi_k^\tau P_{k-1}\varphi_k}\right\}. \tag{8}$$

c) *Kalman Filter (KF)* based algorithm

$$F_k = \frac{P_{k-1}\varphi_k \varphi_k^\tau}{R + \mu \varphi_k^\tau P_{k-1}\varphi_k} \tag{9}$$

$$P_k = P_{k-1} - \frac{\mu P_{k-1}\varphi_k \varphi_k^\tau P_{k-1}}{R + \mu \varphi_k^\tau P_{k-1}\varphi_k} + \mu Q, \quad (R > 0, Q > 0). \tag{10}$$

The KF algorithm would be the optimal algorithm if Δ_k and v_k in (1)–(2) were white Gaussian noises with covariances Q and R, respectively.

For the above general tracking algorithms, our prime interest is in the values of the covariance matrix of the tracking error $E[\tilde{\theta}_k\tilde{\theta}_k^\tau]$. Unfortunately, except for some trivial cases, precise expressions for this error are extremely difficult to obtain, mainly due to the inherent dependence between the signals $\{\varphi_k, v_k, \Delta_k\}$. In the companion paper [17], we show that this value can be well approximated by an easily computable expression, namely, by

$$\hat{\Pi}_{k+1} = (I - \mu E[F_k])\hat{\Pi}_k(I - \mu E[F_k])^\tau + \mu^2 E v_k^2 E[L_k L_k^\tau] + E[\Delta_{k+1}\Delta_{k+1}^\tau] \quad (11)$$

with the following guaranteed accuracy

$$||E[\tilde{\theta}_{k+1}\tilde{\theta}_{k+1}^\tau] - \hat{\Pi}_{k+1}|| \le \sigma(\mu)||\hat{\Pi}_{k+1}||$$

where $\sigma(\mu) \to 0$ as $\mu \to 0$, provided that the signals $\{\varphi_k, v_k, \Delta_k\}$ satisfy a certain weak dependence condition and that the homogeneous equation of (5)

$$x_{k+1} = (I - \mu F_k)x_k \quad (12)$$

together with its "averaged" equation

$$\overline{x}_{k+1} = (I - \mu E[F_k])\overline{x}_k \quad (13)$$

is exponentially stable in stochastic and deterministic senses, respectively.

It may be shown, using the stochastic BIBO (bounded-input/bounded-output) results in [16], that exponential stability of (12) and (13) are necessary in general to guarantee the boundedness of $E||\tilde{\theta}_k||^2$ and $||\hat{\Pi}_k||$. Exponential stability of time-varying equations has been recognized to be a tough problem, and few existing methods and results are available, especially in the stochastic case. In general, the deterministic equation (13) is relatively easy to handle compared to the stochastic equation (12), and hence our main efforts in this paper will be focused on stochastic stability of (12), which essentially amounts to studying a product of random matrices.

The investigation of products of random matrices and of the related random linear differential equations has a long history (e.g., [10], [1], [3], and [4] and the references therein) where stationarity and ergodicity of the random matrices are usually assumed and only the asymptotic stability (a weaker property than exponential stability) has been studied. Related results for continuous-time random linear equations are also available (see, e.g., [7] and [22]). Concerning the exponential stability, some interesting results for continuous-time random differential equations are presented in [12] and [13], under the assumption that the coefficient matrices are uniformly bounded and satisfy a certain mixing conditions. Without imposing any specific assumptions on $\{F_k\}$, some stability results on (12) are obtained in [16] by transferring the exponential stability of the vector equation to that of a relatively simple scalar random equation via a random Lyapunov equation (but the dependence of the Lyapunov exponent on μ is not quantified when applied to KF). A detailed study for products of bounded nonnegative definite random matrices is also presented in [16]. These results, as mentioned in [16], are just preliminaries for parts of the results to be established in the present paper.

Tracking algorithms differ mainly in their form of adaptation gains. In general, the gains can be roughly divided into two classes: gradient-based gains (e.g., LMS) and Riccati equation-based gains (e.g., KF and RLS). Thus, exponential stability of (12) will naturally be studied separately for each cases.

In this paper, we shall prove that:

1) For a quite general class of $\{F_k\}$, the (stochastic) exponential stability of (12) is equivalent to the (deterministic) exponential stability of (13) without necessarily requiring boundedness, stationary, and strong mixing conditions on the coefficient matrix process. This may be directly applied to a general class of gradient-based algorithms, and
2) For algorithms with gains based on Riccati equations, especially for the standard KF algorithm, we shall prove that by reducing the dimension of the stochastic equations in question, the exponential stability can be established by resorting to a general stochastic excitation condition.

Our main result will be formulated as Theorem 3.2 in Section III. It states that exponential stability of the averaged equation (13) implies exponential stability of the actual one, (12). The conditions for this are essentially that i) the distributions of the stochastic matrix F_k have rapidly decaying tails and ii) that the dependence among the different F_k decays sufficiently fast. These assumptions are quite weaker than boundedness and strong mixing, respectively.

The result in itself is of a technical nature, but has an independent interest, in addition to being a building block for more practical performance analysis. Even more so, the reader will find that the proof is quite technical and the development in several of the sections is terse. This seems to be unavoidable, given the character of the problem.

In a companion paper, [17], the exponential stability results are applied to analysis of the tracking performance of a broad class of algorithms.

II. NOTATION

For convenience of reference, here we collect the main notations to be used throughout the paper.

a) The maximum eigenvalue of a matrix X is denoted by $\lambda_{\max}(X)$, and the Euclidean norm of X is defined as its maximum singular value, i.e.,

$$||X|| \triangleq \{\lambda_{\max}(XX^\tau)\}^{1/2}$$

and the L_p-norm of a random matrix X is defined as

$$||X||_p \triangleq \{E(||X||^p)\}^{1/p}, \qquad p \ge 1.$$

b) For any square random matrix sequence $F = \{F_k\}$ and real numbers $p \ge 1, \mu^* \in (0, 1)$, the stochastic exponentially stable family is defined by

$$\mathcal{S}_p(\mu^*) = \left\{ F: \left\| \prod_{j=i+1}^{k} (I - \mu F_j) \right\|_p \leq M(1-\mu\alpha)^{k-i}, \right.$$
$$\forall \mu \in (0, \mu^*], \forall k \geq i \geq 0,$$
$$\left. \text{for some } M > 0 \quad \text{and} \quad \alpha \in (0,1) \right\}.$$

Likewise, the corresponding deterministic exponentially stable family is defined by

$$\mathcal{S}(\mu^*) = \left\{ F: \left\| \prod_{j=i+1}^{k} (I - \mu E[F_j]) \right\| \leq M(1-\mu\alpha)^{k-i}, \right.$$
$$\forall \mu \in (0, \mu^*], \forall k \geq i \geq 0,$$
$$\left. \text{for some } M > 0, \quad \text{and} \quad \alpha \in (0,1) \right\}.$$

In what follows, it will be convenient to set

$$\mathcal{S}_p \triangleq \bigcup_{\mu^* \in (0,1)} \mathcal{S}_p(\mu^*), \quad \mathcal{S} \triangleq \bigcup_{\mu^* \in (0,1)} \mathcal{S}(\mu^*). \tag{14}$$

These sets thus mean that the stochastic, and deterministic, respectively, equation is exponentially stable for at least one nonzero value of μ (and then also all smaller values).

c) For a scalar sequence $a = \{a_k, k \geq 0\}$ and $\lambda \in (0,1)$, we set

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

Also

$$\mathcal{S}^0 \triangleq \bigcup_{\lambda \in (0,1)} \mathcal{S}^0(\lambda). \tag{15}$$

d) Let $p \geq 1, F \triangleq \{F_i\}$. Set

$$S_j^{(T)} = \sum_{t=jT}^{(j+1)T-1} (F_t - E[F_t]) \tag{16}$$

and define the set

$$\mathcal{M}_p = \{F: \sup_j \|S_j^{(T)}\|_p = o(T), \quad \text{as} \quad T \to \infty\}.$$

The sum $S_j^{(T)}$ consists of T zero-mean terms, and just requiring that its norm increases slower than T means that a rather weak averaging property is imposed on the F_k. As is known, martingale difference sequences, ϕ- and α-mixing sequences, and linear random processes are all in the set $\mathcal{M}_p$ (see [18]).

III. The Main Results

A. Stochastic Averaging Based Results

We first claim that under a very general weak dependence condition, the stability of stochastic equation (12) implies stability of averaged equation (13): $\{F_k\} \in \theta_2$ implies that $\{F_k\} \in \theta$.

Theorem 3.1: Let $\{F_k\}$ be a random matrix process belonging to $\mathcal{M}_2$. Then

$$\{F_k\} \in \mathcal{S}_2 \Rightarrow \{F_k\} \in \mathcal{S}.$$

The proof is given in Section IV.

Remark: By Jensen's or Lyapunov's inequality it follows that $\{F_k\}$ belongs to $\mathcal{S}_2$ if it belongs to $\mathcal{S}_p$ for any $p > 2$. The theorem thus holds for any $p \geq 2$. If some further assumptions (like independence among the F_k) are introduced, it can be shown along the lines of the current proof that it will also hold for $p \geq 1$.

Next, our main task is to establish the following "converse" assertion together with its consequences.

Theorem 3.2: Let $\{F_k\}$ be a random matrix process. Then

$$\{F_k\} \in \mathcal{S} \Rightarrow \{F_k\} \in \mathcal{S}_p, \forall p \geq 1$$

provided that the following two conditions are satisfied:

i) There exist positive constants ε, δ, M and K such that for any $n \geq i \geq 0$

$$E \exp\left\{ \varepsilon \sum_{j=i+1}^{n} \|F_j\|^{1+\delta} \right\} \leq M \exp\{K(n-i)\}.$$

ii) There exist a constant M and a nondecreasing function $g(T)$ with $g(T) = o(T)$ such that for any fixed T, all small $\mu > 0$ and any $n \geq i \geq 0$

$$E \exp\left\{ \mu \sum_{j=i+1}^{n} \|S_j^{(T)}\| \right\} \leq M \exp\{[\mu g(T) + o(\mu)](n-i)\}$$

where $S_j^{(T)}$ is defined by (16).

The proof is given in Section V.

Now, some remarks and discussions on the above two conditions are in order.

Remark 3.1: Note that i) and ii) of the theorem are trivially satisfied if F_k is deterministic. In that case, the theorem is of course vacuous, but the point is that no too strong assumptions have been introduced.

Remark 3.2: Condition i) is immediately satisfied if $\|F_j\| \leq L, \forall j$, for some constant $L > 0$. This is the case when, for example, a normalized version of the LMS algorithm is considered (cf., [16] and [24]) or when a weighted version of LMS with $L_k = P_k \varphi_k$ is studied (cf., [25]), where φ_k is a bounded random regressor and P_k is any bounded deterministic weighting matrix (the standard LMS corresponds to $P_k = I$). In general, Condition i) implies that the distribution of the random process $\{F_k\}$ has an exponentially decaying tail. This kind of condition has been previously used in stability analysis of stochastic differential equations (cf., [20]) and in stochastic adaptive control (see [27]). We now give two examples to illustrate this condition.

Example 3.1: If there are positive constants $\varepsilon, \delta, M, h$ such that

$$E \exp\{\varepsilon ||F_{k+h}||^{1+\delta} | \{\mathcal{F}_k\} \le M, \qquad \forall k$$

then similar to the proof in [15], it can be verified that Condition i) holds.

Example 3.2: Let $\{F_k\}$ be dominated by a linear process

$$||F_k|| \le \sum_{j=-\infty}^{\infty} \lambda_j w_{k-j}, \qquad \sum_{j=-\infty}^{\infty} \lambda_j < \infty$$

with $\lambda_j \ge 0$ and with $\{w_k\}$ being a nonnegative independent process satisfying

$$\sup_k E \exp\{\alpha w_k^{1+\delta}\} \le e^b < \infty$$

for some positive constants α, δ and b. Then Condition i) is also true. (The proof is given in Appendix A.)

Remark 3.3: Let $\{F_k\}$ be a sequence in $\mathcal{M}_p$, then with $S_j^{(T)}$ defined by (16) we know that $\sup_j ||S_j^{(T)}||_p = o(T)$ as $T \to \infty$. This gives an intuitive explanation for the upper bound in Condition ii). This condition mainly reflects the weak dependence property of the coefficient matrices and cannot be removed in general. To see this, let $F_k = f, \forall k$, with f being a random variable uniformly distributed on [0, 1]. Then, clearly, $\{F_k\} \in \mathcal{S}$ and Condition i) holds, but for any $p \ge 1, \{F_k\} \notin \mathcal{S}_p$, because for any $\mu \in (0,1)$

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

which does not tend to zero exponentially fast. This example also shows that even for stationary cases, the result of Theorem 3.2 may not be true if no weak dependence condition is imposed.

Remark 3.4: From the proof of Theorem 3.2, it follows that if Condition ii) is replaced by a weaker one $(r > 1)$

$$\left|\left| \prod_{j=i+1}^{n} (1 + \mu ||S_j^{(T)}||) \right|\right|_r \le M[1 + \mu g(T) + o(\mu)]^{n-i}, \qquad \forall n \ge i \ge 0 \quad (17)$$

then we have $\{F_k\} \in \mathcal{S}_p, \forall p < r$. Moreover, under Condition i), (17) can be guaranteed if for some $q > r$

$$\left|\left| \prod_{j=i+1}^{n} \{1 + \mu ||S_j^{(T)}|| I(||S_j^{(T)}|| \le c_T)\} \right|\right|_q \le M\{1 + \mu g(T) + o(\mu)\}^{n-i}, \qquad \forall n \ge i \ge 0 \quad (18)$$

where c_T satisfies $T = o(c_T)$ as $T \to \infty$, and $I(\cdot)$ is the indicator function. (The proof of this fact is given in Appendix B.)

Since the truncated random process $\{||S_j^{(T)}|| I(||S_j^{(T)}|| \le c_T), j \ge 1\}$ is uniformly bounded, (18) will considerably ease the task of verifying Condition ii), as evidenced by the proof of the following corollary (see Section IV).

Corollary 3.1: Let $\{F_k\}$ be a ϕ-mixing random process satisfying Condition i) of Theorem 3.2. Then

$$\{F_k\} \in \mathcal{S} \Rightarrow \{F_k\} \in \mathcal{S}_p, \qquad \forall p \ge 1.$$

The next corollary shows that Theorem 3.2 is also applicable to a wide class of matrix processes other than ϕ-mixing (see Section VI for the proof).

Corollary 3.2: Let $\{F_k\}$ be expanded as

$$F_k = \sum_{j=-\infty}^{\infty} A_j V_{k-j} + \xi_k, \qquad \sum_{j=-\infty}^{\infty} ||A_j|| < \infty$$

(compare this with the well-known Wold decomposition for stationary processes in, for example, [8]), where $\{V_k\}$ is an independent sequence satisfying

$$\sup_k E \exp\{\alpha ||V_k||^{1+\delta}\} < \infty, \quad \text{for some} \quad \alpha > 0, \delta > 0$$

and $\{\xi_k\}$ is a bounded deterministic process. Then

$$\{F_k\} \in \mathcal{S} \Rightarrow \{F_k\} \in \mathcal{S}_p, \qquad \forall p \ge 1.$$

Now, we show how Theorem 3.2 can be conveniently applied to the analysis of LMS.

Theorem 3.3: Let $\{F_k\}$ be a sequence of nonnegative definite random matrices (e.g., the LMS case). If $\{F_k\}$ satisfies the conditions in either Corollary 3.1 or Corollary 3.2 and if there are constants $h > 0$ and $\delta > 0$ such that

$$\sum_{i=k+1}^{k+h} E[F_i] \ge \delta I, \qquad \forall k$$

then we have

$$\{F_k\} \in \mathcal{S}_p \cap \mathcal{S}, \qquad \forall p \ge 1.$$

Proof: By Corollaries 3.1 and 3.2, we need only to show that $\{F_k\} \in \mathcal{S}(\mu^*)$, for some $\mu^* \in (0,1)$. This follows directly from Theorem 2.1 in [16], however, if we take the matrix A_k there as the deterministic matrix $\mu E[F_k]$ and take $\mu^* < (\sup_k ||E[F_k]||)^{-1}$.

B. Stochastic Lyapunov Equation Based Results

In general, given a random matrix process $\{F_k\}$, we may always transfer the study of $\{F_k\} \in \mathcal{S}_p$ to that of a relatively simple scalar random sequence in $\mathcal{S}^0$, via the following random Lyapunov equation

$$P_{k+1} = (I - \mu F_k) P_k (I - \mu F_k^T) + Q_k, \qquad P_0 > 0 \quad (19)$$

where $\{Q_k\}$ is a sequence of nonnegative definite random matrices (may depend on μ). The following result is proved in Theorem 2.4 of [16]. (Recall the definition of $\mathcal{S}^0$ in (15).)

Theorem 3.4: Let $\{F_k\}$ be an arbitrary matrix random process and $\{Q_k\}$ be any sequence of positive definite random matrices. If $\{P_k\}$ defined by (19) satisfies the following two conditions (for some $\mu^* \in (0,1), \alpha \in (0,1), p \geq 1$)

$$\text{i)} \quad \left\{\frac{1}{1+\|Q_k^{-1}P_{k+1}\|}\right\} \in \mathcal{S}^0(1-\alpha\mu), \qquad \forall \mu \in (0,\mu^*),$$

$$\text{ii)} \quad \sup_{\mu \in (0,\mu^*)} \sup_{n \geq m \geq 0} \|(\|P_n\| \cdot \|P_m^{-1}\|)\|_p < \infty$$

then $\{F_k\} \in \mathcal{S}_p(\mu^*)$.

Note that the essence of both i) and ii) is that P_k stays bounded even if Q_k does not vanish.

Based on this result, we may establish the exponential stability of the KF algorithm. The RLS algorithm may also be analyzed within this framework. Due to its special form, however, simpler methods exist (cf., [18]), and we shall not discuss that here.

Now, let $h > 0$ be an integer and let $\mathcal{G}_k$ be the sigma-algebra generated by $\{\varphi_i, i \leq k\}$. Introduce

$$\lambda_k \triangleq \lambda_{\min}\left\{E\left[\frac{1}{1+h}\sum_{i=kh+1}^{(k+1)h}\frac{\varphi_i\varphi_i^\tau}{1+\|\varphi_i\|^2}\middle|\mathcal{G}_{kh}\right]\right\} \tag{20}$$

$$b_k = (1-\delta)b_{k-1} + \delta(\|\varphi_k\|^2+1), \qquad \delta \in (0,1). \tag{21}$$

Theorem 3.5: Consider the KF algorithm defined by (3), (9), and (10). Assume that there are constants $\lambda \in (0,1)$ (independent of δ) and $h > 0$ such that

$$\left\{\frac{\lambda_k}{1+b_{kh+1}}\right\} \in \mathcal{S}^0(\lambda) \tag{22}$$

where $\mathcal{S}^0(\lambda)$ is defined by (15), and λ_l and b_k are defined by (20) and (21). Assume also that for some $p \geq 1, \|\varphi_k\|_{2p} = O(1)$. Then $\{F_k\} \in \mathcal{S}_t, \forall t < p$.

The proof is given in Section VII.

Let us comment on condition (22). It is a variation of standard persistently excitation (PE) conditions for the regression sequence $\{\varphi_i\}$. The usual PE condition contains two parts: a lower bound on the "richness" of the regressors and an upper bound on their growth rate. Condition (22) means that the smallest eigenvalue λ_k is not "too small" and that b_k (and hence φ_k) is not "too large." Thus, (22) is similar in spirit as the standard PE conditions.

Remark 3.5: A simple case where (22) holds is when φ_k is bounded and persistently gets "full rank contributions." The latter would mean that

$$\varphi_{k+1} = f(\varphi_k, \varphi_{k-1}, \cdots) + v_{k+1}$$

where v_{k+1} is independent of the past and of uniformally full rank. That means that λ_k can be taken as a strictly positive, deterministic scalar (<1), and b_k is bounded, for which trivially (22) holds.

Remark 3.6: When $\{\varphi_k\}$ is bounded, so is $\{b_k\}$ defined by (21), and hence the excitation condition (22) is implied by $\{\lambda_k\} \in \mathcal{S}^0(\lambda)$ (cf., [16]). This condition has been studied in detail in [16]. Also, when $\{\varphi_k\}$ is unbounded, but is the output of a linear state-space model, then (22) can also be verified (cf. [16]).

IV. The Proof of Theorem 3.1

We start with lemmas.

Lemma 4.1: Let $a_0, a_1, \cdots, a_T$ be random variables independent of μ. If there exist $\mu^* > 0$, and a function $M(\mu)$ such that $\forall \mu \in (0, \mu^*]$

$$\|a_0 + \mu a_1 + \cdots + \mu^T a_T\|_p \leq M(\mu) < \infty, \quad \text{for some} \quad p \geq 1$$

then there is a constant M (independent of μ) such that

$$\|a_i\|_p \leq M < \infty, \qquad \forall i = 0, 1, \cdots, T.$$

Proof: Take $(T+1)$ distinct real numbers $\mu_i \in (0, \mu^*)$, $i = 1, 2, \cdots, T+1$, and set $A = (a_0, a_1, \cdots, a_T)^\tau$, $U_i = (1, \mu_i, \cdots, \mu_i^T)^\tau, U = [U_1, \cdots, U_{T+1}]$.

Then by the assumption, we have

$$\|A^\tau U_i\|_p \leq M(\mu_i), \qquad i = 1, \cdots, T+1$$

consequently, by the Minkowski inequality

$$\begin{aligned}\|A^\tau U\|_p &= \|A^\tau[U_1, \cdots, U_{T+1}]\|_p \\ &\leq \|A^\tau U_1\|_p + \cdots + \|A^\tau U_{T+1}\|_p \leq \sum_{i=1}^{T+1} M(\mu_i).\end{aligned}$$

But U is a nonsingular Vandermonde matrix, so we have

$$\begin{aligned}\|A\|_p &= \|AUU^{-1}\|_p \leq \|AU\|_p \cdot \|U^{-1}\| \leq \|U^{-1}\| \\ &\quad \cdot \sum_{i=1}^{T+1} M(\mu_i).\end{aligned}$$

#

Lemma 4.2: Let $T \geq 2$ be an integer. Denote

$$G_n \triangleq \prod_{i=nT}^{(n+1)T-1}(I - \mu F_i), \quad \overline{G}_n = \prod_{i=nT}^{(n+1)T-1}(I - \mu E[F_i]).$$

Assume that $\{F_k\} \in \mathcal{S}_2(\mu^*) \cap \mathcal{M}_2$ for some $\mu^* > 0$. Then for any $\mu \in (0, \mu^*]$ and all $n \geq 1$

$$\|G_n - \overline{G}_n\|_2 \leq \mu T f(T) + O(\mu^2)$$

where the "O" constant may depend on T, and $f(T) \to 0$ as $T \to \infty$.

Proof: By the definition of $\mathcal{S}_2(\mu^*)$, it is seen that

$$\|I - \mu F_k\|_2 \leq M, \qquad \forall \mu \in (0, \mu^*], \qquad \forall k \geq 1 \tag{23}$$

and

$$\|G_n\|_2 \leq M(1-\mu\alpha)^T \leq M, \qquad \forall \mu \in (0, \mu^*], \quad \forall n \geq 1. \tag{24}$$

Note that (23) implies

$$\|F_k\|_2 \leq (\mu^*)^{-1}(M+1) < \infty, \qquad \forall k. \tag{25}$$

Now, a direct expansion of the products yields

$$\begin{aligned}G_n - \overline{G}_n = \mu &\sum_{i=nT}^{(n+1)T-1}\{[F_i - E(F_i)] \\ &+ \mu^2[A_0 + \mu A_1 + \cdots + \mu^{T-2}A_{T-2}]\}\end{aligned} \tag{26}$$

for some matrices $A_0, A_1, \cdots, A_{T-2}$ depending on $\{F_i\}$ and T. From (24)–(26) we know that there is $M(\mu)$ such that $\forall \mu \in (0, \mu^*]$

$$||A_0 + \mu A_1 + \cdots + \mu^{T-2} A_{T-2}||_2 \le M(\mu) < \infty.$$

Thus by applying Lemma 4.1 (to each element of the matrix under consideration), we know that

$$||A_i||_2 \le M < \infty, \qquad i = 0, 1, \cdots, T-2 \tag{27}$$

where M may depend on T (but not on μ).

Now, by using (27), it follows from (26) that

$$||G_n - \overline{G}_n||_2 \le \mu \left|\left| \sum_{i=nT}^{(n+1)T-1} [F_i - E(F_i)] \right|\right|_2 + O(\mu^2).$$

Hence Lemma 4.2 follows immediately from the assumption $\{F_k\} \in \mathcal{M}_2$.

Proof of Theorem 3.1: For any $T \ge 2$, let G_n and $\overline{G}_n$ be defined as in Lemma 4.2. For any vector x with $||x|| = 1$, introduce a sequence $\{x_n\}$ as follows

$$x_{n+1} = \overline{G}_n x_n, \qquad \forall n \ge m, x_m = x.$$

Then $\forall n > m$

$$\begin{aligned} x_{n+1} &= G_n x_n + (\overline{G}_n - G_n) x_n \\ &= \left(\prod_{i=m}^{n} G_i\right) x_m - \sum_{i=m}^{n} \left[\prod_{j=i+1}^{n} G_j\right] (G_i - \overline{G}_i) x_i. \end{aligned}$$

Note that $\{x_i\}$ is a deterministic process, by the Schwarz inequality it follows that

$$||x_{n+1}|| \le \left|\left| \prod_{i=m}^{n} G_i \right|\right|_2 + \sum_{i=m}^{n} \left|\left| \prod_{j=i+1}^{n} G_j \right|\right|_2 ||G_i - \overline{G}_i||_2 ||x_i||.$$

Since $\{F_k\} \in \mathcal{S}_2(\mu^*)$, we have $\forall n \ge i, \forall \mu \in (0, \mu^*]$

$$\left|\left| \prod_{j=i+1}^{n} G_j \right|\right|_2 = \left|\left| \prod_{(i+1)T}^{(n+1)T-1} (I - \mu F_j) \right|\right|_2 \le M(1-\mu\alpha)^{(n-i)T}$$

for some $M > 0, \alpha \in (0, 1)$.

Therefore, from this and Lemma 4.2, it follows that

$$\begin{aligned} ||x_{n+1}|| \le M(1-\mu\alpha)^{(n-m+1)T} &+ M \sum_{i=m}^{n} (1-\alpha\mu)^{(n-i)T} \\ &\times \{\mu T f(T) + O(\mu^2)\} ||x_i||. \end{aligned} \tag{28}$$

Now, let T be suitably large so that

$$f(T) \le \frac{\alpha}{2M}. \tag{29}$$

Then by denoting

$$\begin{aligned} z_n &= ||x_n|| (1-\mu\alpha)^{-nT} \\ g(\mu) &= \left[\frac{\alpha}{2}\mu T + O(\mu^2)\right] (1-\alpha\mu)^{-T} \end{aligned} \tag{30}$$

we have from (28) and (29)

$$z_{n+1} \le M(1-\mu\alpha)^{-mT} + g(\mu) \sum_{i=m}^{n} z_i.$$

Consequently, by the Bellman–Gronwall inequality we have

$$\begin{aligned} z_{n+1} &\le M(1-\mu\alpha)^{-mT} \left\{ 1 + \sum_{i=m}^{n} (1+g(\mu))^{n-i} \cdot g(\mu) \right\} \\ &\le 2M(1-\mu\alpha)^{-mT} [1+g(\mu)]^{n-m+1}. \end{aligned}$$

Therefore, by using (30) and the inequality $(1+x)^a \le 1 + ax, x \ge 0, a \in (0, 1]$, we get

$$\begin{aligned} ||x_{n+1}|| &\le 2M(1-\mu\alpha)^{(n-m+1)T} \\ &\quad \cdot \{[1+g(\mu)]^{1/T}\}^{(n-m+1)T} \\ &\le 2M\{(1-\mu\alpha)(1+T^{-1}g(\mu))\}^{(n-m+1)T} \\ &= 2M\Big\{(1-\mu\alpha)\Big[1 + \Big(\frac{\alpha}{2}\mu + O(\mu^2)\Big) \\ &\quad \cdot (1-\alpha\mu)^{-T}\Big]\Big\}^{(n-m+1)T}. \end{aligned}$$

From this, it is obvious that there is $\mu_1 \in (0, \mu^*]$ such that $\forall n \ge m$

$$||x_{n+1}|| \le 2M\Big(1 - \frac{\alpha}{3}\mu\Big)^{(n-m+1)T}, \qquad \forall \mu \in (0, \mu_1].$$

From here it is not difficult to conclude that $\{F_k\} \in \mathcal{S}(\mu_1)$. This completes the proof. #

V. The Proof of Theorem 3.2

First of all, by $\{F_k\} \in \mathcal{S}$, there exist constants $M > 0, \alpha \in (0, 1), \mu^* \in (0, 1)$ such that $\forall k \ge i \ge 0$

$$\left|\left| \prod_{j=i+1}^{k} (I - \mu E[F_j]) \right|\right| \le M(1-\mu\alpha)^{k-i}, \qquad \forall \mu \in (0, \mu^*]. \tag{31}$$

The proof is then divided into two lemmas.

Lemma 5.1: Let G_n and $\overline{G}_n$ be defined as in Lemma 4.2 and $\{\delta_i\}$ be defined by

$$\delta_i = M(1-\alpha\mu)^{-T} ||G_i - \overline{G}_i||, \qquad T > 0$$

where the constants M and α are defined in (31). Under Condition i) if (17) holds for some $r > 1$, then there exists a function $g(T)$ with $g(T) = o(T)$ such that for any fixed T and for all small μ

$$\left|\left| \prod_{j=i+1}^{n} (1+\delta_j) \right|\right|_q \le M[1 + \mu g(T) + o(\mu)]^{n-i}, \qquad \forall n \ge i \ge 0, \quad \forall q < r. \tag{32}$$

Proof: First, note that by (26)

$$G_i - \overline{G}_i = \mu S_i^{(T)} + \mu^2 H_i \tag{33}$$

where $S_i^{(T)}$ is defined by (16). We shall drop the argument (T) in the sequel. Moreover

$$\mu^2 H_i = \mu^2 H_i(0) + \mu^3 H_i(1) + \cdots + \mu^T H_i(T-2) + O(\mu^2)$$

with

$$H_i(d) = \sum F_{j_{d+2}} \cdots F_{j_1}, \qquad d = 0, \cdots, T-2$$

where the summation is over

$$iT \leq j_1 < j_2 < \cdots < j_{d+2} \leq (i+1)T - 1.$$

Denote $c = M(1-\mu\alpha)^{-T}$ and $t = (q^{-1} - r^{-1})^{-1}$, we have by the definition of δ_i and (33)

$$\begin{aligned}\left\| \prod_{j=i+1}^{n} (1+\delta_j) \right\|_q &= \left\| \prod_{j=i+1}^{n} (1 + c\|G_j - \overline{G}_j\|) \right\|_q \\ &\leq \left\| \prod_{j=i+1}^{n} (1 + \mu c\|S_j\| + \mu^2 c\|H_j\|) \right\|_q \\ &\leq \left\| \prod_{j=i+1}^{n} (1 + \mu c\|S_j\|) \prod_{j=i+1}^{n} (1 + \mu^2 c\|H_j\|) \right\|_q \\ &\leq \left\| \prod_{j=i+1}^{n} (1 + \mu c\|S_j\|) \right\|_r \\ &\quad \cdot \left\| \prod_{j=i+1}^{n} (1 + \mu^2 c\|H_j\|) \right\|_t. \end{aligned} \tag{34}$$

We now proceed to estimate the last term in the above inequality. Without loss of generality, we may assume in Condition i) that $0 < \delta < 1$. Then, by the relationship between geometric mean and arithmetic mean and the following inequality

$$1 + x \leq \exp(\beta x^{1/\beta}), \qquad \forall \beta \geq 1$$

with $\beta = k/1+\delta, k \geq 2$, we have for any nonnegative numbers $x_i, i \geq 1$

$$\begin{aligned} 1+\mu^k x_1 \cdots x_k &\leq \exp\left\{ \frac{k}{1+\delta} \cdot \mu^{1+\delta} (x_1 \cdots x_k)^{1+\delta/k} \right\} \\ &\leq \exp\left\{ \frac{k}{1+\delta} \cdot \mu^{1+\delta} \left(\frac{x_1 + \cdots + x_k}{k} \right)^{1+\delta} \right\} \\ &\leq \exp\left\{ \frac{\mu^{1+\delta}}{(1+\delta)k^{\delta}} (x_1 + \cdots + x_k)^{1+\delta} \right\} \\ &\leq \exp\left\{ \frac{\mu^{1+\delta}}{1+\delta} (x_1^{1+\delta} + \cdots + x_k^{1+\delta}) \right\} \end{aligned} \tag{35}$$

and so for $c \geq 1$

$$1 + \mu^k c x_1 \cdots x_k \leq \exp\left\{ \frac{c\mu^{1+\delta}}{1+\delta} (x_1^{1+\delta} + \cdots + x_k^{1+\delta}) \right\}.$$

Applying this to each $H_i(k)$ gives

$$\begin{aligned} &(1 + \mu^2 c\|H_i\|) \\ &\leq \prod_{k=0}^{T-2} (1 + \mu^{k+2} c\|H_i(k)\|)(1 + O(\mu^2)) \\ &\leq \prod_{k=0}^{T-2} \prod (1 + \mu^{k+2} c\|F_{j_{k+2}} \cdots F_{j_1}\|)(1 + O(\mu^2)) \\ &\leq \prod_{k=0}^{T-2} \prod \exp\left\{ \frac{c\mu^{1+\delta}}{1+\delta} \cdot \sum_{s=1}^{k+2} \|F_{j_s}\|^{1+\delta} \right\} \\ &\quad \cdot (1 + O(\mu^2)) \\ &\leq \exp\left\{ \frac{2^T c}{1+\delta} \cdot \mu^{1+\delta} \sum_{j=iT}^{(i+1)T-1} \|F_j\|^{1+\delta} \right\} \\ &\quad \cdot (1 + O(\mu^2)). \end{aligned}$$

The second products in the intermediary steps are taken over $iT \leq j_1 < j_2 \cdots < j_{k+2} \leq (i+1)T - 1$. Consequently

$$\begin{aligned} E\left\| \prod_{j=i+1}^{n} (1 + \mu^2 c\|H_j\|) \right\|^t &\leq \exp\left\{ \frac{t2^T c}{1+\delta} \cdot \mu^{1+\delta} \sum_{j=(i+1)T}^{(n+1)T-1} \|F_j\|^{1+\delta} \right\} \\ &\quad \times (1 + O(\mu^2))^{t(n-i)}. \end{aligned} \tag{36}$$

Now, for fixed T take μ small enough so that

$$\frac{t2^T c}{1+\Delta} \cdot \mu^{1+\delta} \leq \varepsilon$$

then by Condition i) we have from (36)

$$\begin{aligned} \left\| \prod_{j=i+1}^{n} (1 + \mu^2 c\|H_i\|) \right\|_t &\leq \exp\left\{ \frac{2^T K c}{\varepsilon(1+\delta)} \cdot \mu^{1+\delta} (n-i) T \right\} (1 + O(\mu^2))^{n-i} \\ &\leq \left(1 + \frac{2^T K T c}{\varepsilon(1+\delta)} \cdot \mu^{1+\delta} + O(\mu^2) \right)^{n-i}. \end{aligned}$$

Finally, putting this into (34) and using condition (17) we see that (32) is true.

Lemma 5.2: Let Condition i) be satisfied. If (17) holds for some $r > 1$, then

$$\{F_k\} \in \mathcal{S} \Rightarrow \{F_k\} \in \mathcal{S}_p, \qquad \forall p < r$$

and consequently Theorem 3.2 holds.

Proof: For any vector x with $\|x\| = 1$, similar to the proof of Theorem 3.1, introduce a sequence $\{y_n\}$ as follows

$$y_{n+1} = G_n y_n, \qquad \forall n \geq m, y_m = x.$$

(Note, however, this is a random equation and is no longer deterministic.) Then $\forall n > m$

$$y_{n+1} = \overline{G}_n y_n + (G_n - \overline{G}_n) y_n = \left(\prod_{i=m}^{n} \overline{G}_i\right) x + \sum_{i=m}^{n} \left[\prod_{j=i+1}^{n} \overline{G}_j\right] (G_i - \overline{G}_i) y_i.$$

Hence by (31)

$$\|y_{n+1}\| \le M(1-\mu\alpha)^{(n+1-m)T} + M \sum_{i=m}^{n} (1-\mu\alpha)^{(n-i)T} \|G_i - \overline{G}_i\| \cdot \|y_i\| \quad (37)$$

consequently, by the definition of δ_i in Lemma 5.1 and the Gronwall inequality

$$\|y_{n+1}\| \le M(1-\mu\alpha)^{(n+1-m)T} \times \left\{1 + \sum_{i=m}^{n} \prod_{j=i+1}^{n} (1+\delta_j)\delta_i\right\}. \quad (38)$$

Since $p < r$, we can always find a number q such that $p < q < r$, then from (38) and by the Hölder inequality (with $s = (p^{-1} - q^{-1})^{-1}$)

$$\|y_{n+1}\|_p \le M(1-\mu\alpha)^{(n+1-m)T} \times \left\{1 + \sum_{i=m}^{n} \left\|\prod_{j=i+1}^{n} (1+\delta_j)\right\|_q \|\delta_i\|_s\right\}. \quad (39)$$

Note that by Condition i), $\|\delta_i\|_s = O(\mu), \forall s > 1$, so by (32) and (39)

$$\begin{aligned}\|y_{n+1}\|_p &\le M(1-\mu\alpha)^{(n+1-m)T} \times \{1 + O([1+\mu g(T) + o(\mu)]^{n-m+1})\} \\ &= O(\{(1-\mu\alpha)[1+\mu T^{-1} g(T) + o(\mu)]\}^{(n-m+1)T}) \\ &= O(\{1 - [\alpha - T^{-1}g(T)]\mu + o(\mu)\}^{(n-m+1)T}) \\ &= O\left(\left\{1 - \frac{\alpha}{2}\mu + o(\mu)\right\}^{(n-m+1)T}\right)\end{aligned}$$

provided that T is so large that $T^{-1}g(T) < \alpha/2$. Hence from here it is easy to see that Lemma 5.2 is true. #

VI. The Proofs of Corollaries 3.1 and 3.2

The proof of Corollary 3.1 is prefaced by the following two lemmas, whose proofs are given in Appendix C.

Lemma 6.1: Let $\{x_i\}$ be a scalar ϕ-mixing process with mixing rate $\phi(i)$. If there are constants $\delta \ge 0, c \ge 0$, such that

$$E|x_i| \le \delta, \quad |x_i| \le c, \quad \forall i$$

then for any $n \ge m \ge 0$

$$E \prod_{i=m}^{n} (1 + |x_i|) \le 2(1+\beta)^{n-m+1}$$

where $\beta = \delta + 2c\phi(1)$.

Lemma 6.2: If, in addition to the condition of Lemma 6.1, $|x_i| \le c \le 1$, then for any $p \ge 1$

$$\left\|\prod_{i=m}^{n} (1 + |x_i|)\right\|_p \le 2^{1/p}(1+\beta_1)^{n-m+1}, \qquad \forall n \ge m \ge 1$$

where $\beta_1 = 2^{2p-1}(\delta + 2c\phi(2))$.

Proof of Corollary 3.1: Let the mixing rate of $\{F_k\}$ be $\{\phi(k)\}$. By Remark 3.4 we need only to prove (18) for any $q > 1$, by taking $c_T = T/\sqrt{\phi(T+1)}$. To this end, set

$$x_i = \mu \|S_i\| I(\|S_i\| \le c_T).$$

Since $S_i = S_i^{(T)}$ is defined by (16), it is obvious that $\{x_i\}$ is also a ϕ-mixing process with mixing rate $\{\phi((i-1)T+1)\}$. Moreover

$$x_i \le \mu \frac{T}{\sqrt{\phi(T+1)}}$$

which is less than one for suitably small μ (with T fixed).

Also, since $\{F_i\}$ is ϕ-mixing and has bounded moments, by the ϕ-mixing inequality of Ibragimov [21], we know that

$$\|E\{[F_t - EF_t][F_s - EF_s]^\tau\}\| \le c\sqrt{\phi(|t-s|)}, \qquad \forall t\ \forall s$$

where c is a constant. Consequently, by the definition of S_i in (16) it is readily verified that there exists a function $f(T)$ with $f(T) = o(T)$ such that

$$E\|S_i\| \le \|S_i\|_2 \le f(T), \qquad \forall i, \forall T.$$

Hence, by the definition of x_i it follows that

$$\sup_i Ex_i \le \mu f(T).$$

Thus, applying Lemma 6.2, we have for any $q \ge 1$

$$\left\|\prod_{i=m}^{n} (1 + x_i)\right\|_q \le 2\{1 + \mu[f(T) + 2T\sqrt{\phi(T+1)}]2^{2q-1}\}^{n-m+1}$$

and hence (18) is true for any $q > 1$. This completes the proof of Corollary 3.1.

Proof of Corollary 3.2: By Example 3.2, we need only to verify Condition ii) of Theorem 3.2.

First, for any random variable $x \ge 0$, if $x \le 1$, then it is readily verified that (cf., [15])

$$E\exp(x) \le \exp(2Ex).$$

By this, we may estimate $E\exp(x)$ in the general case as follows

$$\begin{aligned}E\exp(x) &= E\exp\{x[I(x \le \tfrac{1}{2}) + I(x > \tfrac{1}{2})]\} \\ &\le \|\exp\{xI(x \le \tfrac{1}{2})\}\|_2 \cdot \|\exp\{xI(x > \tfrac{1}{2})\}\|_2 \\ &\le \exp(2Ex) \cdot \{1 + \|[\exp(x)]I(x > \tfrac{1}{2})\|_2\}. \quad (40)\end{aligned}$$

Now, denote

$$\varepsilon_t = \sum_{i=t}^{t+T-1} \tilde{V}_i \quad \tilde{V}_i = V_i - EV_i \quad (41)$$

then there exits a constant $c > 0$ such that

$$E\|\varepsilon_t\| \le c\sqrt{T} \qquad \forall t, \forall T.$$

Moreover, for any constant $b > 0$ and small $\mu > 0$

$$\begin{aligned} &E\exp(2\mu b\|\varepsilon_t\|)I\left(\mu b\|\varepsilon_t\| > \frac{1}{2}\right) \\ &\quad \le E\exp(2\mu b\|\varepsilon_t\| + \sqrt{\mu b}\|\varepsilon_t\|)\exp\left(-\frac{1}{2\sqrt{\mu b}}\right) \\ &\quad = O(\mu^2 b). \end{aligned}$$

Hence, by (40) we have

$$E\exp(\mu b\|\varepsilon_t\|) \le \exp\{2\mu bc[\sqrt{T} + O(\mu)]\}. \tag{42}$$

Note that

$$S_i = \sum_{k=iT}^{(i+1)T-1}(F_k - E[F_k]) = \sum_{j=-\infty}^{\infty} A_j\left[\sum_{k=iT}^{(i+1)T-1}\tilde{V}_{k-j}\right]$$

we have

$$\begin{aligned} &E\exp\left\{\mu\sum_{i=m+1}^{n}\|S_i\|\right\} \\ &\le E\exp\left\{\mu\sum_{i=m+1}^{n}\sum_{j=-\infty}^{\infty}\|A_j\|\cdot\left\|\sum_{k=iT}^{(i+1)T-1}\tilde{V}_{k-j}\right\|\right\} \\ &= E\exp\left\{\mu\sum_{i=m+1}^{n}\sum_{j=-\infty}^{\infty}\|A_j\|\cdot\left\|\sum_{s=iT-j}^{(i+1)T-1-j}\tilde{V}_s\right\|\right\} \\ &= E\exp\left\{\mu\sum_{i=m+1}^{n}\sum_{t=-\infty}^{\infty}\|A_{iT-t}\|\cdot\left\|\sum_{s=t}^{t+T-1}\tilde{V}_s\right\|\right\} \\ &= E\exp\left\{\mu\sum_{t=-\infty}^{\infty}\left(\sum_{i=m+1}^{n}\|A_{iT-t}\|\right)\|\varepsilon_t\|\right\} \\ &= E\exp\left\{\mu\sum_{t=-\infty}^{\infty}f_t\|\varepsilon_t\|\right\}, \left(f_t = \sum_{i=m+1}^{n}\|A_{iT-t}\|\right) \\ &= E\exp\left\{\mu\sum_{j=0}^{T-1}\sum_{t=-\infty}^{\infty}f_{tT+j}\|\varepsilon_{tT+j}\|\right\} \\ &\le \left\{\prod_{j=0}^{T-1}E\exp\left(\mu T\sum_{t=-\infty}^{\infty}f_{tT+j}\|\varepsilon_{tT+j}\|\right)\right\}^{1/T} \\ &= \left\{\prod_{j=0}^{T-1}\prod_{t=-\infty}^{\infty}E\exp(\mu Tf_{tT+j}\|\varepsilon_{tT+j}\|)\right\}^{1/T}. \end{aligned} \tag{43}$$

But, by (42) we know that for some constant $c > 0$

$$E\exp(\mu Tf_{tT-j}\|\varepsilon_{tT-j}\|) \le \exp\{\mu f_{tT+j}c[T^{3/2} + O(\mu)]\}.$$

Hence by (43) we have

$$\begin{aligned} &E\exp\left\{\mu\sum_{i=m+1}^{n}\|S_i\|\right\} \\ &\le \exp\left\{\sum_{j=0}^{T-1}\sum_{t=-\infty}^{\infty}\mu cf_{tT+j}[\sqrt{T} + O(\mu)]\right\} \\ &= \exp\left\{\sum_{t=-\infty}^{\infty}\mu cf_t[\sqrt{T} + O(\mu)]\right\} \\ &= \exp\left\{\sum_{i=m+1}^{n}\mu c\left(\sum_{t=-\infty}^{\infty}\|A_{iT-t}\|\right)[\sqrt{T} + O(\mu)]\right\} \\ &= \exp\left\{\left[\mu c\left(\sum_{t=-\infty}^{\infty}\|A_t\|\right)\sqrt{T} + O(\mu^2)\right](n-m)\right\}. \end{aligned}$$

Hence Condition ii) of Theorem 3.2 holds.

VII. The Proof of Theorem 3.5

The proof is prefaced with two lemmas.

Lemma 7.1: Let $\{P_k\}$ be defined by (10). If $\|\varphi_k\|_{2p} = O(1)$, $p \ge 1$, then

$$\sup_{\mu\in(0,1)}\sup_{k\ge 1}\|P_k^{-1}\|_p < \infty.$$

Proof: By (10) and the matrix inversion formula

$$P_k = [P_{k-1}^{-1} + \mu R^{-1}\varphi_k\varphi_k^\tau]^{-1} + \mu Q. \tag{44}$$

Denote $u_k = \lambda_{\max}(P_k^{-1}) = \frac{1}{\lambda_{\min}(P_k)}$, then by (44)

$$\lambda_{\min}(P_k) \ge [u_{k-1} + \mu R^{-1}\|\varphi_k\|^2]^{-1} + \mu\lambda_{\min}(Q).$$

Consequently

$$\begin{aligned} u_k &\le \{[u_{k-1} + \mu R^{-1}\|\varphi_k\|^2]^{-1} + \mu\lambda_{\min}(Q)\}^{-1} \\ &= \frac{u_{k-1} + \mu R^{-1}\|\varphi_k\|^2}{1 + \mu(u_{k-1} + \mu R^{-1}\|\varphi_k\|^2)\lambda_{\min}(Q)} \\ &\le \frac{u_{k-1} + \mu R^{-1}\|\varphi_k\|^2}{1 + \mu u_{k-1}\lambda_{\min}(Q)} \\ &\le \frac{u_{k-1}}{1 + \mu u_{k-1}\lambda_{\min}(Q)} + \mu R^{-1}\|\varphi_k\|^2 \\ &= \frac{u_{k-1}}{1 + \mu u_{k-1}\lambda_{\min}(Q)} \\ &\quad \times [I(u_{k-1} \ge \lambda_{\min}^{-1}(Q)) + I(u_{k-1} < \lambda_{\min}^{-1}(Q))] \\ &\quad + \mu R^{-1}\|\varphi_k\|^2 \\ &\le \frac{u_{k-1}}{1+\mu}I(u_{k-1} \ge \lambda_{\min}^{-1}(Q)) \\ &\quad + u_{k-1}I(u_{k-1} < \lambda_{\min}^{-1}(Q)) + \mu R^{-1}\|\varphi_k\|^2 \\ &= \frac{1}{1+\mu}u_{k-1} + \left(1 - \frac{1}{1+\mu}\right)u_{k-1}I(u_{k-1} < \lambda_{\min}^{-1}(Q)) \\ &\quad + \mu R^{-1}\|\varphi_k\|^2 \\ &\le \left(\frac{1}{1+\mu}\right)u_{k-1} + \mu(\lambda_{\min}^{-1}(Q) + R^{-1}\|\varphi_k\|^2). \end{aligned} \tag{45}$$

From this it is evident that Lemma 7.1 holds.

Lemma 7.2: Let condition (22) hold. Then for $\{P_k\}$ defined by (10)

$$\sup_{\mu\in(0,1)}\sup_{k\geq 1}\|P_k\|_p<\infty, \qquad \forall p\geq 1.$$

Proof: Denote

$$T_k=\sum_{i=(k-1)h+1}^{kh}\operatorname{tr}(P_{i+1}), \qquad T_0=1. \tag{46}$$

Then similar to (4.4) in [16], we have

$$T_{k+1}\leq(1-\mu a_{k+1})T_k+O(\mu) \tag{47}$$

where

$$a_{k+1}=\frac{\operatorname{tr}\left[(P_{kh+1}+\mu hQ)^2\sum_{i=kh+1}^{(k+1)h}\frac{\varphi_i\varphi_i^\tau}{1+\|\varphi_i\|^2}\right]}{h(R+1)[1+\lambda_{\max}(P_{kh+1}+\mu hQ)]\operatorname{tr}(P_{kh+1}+\mu hQ)}.$$

Note that

$$\begin{aligned}
&E[a_{k+1}|\mathcal{G}_{kh}]\\
&\geq\frac{(1+h)\operatorname{tr}(P_{kh+1}+\mu hQ)^2\lambda_k}{h(R+1)[1+\lambda_{\max}(P_{kh+1}+\mu hQ)]\operatorname{tr}(P_{kh+1}+\mu hQ)}\\
&\geq\frac{(1+h)\lambda_k}{dh(R+1)}\cdot\frac{\lambda_{\max}(P_{kh+1}+\mu hQ)}{1+\lambda_{\max}(P_{kh+1}+\mu hQ)}, \quad (d=\dim(\varphi_k))\\
&\geq\frac{(1+h)\lambda_k}{dh(R+1)}\cdot\frac{\lambda_{\min}(P_{kh+1})}{1+\lambda_{\min}(P_{kh+1})},\\
&\quad\left(\frac{x}{1+x}, x>0, \text{ is increasing}\right)\\
&=\frac{(1+h)\lambda_k}{dh(R+1)}\cdot\frac{1}{1+\lambda_{\max}(P_{kh+1}^{-1})}\\
&\geq\frac{1+h}{dh(R+1)}\cdot\frac{\lambda_k}{1+u_{kh+1}}
\end{aligned} \tag{48}$$

where u_k is the same as that in (45). By (45), there is a constant $c>1+\|P_0^{-1}\|$ such that

$$u_k\leq\left(1-\frac{\mu}{2}\right)u_{k-1}+\frac{\mu}{2}c(\|\varphi_k\|^2+1), \qquad k\geq 1, \quad \mu\in(0,1).$$

Consequently, $u_k\leq cb_k, \forall k\geq 1$, where b_k is defined by (21) with $\delta=\mu/2$. Thus (48) gives

$$E[a_{k+1}|\Gamma_{kh}]\geq\frac{1+h}{dch(R+1)}\cdot\frac{\lambda_k}{1+b_{kh+1}}.$$

From this, by Lemmas 2.1 and 2.3 in [16] and condition (22) we know that

$$\{a_k\}\in\mathcal{S}^0(\lambda), \quad \text{for some} \quad \lambda\in(0,1).$$

Note that $a_k\in[0.1/1+R]$. Then by using Lemma 2.3 in [16] again, we derive

$$\{\mu a_k\}\in\mathcal{S}^0(\lambda^{\mu R(1+R)^{-1}}), \qquad \forall\mu\in(0,1). \tag{49}$$

From this and (47) it is evident that $\{T_k\}$ and hence $\{P_k\}$ is uniformly bounded in $L_p, \forall p\geq 1$; this completes the proof. #

Proof of Theorem 3.5: First, (10) can be rewritten as

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

where

$$Q_k=\mu RL_kL_k^\tau+\mu Q.$$

Hence, the desired result will follow directly from Theorem 3.4 if we can prove that Conditions i) and ii) are true.

By Lemmas 7.1 and 7.2 and the Hölder inequality, it is easy to see that Condition ii) of Theorem 3.4 holds for $\forall t<p$, so we need only to verify Condition i).

Note that

$$\begin{aligned}
\frac{1}{1+\|Q_k^{-1}P_{k+1}\|}&\geq\frac{1}{1+\|Q_k^{-1}\|\cdot\|P_{k+1}\|}\\
&\geq\frac{1}{1+\mu^{-1}\|Q^{-1}\|\cdot\|P_{k+1}\|}\\
&\geq\frac{\mu}{1+\|Q^{-1}\|\cdot\|P_{k+1}\|}.
\end{aligned} \tag{50}$$

Now, denote

$$x_k=(h+\|Q^{-1}\|T_k)\mu^{-1}$$

where T_k is defined by (46). Then it follows from (47) that

$$x_{k+1}\leq(1-\mu a_{k+1})x_k+O(1).$$

Consequently, by Lemma 3.1 in [16] and (49), we get

$$\left\{\frac{1}{x_k}\right\}\in\mathcal{S}^0(1-\alpha\mu), \quad \text{for some} \quad \alpha>0, \qquad \forall\mu\in(0,1). \tag{51}$$

Note, however, that

$$x_k=\sum_{i=(k-1)h+1}^{kh}\mu^{-1}[1+\|Q^{-1}\|\operatorname{tr}(P_{i+1})]$$

from (51) it can be derived that (cf., [14, Lemma 5])

$$\left\{\frac{\mu}{1+\|Q^{-1}\|\cdot\|P_{k+1}\|}\right\}\in\mathcal{S}^0(1-\beta\mu)$$

for some $\beta>0$, and $\forall\mu\in(0,1)$, from this and (50) it is evident that Condition i) of Theorem 3.4 holds, and this completes the proof of Theorem 3.5. #

VIII. Conclusions

Exponential stability of a tracking algorithm is the key both to its practical usefulness and to its analysis. In many cases the properties of the averaged dynamic equation (13) are easy to study. Exponential stability of this deterministic, linear difference equation often follows in a straightforward way from persistence of excitation conditions for the regressors in the underlying estimation problem. The stability of the actual, stochastic, difference equation (12), however, is much more difficult to establish. The most relevant earlier results, [12] and [13], have dealt with continuous time, stochastic differential equations. The assumptions used in these articles include that the coefficient matrices are bounded and satisfy a certain mixing condition.

Our main result, Theorem 3.2, deals with the discrete-time case. It establishes that the stability of the averaged equation implies the stability of the stochastic one under two quite general conditions: one being that the tails of the distributions of the random matrices decay sufficiently fast and the other that the dependence between the matrices also decays sufficiently fast. This is the case, for example, when they are linearly generated using a stable filter (Corollary 3.2).

Exponential stability of the homogeneous tracking algorithm is sufficient to directly obtain useful results, e.g., about error propagation in the algorithm and bounded tracking error. It is also a necessary building block in establishing more accurate measurements of the tracking performance.

This requires additional work, however, and we refer to the companion paper [17] for such results.

Appendix A
Proof of Example 3.2

Denote $\lambda = \Sigma_{j=-\infty}^{\infty} \lambda_j$, by the Hölder inequality

$$\|F_k\|^{1+\delta} \le \lambda^\delta \left(\sum_{j=-\infty}^{\infty} \lambda_j w_{k-j}^{1+\delta} \right).$$

Then for small $\varepsilon > 0$ we have

$$\begin{aligned} &E \exp\left\{ \varepsilon \sum_{k=i+1}^{n} \|F_k\|^{1+\delta} \right\} \\ &\le \prod_{j=-\infty}^{\infty} E \exp\left\{ \varepsilon \lambda^\delta \left(\sum_{k=i+1}^{n} \lambda_{k-j} \right) w_j^{1+\delta} \right\} \\ &\le \prod_{j=-\infty}^{\infty} \exp\left\{ b\varepsilon\lambda^\delta \alpha^{-1} \sum_{k=i+1}^{n} \lambda_{k-j} \right\} \\ &= \exp\left\{ b\varepsilon\lambda^\delta\alpha^{-1} \sum_{k=i+1}^{n} \sum_{j=-\infty}^{\infty} \lambda_{k-j} \right\} \\ &= \exp\{ b\varepsilon\lambda^{1+\delta}\alpha^{-1}(n-i)\}. \end{aligned}$$

Appendix B
The Proof of "(18) $\Rightarrow$ (17)"

By Condition i) of Theorem 3.2, we have for any $\beta > 0$

$$\begin{aligned} &E \prod_{j=i+1}^{n} [1 + \mu\|S_j\| I(\|S_j\| > c_T)]^\beta \\ &\le E \exp\left\{ \mu\beta \sum_{j=i+1}^{n} \|S_j\| I(\|S_j\| > c_T) \right\} \\ &\le E \exp\left\{ \mu\beta \sum_{j=i+1}^{n} \|S_j\|^{1+\delta} c_T^{-\delta} \right\} \\ &\le E \exp\left\{ \mu\beta \left(\frac{T}{c_T}\right)^\delta \sum_{j=(i+1)T}^{(n+1)T-1} \|F_j\|^{1+\delta} \right\} \\ &\le E \exp\left\{ \varepsilon^{-1}\mu\beta K \left(\frac{T}{c_T}\right)^\delta T(n-i) \right\}. \end{aligned} \tag{B.1}$$

Note that as $T \to \infty, T/c_T \to 0$, so by the Hölder inequality, (17) follows from (18) and (B.1) immediately.

Appendix C
Proofs of Lemmas 6.1 and 6.2

Proof of Lemma 6.1: Denote $z_n = \Pi_{i=m}^{n}(1 + |x_i|)$, $z_{m-1} = 1$. Then

$$z_n = z_{n-1} + |x_n| z_{n-1} = 1 + \sum_{i=m}^{n} |x_i| z_{i-1}. \tag{C.1}$$

By Theorem A.6 in [19] it is known that

$$|E(|x_i| z_{i-1}) - E|x_i| E z_{i-1}| \le 2\phi(1) \cdot c \cdot E z_{i-1}.$$

Consequently, it follows that

$$E[|x_i| z_{i-1}] \le \beta E z_{i-1}.$$

Substituting this into (C.1) we get

$$E z_n \le 1 + \beta \sum_{i=m}^{n} E z_{i-1}$$

hence by the Bellman–Gronwall inequality

$$E z_n \le 1 + \sum_{i=m-1}^{n-1} (1+\beta)^{n-i+1} \beta \le 1 + (1+\beta)^{n-m+1}.$$

This completes the proof. #

Proof of Lemma 6.2: By the Schwarz inequality and the elementary inequality $(1+x)^{2p} \le 1 + p2^{2p}x, 0 \le x \le 1$, it follows that

$$\begin{aligned} &\left\| \prod_{i=m}^{n} (1 + |x_i|) \right\|_p \\ &\le \left\| \prod_{\substack{m \le i \le n \\ i \text{ odd}}} (1 + |x_i|) \right\|_{2p} \cdot \left\| \prod_{\substack{m \le i \le n \\ i \text{ even}}} (1 + |x_i|) \right\|_{2p} \\ &\le \left\{ E \prod_{\substack{m \le i \le n \\ i \text{ odd}}} (1 + p2^{2p}|x_i|) \right\}^{1/2p} \\ &\times \left\{ E \prod_{\substack{m \le i \le n \\ i \text{ even}}} (1 + p2^{2p}|x_i|) \right\}^{1/2p}. \end{aligned} \tag{C.2}$$

Since the process $\{x_{2k}\}$ is also ϕ-mixing with mixing rate $\phi(2k)$, by Lemma 6.1, we have $\forall n \ge m$

$$\begin{aligned} &\left\{ E \prod_{k=m}^{n} (1 + p2^{2p}|x_{2k}|) \right\}^{1/2p} \\ &\le \{2(1 + 2p\beta_1)^{n-m+1}\}^{1/2p} \\ &\le 2^{1/2p}(1+\beta_1)^{n-m+1} \end{aligned} \tag{C.3}$$

where $\beta_1 = 2^{2p-1}(\delta + 2c\phi(2))$.

Similarly, $\forall n \geq m$

$$\left\{ E \prod_{k=m}^{n} (1 + p2^{2p}|x_{2k+1}|) \right\}^{1/2p} \leq 2^{1/2p}(1+\beta_1)^{n-m+1}. \tag{C.4}$$

Finally, by (C.3) and (C.4), the desired result follows from (C.2). #

REFERENCES

[1] L. Arnold and W. Wihstutz, *Lyapunov Exponents* (Lecture Notes in Mathematics). vol. 1186. Berlin: Springer-Verlag, 1987.

[2] A. Benveniste, M. Metivier, and P. Priouret, *Algorithmes adaptifs et approximations stochastiques*, Masson, 1987.

[3] R. R. Bitmead, "Persistence of excitation conditions and the convergence of adaptive schemes," *IEEE Trans. Info. Theory*, vol. IT-30, no. 2, pp. 183–191, 1984.

[4] R. R. Bitmead, B. D. O. Anderson, and T. S. Ng, "Convergence rate determination for gradient based adaptive estimators," *Automatica*, vol. 25, no. 4, pp. 185–191, 1980.

[5] S. Bittanti and M. Campi, "Tuning the forgetting factor in RLS identification algorithms," in *Proc. IEEE Conf. Decis. Contr.*, Brighton, England, 1991, pp. 1688–1689.

[6] ——, "A fake information matrix approach to the analysis of finite memory RLS identification techniques," in *Proc. IEEE Conf. Decis. Contr.*, San Antonio, TX, 1993, pp. 2026–2031.

[7] G. Blankenship, "Stability of linear differential equations with random coefficients," *IEEE Trans. Automat. Contr.*, vol. AC–22, pp. 834–838, 1977.

[8] P. E. Caines, *Linear Stochastic Systems*. New York: Wiley, 1988.

[9] E. Eweda and O. Macchi, "Tracking error bounds of adaptive nonstationary filtering," *Automatica*, vol. 21, no. 3, pp. 293–302, 1985.

[10] H. Furstenberg and H. Kesten, "Products of random matrices," *Ann. Math. Stat.*, vol. 31, pp. 475–469, 1960.

[11] W. Gardner, "Nonstationary learning characteristics of the LMS algorithms: A general study, analysis and critique," *IEEE Trans. Circuits Syst.*, vol. CAS-34, no. 10, pp. 1199–1207, 1987.

[12] S. Geman, "Some averaging and stability results for random linear differential equations," *SIAM J. Appl. Math.*, vol. 36, 1979.

[13] L. Gerencser, "Almost sure exponential stability of random linear differential equations," *Stochstics and Stochastics Reports*, vol. 36, pp. 91–107, 1991.

[14] L. Guo, "Estimating time-varying parameters by Kalman filter based algorithms: Stability and convergence," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 141–147, 1990.

[15] ——, "On adaptive stabilization of time-varying stochastic systems," *SIAM J. Contr. Optim.*, vol. 28, pp. 1432–1451, 1990.

[16] ——, "Stability of recursive stochastic tracking algorithms," *SIAM J. Contr. Optim.*, vol. 32, pp. 1195–1225, 1994.

[17] L. Guo and L. Ljung, "Performance analysis of general tracking algorithms," *IEEE Trans. Automat. Contr.*, vol. 40, vol. 8, pp. 1388–1402, 1995.

[18] L. Guo, L. Ljung, and P. Priouret, "Performance analysis of the forgetting factor RLS algorithm," *Int. J. Adaptive Contr. Sig. Processing*, vol. 7, pp. 525–537, 1993.

[19] P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Applications*. New York: Academic, 1980.

[20] R. Z. Hasminskii, *Stochastic Stability of Differential Equations*. Alphen ann den Rijn, The Netherlands: Sijthoff and Nordhoff, 1980.

[21] I. A. Ibragimov, "Some limit theorems for staionary processes," *Theory Prob. Appl.*, vol. 7, pp. 349–382, 1962.

[22] E. F. Infante, "On the stability of some linear nonautonomous random systems," *ASME*, vol. 35, 1968.

[23] H. J. Kushner and H. Huang, "Asymptotic properties of stochastic approximations with constant coefficients," *SIAM J. Contr. Optim.*, vol. 19, no. 5, pp. 87–105, 1981.

[24] L. Ljung and S. Gunnarsson, "Adaptive tracking in system identification—A survey," *Automatica*, vol. 26, no. 1, pp. 7–22, 1990.

[25] L. Ljung and P. Priouret, "A result of the mean square error obtained using general tracking algorithms," *Int. J. Adaptive Contr.*, vol. 5, no. 4, pp. 231–250, 1991.

[26] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT press, 1983.

[27] S. P. Meyn and L. Guo, "Adaptive control of time-varying stochastic systems," in *Proc. 11th IFAC World Congress*, Tallin, USSR, vol. 3, 1990, pp. 198–202.

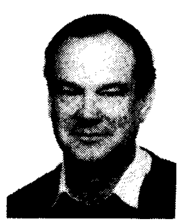
[28] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms: Stability and Performance*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[29] B. Widrow and S. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

Lei Guo was born in Shandong, China, in November, 1961. He received the B.S. degree in mathematics from Shanding University in 1982 and the M.S. and Ph.D. degrees in control theory from the Chinese Academy of Sciences in 1984 and 1987, respectively.

From1987 to 1989, he was with the Australian National University. Since 1992, he has been Professor with the Institute of Systems Science of the Chinese Academy of Sciences. He is the author of the book, *Time-Varying Stochastic Systems: Stability, Estimation, and Control* (in Chinese) and coauthor of the book, *Identification and Stochastic Adaptive Control* (Boston, MA: Birkhäuser, 1991).

Dr. Guo is the recipient of the Young Scientist Award of China (1994), the National Science Prize of the Chinese Academy of Sciences (1994), the Young Author Prize of the 12th IFAC World Congress (1993), and (with his coauthor) the National Natural Science Prize of China (1987). He was an Associate Editor of the *SIAM Journal on Control and Optimization* from 1991–1993 and is currently a Vice Chairman of the Theory Committee of the Chinese Association of Automation.



Lenart Ljung (S'74–M'75–SM'83–F'85) was born in 1946. He received the Ph.D. degree in automatic control from the Lund Institute of Technology, Lund, Sweden, in 1974.

Since 1976, he has been Professor of the Chair of Automatic Control at Linköping University, Linköping, Sweden. He is the author/coauthor of six textbooks.

Dr. Ljung is a member of the Royal Swedish Academy of Engineering Sciences and the Royal Swedish Academy of Sciences. He is an IFAC Advisor and an Associate Editor of several journals.