

Necessary and Sufficient Conditions for Stability of LMS

Lei Guo, *Senior Member, IEEE*, Lennart Ljung, *Fellow, IEEE*, and Guan-Jun Wang

Abstract—In a recent work [7], some general results on exponential stability of random linear equations are established which can be applied directly to the performance analysis of a wide class of adaptive algorithms, including the basic LMS ones, without requiring stationarity, independency, and boundedness assumptions of the system signals. The current paper attempts to give a complete characterization of the exponential stability of the LMS algorithms by providing a necessary and sufficient condition for such a stability in the case of possibly unbounded, nonstationary, and non- ϕ -mixing signals. The results of this paper can be applied to a very large class of signals, including those generated from, e.g., a Gaussian process via a time-varying linear filter. As an application, several novel and extended results on convergence and the tracking performance of LMS are derived under various assumptions. Neither stationarity nor Markov-chain assumptions are necessarily required in the paper.

Index Terms—Exponential stability, LMS algorithm, nonstationary signals, tracking performance.

I. INTRODUCTION

A. The Contribution

THE WELL-KNOWN least mean squares (LMS) algorithm, aiming at tracking the “best linear fit” of an observed (or desired) signal $\{y_k\}$ based on a measured d -dimensional (input) signal $\{\phi_k\}$, is defined recursively by

$$x_{k+1} = x_k + \mu\phi_k(y_k - \phi_k^T x_k), \quad x_0 \in \mathcal{R}^d \quad (1)$$

where $\mu > 0$ is a step-size.

Due to its simplicity, robustness, and ease of implementation, the LMS algorithm is known to be one of the most basic adaptive algorithms in many areas, including adaptive signal processing, system identification, and adaptive control, and it has received considerable attention in both theory and applications over the past several decades (see, among many others, books [20], [19], and [2], the survey [14], and the references therein). Also, it has been found recently that the LMS is H^∞ optimal in the sense that it minimizes the energy gain from the disturbances to the predicted errors, and it is also

Manuscript received November 16, 1995; revised September 6, 1996. 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, P.R. China (e-mail: Lguo@iss03.iss.ac.cn).

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

G.-J. Wang is with the Department of Mathematics, The Central University for Nationalities, Beijing 100081, P.R. China.

Publisher Item Identifier S 0018-9286(97)04290-6.

risk-sensitive optimal and minimizes a certain exponential cost function (see [11]).

In many situations, it is desirable to know at least the answers to the following questions.

- Is the LMS stable in the mean-squares sense?
- Does the LMS have good tracking ability?
- How does one calculate and minimize the tracking errors?

Now, for a given sequence $\{\phi_k\}$, (1) is a linear, time-varying difference equation. The properties of this equation are essentially determined by the homogeneous equation

$$x_{k+1} = (I - \mu\phi_k\phi_k^T)x_k \quad (2)$$

with fundamental matrix

$$\beta_\mu(t, k) = \prod_{j=k}^t (I - \mu\phi_j\phi_j^T). \quad (3)$$

The expression for tracking errors will then be of the form

$$\sum_{k=1}^t \beta_\mu(t, k)v(k) \quad (4)$$

where $\{v(k)\}$ describes the error sources (measurement noise, parameter variations, etc.). As elaborated, e.g., in [8] and [6], the essential key to the analysis of (4) is to prove exponential stability of (3). This was also the motivation behind the work of [1]. We shall establish such exponential stability in the sense that for any $p \geq 1$ there exist positive constants M , α , and μ^* such that

$$[E\|\beta_\mu(t, k)\|^p]^{1/p} \leq M(1 - \mu\alpha)^{t-k}, \quad \forall t \geq k, \quad \forall \mu \in (0, \mu^*). \quad (5)$$

The expectation E , here, is with respect to the sequence $\{\phi_k\}$.

Clearly, (5) is a property of the sequence $\{\phi_k\}$ only. We shall here establish (5) under very general conditions on $\{\phi_k\}$. These are of the kind (precise conditions are given in Theorem 2):

- *restrictions on the dependence among the ϕ_k* : This takes the form that ϕ_k is formed by possibly time-varying, but uniformly stable, filtering of a noise source ε_j which is mixing and obeys an additional condition on the rate of decay of dependence;
- *restrictions on the tail of the distribution of ϕ_k* : This takes the form that

$$E[\exp(\alpha\|\varepsilon_k\|^2)] < C, \quad \forall k \quad (6)$$

for some $\alpha > 0$ and some constant C . Here ε_k is the “source” from which ϕ_k was formed.

Both of these restrictions are very mild and allow, for example, the Gaussian, dependent case (unlike most previous treatments). Now, for sequences ϕ_k subject to these two restrictions, the *necessary and sufficient* condition for (5) to hold is that

$$\sum_{i=k+1}^{k+h} E[\phi_i \phi_i^\tau] \geq \delta I, \quad \forall k \geq 0 \quad (7)$$

for some $h > 0$ and $\delta > 0$. This is the “persistence of excitation” or “full rank” condition on ϕ_k .

This result is the main contribution of this paper. Furthermore, several direct applications of the stability result to adaptive tracking will be given under various noise assumptions, which in particular, yield more general results on LMS than those established recently in [8].

B. Earlier Work

Most of the existing work related to exponential stability of (2) is concerned with the case where the signals $\{\phi_k\}$ are independent or M -dependent (cf., e.g., [20], [19], [4], [1], and [2]). This independence assumption can be relaxed considerably if we assume that the signals $\{\phi_k\}$ are bounded as in, e.g., [6], [18], and [12]. Note that the boundedness assumption is suitable for the study of the so-called normalized LMS algorithms (cf. [19], [6], and [15]), since the normalized signals are automatically bounded. In this case, some general results together with a very weak (probably the weakest ever known) excitation condition for guaranteeing the exponential stability of LMS can be found in [6]. Moreover, in the bounded ϕ -mixing case, a complete characterization of the exponential stability can also be given. Indeed, in that case it has been shown in [6] that (7) is the necessary and sufficient condition for (2) to be exponentially stable.

For general unbounded and correlated random signals, the stability analysis for the standard LMS algorithm (1) becomes more complex as to have defied complete solution for over 30 years. Recently, some general stability results applicable to unbounded nonstationary-dependent signals are established in [7], based on which a number of results on the tracking performance of the LMS algorithms can be derived (see [8]). In particular, the result of [7] can be applied to a typical situation, where the signal process is generated from a white noise sequence through a stable linear filter

$$\phi_k = \sum_{j=-\infty}^{\infty} A_j \varepsilon_{k-j} + \xi_k, \quad \sum_{j=-\infty}^{\infty} \|A_j\| < \infty \quad (8)$$

where $\{\varepsilon_k\}$ is an independent sequence satisfying

$$\sup_k E[\exp(\alpha \|\varepsilon_k\|^\delta)] < \infty, \quad \text{for some } \alpha > 0, \delta > 2 \quad (9)$$

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

It is obvious that (8) has a similar form as the well-known Wold decomposition for wide-sense stationary processes. Note,

however, that the signal process $\{\phi_k\}$ defined by (8) need not be a stationary process nor a Markov chain, in general.

Unfortunately, (9) with $\delta > 2$ excludes the case where $\{\varepsilon_k\}$ is a Gaussian process, since such signals could only satisfy a weaker condition

$$\sup_k E[\exp(\alpha \|\varepsilon_k\|^2)] < \infty, \quad \text{for some } \alpha > 0. \quad (10)$$

The motivation of this paper has thus been to relax (9) so that at least the signal process $\{\phi_k\}$ defined by (8) and (10) can be included. This will be done in a more general setting, based on a relaxation of the moment condition used in [7, Th. 3.2].

II. THE MAIN RESULTS

A. Notations

Here we adopt the following notations introduced in [7].

- 1) 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.,

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

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

$$\|X\|_p \triangleq \{E(\|X\|^p)\}^{\frac{1}{p}}, \quad p \geq 1.$$

- 2) For any square random matrix sequence $F = \{F_k\}$, and real numbers $p \geq 1, \mu^* \in (0, 1)$, the L_p -exponentially stable family $\mathcal{S}_p(\mu^*)$ is defined by

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

Likewise, the averaged exponentially stable family $\mathcal{S}(\mu^*)$ is defined by

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

In what follows, it will be convenient to set

$$\mathcal{S}_p \triangleq \bigcup_{\mu^* \in (0, 1)} \mathcal{S}_p(\mu^*), \quad \mathcal{S} \triangleq \bigcup_{\mu^* \in (0, 1)} \mathcal{S}(\mu^*). \quad (11)$$

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

$$\mathcal{M}_p = \left\{ F : \sup_i \|S_i^{(T)}\|_p = o(T), \text{ as } T \rightarrow \infty \right\} \quad (12)$$

where

$$S_i^{(T)} = \sum_{j=iT}^{(i+1)T-1} (F_j - E[F_j]). \quad (13)$$

The definition of \mathcal{M}_p is reminiscent of the law of large numbers. As shown by [9, Lemma 3], it includes a large class of random processes.

B. The Main Results

We first present a preliminary theorem.

Theorem 1: Let $\{F_k\}$ be a random matrix process. Then

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

provided that the following *two conditions* are satisfied.

1) There exist positive constants ε, M , and K such that for any $n \geq 1$

$$E \left[\exp \left(\varepsilon \sum_{i=1}^n \|F_{j_i}\| \right) \right] \leq M \exp(Kn)$$

holds for any integer sequence $0 \leq j_1 < j_2 < \dots < j_n$.

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

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

where $S_j^{(T)}$ is defined by (13).

The proof is given in Section IV.

Remark 1: The form of Theorem 1 is similar to that of [7, Th. 3.2]. The key difference lies in Condition 1). This condition was introduced in [5, p. 112] and is, in a certain sense, a relaxation of the corresponding condition used in [7, Th. 3.2]. Such a relaxation enables us to include Gaussian signals as a special case when the LMS algorithms are in consideration, as will be shown shortly.

Based on Theorem 1, we may prove that for a large class of unbounded nonstationary signals including (8), (7) is also necessary and sufficient for the exponential stability of LMS.

Let us start with the following decomposition which is more general than that in (8):

$$\phi_k = \sum_{j=-\infty}^{\infty} A(k, j) \varepsilon_{k-j} + \xi_k, \quad \sum_{j=-\infty}^{\infty} \sup_k \|A(k, j)\| < \infty \quad (14)$$

where $\{\xi_k\}$ is a d -dimensional bounded deterministic process, and $\{\varepsilon_k\}$ is now a general m -dimensional ϕ -mixing sequence. The weighting matrices $A(k, j) \in \mathcal{R}^{d \times m}$ are assumed to be deterministic.

We remark that the summability condition in (14) is precisely the standard definition for uniform stability of time-varying linear filters (cf., e.g., [13]). Also, recall that a random sequence $\{\varepsilon_k\}$ is called ϕ -mixing if there exists a nonincreasing function $\phi(m)$ (called the mixing rate) with $\phi(m) \in [0, 1]$, $\forall m \geq 0$, and $\phi(m) \rightarrow 0$ as $m \rightarrow \infty$ such that

$$\sup_{A \in \mathcal{F}_{-\infty}^-, B \in \mathcal{F}_{k+m}^{\infty}} |P(B | A) - P(B)| \leq \phi(m), \quad \forall m \geq 0, k \in (-\infty, \infty)$$

where by definition \mathcal{F}_i^j , $-\infty \leq i \leq j \leq \infty$ is the σ -algebra generated by $\{\varepsilon_k, i \leq k \leq j\}$.

The ϕ -mixing concept is a standard one in the literature for describing weakly dependent random processes. As is well known, the ϕ -mixing property is satisfied by, for example, any M -dependent sequences, sequences generated from bounded white noises via a stable linear filter, and stationary aperiodic Markov chains which are Markov ergodic and satisfy Doeblin's condition (cf. [3]).

The main result of this paper is then stated as follows.

Theorem 2: Consider (2). Let the signal process $\{\phi_k\}$ be generated by (14) where $\{\xi_k\}$ is a bounded deterministic sequence, and $\{\varepsilon_k\}$ is a ϕ -mixing process which satisfies for any $n \geq 1$ and any integer sequence $j_1 < j_2 < \dots < j_n$

$$E \left[\exp \left(\alpha \sum_{i=1}^n \|\varepsilon_{j_i}\|^2 \right) \right] \leq M \exp(Kn) \quad (15)$$

where α, M , and K are positive constants. Then for any $p \geq 1$, there exist constants $\mu^* > 0$, $M > 0$, and $\alpha \in (0, 1)$ such that for all $\mu \in (0, \mu^*)$

$$\left[E \left\| \prod_{j=k+1}^t (I - \mu \phi_j \phi_j^T) \right\|^p \right]^{1/p} \leq M(1 - \mu\alpha)^{t-k}, \quad \forall t \geq k \geq 0 \quad (16)$$

if and only if there exists an integer $h > 0$ and a constant $\delta > 0$ such that

$$\sum_{i=k+1}^{k+h} E[\phi_i \phi_i^T] \geq \delta I, \quad \forall k \geq 0. \quad (17)$$

The proof is also given in Section IV.

Remark 2: By taking $A(k, 0) = I, A(k, j) = 0, \forall k, \forall j \neq 0$, and $\xi_k = 0, \forall k$ in (14), we see that $\{\phi_k\}$ coincides with $\{\varepsilon_k\}$, which means that Theorem 2 is applicable to any ϕ -mixing sequences. Furthermore, if $\{\varepsilon_k\}$ is bounded, then (15) is automatically satisfied. This shows that Theorem 2 may include the corresponding result in [6] as a special case.

Note, however, that a linearly filtered ϕ -mixing process like (14) will no longer be a ϕ -mixing sequence in general (because of the possible unboundedness of $\{\varepsilon_k\}$). In fact, Theorem 2 is applicable also to quite a large class of processes other than ϕ -mixing, as shown by the following corollary.

Corollary 1: Let the signal process $\{\phi_k\}$ be generated by (14), where $\{\xi_k\}$ is a bounded deterministic sequence, and $\{\varepsilon_k\}$ is an independent sequence satisfying (10). Then $\{\phi_k \phi_k^T\} \in \mathcal{S}_p$ for all $p \geq 1$ if and only if there exists an integer $h > 0$ and a constant $\delta > 0$ such that (17) holds.

Proof: By Theorem 2, we only need to show that (15) is true. This is obvious since $\{\varepsilon_k\}$ is an independent sequence satisfying (10). \square

Remark 3: Corollary 1 continues to hold if the independence assumption of $\{\varepsilon_k\}$ is weakened to M -dependence. Moreover, (10) used in Corollary 1 may also be further relaxed if additional conditions are imposed. This is the case when, for example, $\{\phi_k\}$ is a stationary process generated by a stable finite-dimensional linear state-space model with the innovation process $\{\varepsilon_k\}$ being an i.i.d. sequence (see [16]).

III. PERFORMANCE OF ADAPTIVE TRACKING

Let us now assume that $\{y_k\}$ and $\{\phi_k\}$ are related by a linear regression

$$y_k = \phi_k^T x_k^* + v_k \quad (18)$$

where $\{x_k^*\}$ is the true or "fictitious" time-varying parameter process, and $\{v_k\}$ represents the disturbance or unmodeled dynamics.

The objective of the LMS algorithm (1) is then to track the time-varying unknown parameter process $\{x_k^*\}$. The tracking error will depend on the parameter variation process $\{\Delta_k\}$ defined by

$$\Delta_k = x_k^* - x_{k-1}^* \quad (19)$$

through the following error equation obtained by substituting (18) and (19) into (1):

$$\tilde{x}_{k+1} = (I - \mu \phi_k \phi_k^T) \tilde{x}_k + \mu \phi_k v_k - \Delta_{k+1} \quad (20)$$

where $\tilde{x}_k \triangleq x_k - x_k^*$.

Obviously, the quality of tracking will essentially depend on the properties of $\{\phi_k, \Delta_k, v_k\}$. The homogeneous part of (20) is exactly (2) and can be dealt with by Theorem 2. Hence, we need only to consider the nonhomogeneous terms in (20). Different assumptions on $\{\Delta_k, v_k\}$ will give different tracking error bounds or expressions, and we shall treat three cases separately in the following.

A. First Performance Analysis

By this, we mean that the tracking performance analysis is carried out under a "worst case" situation, i.e., the parameter variations and the disturbances are only assumed to be bounded in an averaging sense. To be specific, let us make the following assumption.

A1) There exists $r > 2$ such that

$$\sigma \triangleq \sup_k \|v_k\|_r < \infty$$

and

$$\gamma \triangleq \sup_k \|\Delta_k\|_r < \infty.$$

Note that this condition includes any "unknown but bounded" deterministic disturbances and parameter variations as a special case.

Theorem 3: Consider the LMS algorithm (1) applied to (18). Let Condition A1) be satisfied. Also, let $\{\phi_k\}$ be as in Theorem 2 with (17) satisfied. Then for all $t \geq 1$ and all small $\mu > 0$

$$E\|x_t - x_t^*\|^2 = O\left(\sigma^2 + \frac{\gamma^2}{\mu^2}\right) + O([1 - \mu\beta]^t)$$

where $\beta \in (0, 1)$ is a constant.

This result follows immediately from Theorem 2, (20), and the Hölder inequality. We remark that various such "worst case" results for other commonly used algorithms (e.g., recursive least squares and Kalman filter) may be found in [6]. The main implication of Theorem 3 is that the tracking error will be small if both the parameter variation (γ) and the disturbance (σ) are small.

B. Second Performance Analysis

By this, we mean that the tracking performance analysis is carried out for zero mean random parameter variations and disturbances which may be correlated processes in general. To be specific, we introduce the following set for $r \geq 1$:

$$\mathcal{N}_r = \left\{ w : \sup_k \left\| \sum_{i=k+1}^{k+n} w_i \right\|_r \leq c_r^w \sqrt{n}, \quad \forall n \geq 1 \right\} \quad (21)$$

where c_r^w is a constant depending on r and the distribution of $\{w_i\}$ only.

Obviously, \mathcal{N}_r is a subset of \mathcal{M}_r defined by (12). It is known (see [9]) that the martingale difference, zero mean ϕ - and α -mixing sequences can all be included in \mathcal{N}_r . Also, from the proof of [9, Lemma 3], it is known that the constant c_r^w can be dominated by $\sup_k \|w_k\|_r$ in the first two cases and by $\sup_k \|w_k\|_{r+\delta}$ ($\delta > 0$) in the last case.

Moreover, it is interesting to note that \mathcal{N}_r is invariant under linear transformations. This means that if $\{\phi_k\}$ and $\{\varepsilon_k\}$ are related by (8) with $\xi_k \equiv 0$, then $\{\varepsilon_k\} \in \mathcal{N}_r$ implies that $\{\phi_k\} \in \mathcal{N}_r$. This can be easily seen from the following inequality:

$$\begin{aligned} \left\| \sum_{i=k+1}^{k+n} \phi_i \right\|_r &= \left\| \sum_{j=-\infty}^{\infty} A_j \sum_{i=k+1}^{k+n} \varepsilon_{i-j} \right\|_r \\ &\leq \sum_{j=-\infty}^{\infty} \|A_j\| \cdot \left\| \sum_{i=k+1}^{k+n} \varepsilon_{i-j} \right\|_r. \end{aligned}$$

Thus, random processes generated from martingale differences, or ϕ - or α -mixing sequences via an infinite order linear filter, can all be included in \mathcal{N}_r .

Now, we are in a position to introduce the following condition for the second performance analysis.

A2) For some $r > 2$, $\{\Delta_k\} \in \mathcal{N}_r$ and $\{\phi_k v_k\} \in \mathcal{N}_r$.

Theorem 4: Consider the LMS algorithm (1) applied to (18). Let $\{\phi_k\}$ be defined as in Theorem 2 with (17) satisfied, and let the Condition A2) hold for a certain r . Then for all $t \geq 1$ and all small $\mu > 0$

$$E\|x_t - x_t^*\|^2 = O\left(\mu(c_r^{\phi v})^2 + \frac{(c_r^{\Delta})^2}{\mu}\right) + O([1 - \mu\beta]^t)$$

where $c_r^{\phi v}$ and c_r^{Δ} are the constants defined in (21), which depend on the distributions of $\{\phi_k v_k\}$ and $\{\Delta_k\}$, respectively. Moreover, β is the same constant as in Theorem 3.

Proof: By [8, Lemma A.2] and Theorem 2, it is easy to see from (20) that the desired result is true. \square

Note that the upper bound in Theorem 4 significantly improves the ‘‘crude’’ bound given in Theorem 3 for small μ , and it roughly indicates the familiar tradeoff between noise sensitivity and tracking ability.

Theorem 4 can be applied directly to the convergence analysis of some standard filtering problems (cf. [20], [4], and [2]). For example, let $\{y_k\}$ and $\{\phi_k\}$ be two stationary processes, and assume that our purpose is to track the LMS solution

$$x^* = [E(\phi_k \phi_k^T)]^{-1} E(\phi_k y_k)$$

of

$$\min_x E(y_k - x^T \phi_k)^2$$

recursively, based on real-time measurements $\{y_i, \phi_i, i \leq k\}$.

Now, define $\{v_k\}$ by

$$y_k = \phi_k^T x^* + v_k.$$

It is then obvious that $E\phi_k v_k = 0$. Furthermore, in many standard situations it can be verified that $\{\phi_k v_k\} \in \mathcal{N}_r$ for some $r > 2$. Thus, Theorem 4 applied to the above linear regression gives

$$E\|x_t - x^*\|^2 = O(\mu) + O([1 - \mu\beta]^t)$$

which tends to zero as $t \rightarrow \infty$ and $\mu \rightarrow 0$.

Apparently, Theorem 4 is also applicable to nonstationary signals $\{y_k\}$ and $\{\phi_k\}$.

C. Third Performance Analysis

By this, we mean that the analysis is purposed to get an explicit (approximate) expression for the tracking performance rather than just getting an upper bound as in the previous two cases. This is usually carried out under white noise assumptions on $\{\Delta_k, v_k\}$. Roughly speaking, the parameter process in this case will behave like a random walk, and some detailed interpretations of this parameter model may be found in [14] and [8]. We make the following assumptions.

A3) The regressor process is generated by a time-varying causal filter

$$\phi_k = \sum_{j=0}^{\infty} A(k, j) \varepsilon_{k-j} + \xi_k, \quad \sum_{j=0}^{\infty} \sup_k \|A(k, j)\| < \infty \quad (22)$$

where $\{\xi_k\}$ is a bounded deterministic sequence, and $\{\varepsilon_k, \Delta_k, v_{k-1}\}$ is a ϕ -mixing process with mixing rate denoted by $\phi(m)$. Assume also that (15) and (17) hold.

A4) The process $\{\Delta_k, v_k\}$ satisfies the following conditions:

- a) $E[v_k | \mathcal{F}_k] = 0, \quad E[\Delta_{k+1} | \mathcal{F}_k] = E[\Delta_{k+1} v_k | \mathcal{F}_k] = 0;$
- b) $E[v_k^2 | \mathcal{F}_k] = R_v(k), \quad E[\Delta_k \Delta_k^T] = Q(k);$
- c) $\sup_k E[|v_k|^r | \mathcal{F}_k] \leq M, \quad \gamma \triangleq \sup_k \|\Delta_k\|_r < \infty;$

where $r > 2$ and $M > 0$ are constants, and \mathcal{F}_k denotes the σ -algebra generated by $\{\varepsilon_i, \Delta_i, v_{i-1}, i \leq k\}$.

Theorem 5: Consider the LMS algorithm (1) applied to (18). Let Conditions A3) and A4) be satisfied. Then, the tracking error covariance matrix has the following expansion for all $t \geq 1$ and all small $\mu > 0$:

$$E[\tilde{x}_t \tilde{x}_t^T] = \Pi_t + O\left(\sigma(\mu) \left[\mu + \frac{\gamma^2}{\mu} + (1 - \beta\mu)^t\right]\right)$$

where the function $\sigma(\mu) \rightarrow 0$ as $\mu \rightarrow 0$, and Π_t is recursively defined by

$$\Pi_{t+1} = (I - \mu S_t) \Pi_t (I - \mu S_t)^T + \mu^2 R_v(t) S_t + Q(t+1)$$

with $S_t = E[\phi_t \phi_t^T]$ and $R_v(t)$ and $Q(t)$ being defined as in Condition A4).

This theorem relaxes and unifies the conditions used in [8, Th. 5.1]. The proof is given in Section IV. The expression for the function $\sigma(\mu)$ may be found from the proof and from the related formula in [8, Th. 4.1]; see (45).

Note that in the (wide-sense) stationary case, $S_t \equiv S$, $R_v(t) \equiv R_v$, $Q(t) \equiv Q$, and Π_t will converge to a matrix Π defined by the Lyapunov equation (cf. [8])

$$S\Pi + \Pi S = \mu R_v S + \frac{Q}{\mu}.$$

In this case, the trace of the matrix Π , which represents the dominating part of the tracking error $E\|\tilde{x}_t\|^2$ for small μ and large t , can be expressed as

$$\text{tr}(\Pi) = \frac{1}{2} \left[\mu R_v d + \frac{\text{tr}(S^{-1}Q)}{\mu} \right]$$

where $d \triangleq \dim(\phi_k)$. Minimizing $\text{tr}(\Pi)$ with respect to μ , one obtain the following formula for the step-size μ :

$$\mu = \sqrt{\frac{\text{tr}(S^{-1}Q)}{R_v d}}.$$

IV. PROOF OF THEOREMS 1, 2, AND 5

A. Proof of Theorem 1

By the proof of [7, Lemma 5.2] we know that Theorem 1 will be true if [7, eq. (32)] can be established. However, by [7, eq. (34)] and Condition 2), it is easy to see that we only need to show that for any fixed $c \geq 1$, $t \geq 1$, and $T > 1$, and for all small $\mu > 0$

$$\left\| \prod_{j=i+1}^n (1 + \mu^2 c \|H_j\|) \right\|_t \leq M (1 + O(\mu^{\frac{3}{2}}))^{n-i}, \quad \forall n > i \quad (23)$$

where $M > 0$ is a constant and

$$\mu^2 H_j = \mu^2 H_j(2) + \mu^3 H_j(3) + \dots + \mu^T H_j(T) + O(\mu^2)$$

with

$$H_j(k) = \sum_{jT \leq j_1 < j_2 < \dots < j_k \leq (j+1)T-1} F_{j_k} \dots F_{j_1}, \quad k = 2, \dots, T,$$

Now, let us set

$$f_j = \exp \left\{ \mu^{\frac{1}{4}} \sum_{s=jT}^{(j+1)T-1} \|F_s\| \right\}.$$

Then for any $2 \leq k \leq T$ and $jT \leq j_1 < \dots < j_k \leq (j+1)T - 1$, by using the inequalities $k \geq \frac{3}{2} + \frac{k}{4}$ and $x \leq \exp(x)$, we have for $\mu \in (0, 1)$

$$\begin{aligned} \mu^k \|F_{j_k} \cdots F_{j_1}\| &\leq \mu^{\frac{3}{2}} (\mu^{\frac{1}{4}} \|F_{j_k}\|) \cdots (\mu^{\frac{1}{4}} \|F_{j_1}\|) \\ &\leq \mu^{\frac{3}{2}} \exp \{ \mu^{\frac{1}{4}} (\|F_{j_1}\| + \cdots + \|F_{j_k}\|) \} \\ &\leq \mu^{\frac{3}{2}} f_j. \end{aligned}$$

Consequently

$$\begin{aligned} (1 + \mu^2 c \|H_j\|) &\leq \prod_{k=2}^T (1 + \mu^k c \|H_j(k)\|) (1 + O(\mu^2)) \\ &\leq \prod_{k=2}^T \prod_{iT \leq j_1 < j_2 < \dots < j_k \leq (i+1)T-1} (1 + \mu^k c \|F_{j_k} \cdots F_{j_1}\|) \\ &\quad \times (1 + O(\mu^2)) \\ &\leq (1 + \mu^{\frac{3}{2}} c f_j)^{2^T} (1 + O(\mu^2)). \end{aligned} \quad (24)$$

Note that

$$\prod_{j=i+1}^n (1 + \mu^{\frac{3}{2}} c f_j) = \sum_{k=0}^{n-i} (\mu^{\frac{3}{2}} c)^k \sum_{i+1 \leq j_1 < \dots < j_k \leq n} f_{j_1} \cdots f_{j_k}.$$

Now, applying the Minkowski inequality to the above identity, noting the disjoint property of the sets $\{j_i T < j < (j_i + 1)T - 1\}$, $i = 1, 2, \dots$, for $j_1 < j_2 < \dots$, taking μ small enough so that $2^T t \mu^{\frac{1}{4}} \leq \varepsilon$, and using Condition 1), it is evident that

$$\begin{aligned} \left\| \prod_{j=i+1}^n (1 + \mu^{\frac{3}{2}} c f_j) \right\|_{2^{Tt}} &\leq \sum_{k=0}^{n-i} (\mu^{\frac{3}{2}} c)^k \sum_{i+1 \leq j_1 < \dots < j_k \leq n} M^{\frac{1}{2^T t}} \exp \left\{ \left(\frac{KT}{2^T t} \right) k \right\} \\ &\leq M^{\frac{1}{2^T t}} \left(1 + c \mu^{\frac{3}{2}} \exp \left(\frac{KT}{2^T t} \right) \right)^{n-i}. \end{aligned}$$

Finally, from this and (24), we have for any $n > i$

$$\begin{aligned} \left\| \prod_{j=i+1}^n (1 + \mu^2 c \|H_j\|) \right\|_t &\leq \left\| \prod_{j=i+1}^n (1 + \mu^{\frac{3}{2}} c f_j) \right\|_{2^{Tt}} [1 + O(\mu^2)]^{n-i} \\ &\leq M \left\{ \left[1 + c \mu^{\frac{3}{2}} \exp \left(\frac{KT}{2^T t} \right) \right]^{2^T} \right\}^{n-i} [1 + O(\mu^2)]^{n-i} \\ &\leq M [1 + O(\mu^{\frac{3}{2}})]^{n-i}, \quad \text{for all small } \mu > 0 \end{aligned}$$

which is (23). This completes the proof of Theorem 1.

The proof of Theorem 2 is rather involved, and so it is divided (prefaced) by several lemmas.

For the analysis to follow, it is convenient to rewrite (14) as

$$\phi_k = \sum_{j=-\infty}^{\infty} a_j \varepsilon(k, j) + \xi_k, \quad \sum_{j=-\infty}^{\infty} a_j < \infty \quad (25)$$

where by definition

$$a_j \triangleq \sup_k \|A(k, j)\|, \quad \varepsilon(k, j) \triangleq a_j^{-1} A(k, j) \varepsilon_{k-j}. \quad (26)$$

(We set $\varepsilon(k, j) = 0$, $\forall k$, if $a_j = 0$ for some j .)

The new process $\{\varepsilon(k, j)\}$ has the following simple properties.

- 1) For any k and j , $\|\varepsilon(k, j)\| \leq \|\varepsilon_{k-j}\|$.
- 2) For any fixed j , the process $\{\varepsilon(k, j)\}$ is ϕ -mixing with the same mixing rate as $\{\varepsilon_k\}$.
- 3) For any k and j , $\varepsilon(k, j)$ is $\sigma\{\varepsilon_{k-j}\}$ -measurable.

These three properties will be frequently used in the sequel without further explanations.

Lemma 1: Let $\{F_t\}$ be a ϕ -mixing $d \times d$ -dimensional matrix process with mixing rate $\{\phi(m)\}$. Then

$$\sup_i \|S_i^{(T)}\|_2 \leq 2cd \left\{ T \sum_{m=0}^{T-1} \sqrt{\phi(m)} \right\}^{\frac{1}{2}}, \quad \forall T \geq 1$$

where $S_i^{(T)}$ is defined by (13) and c is defined by $c \triangleq \sup_i \|F_i - EF_i\|_2$.

Proof: Denote $G_k = F_k - EF_k$. Then by [10, Th. A.6, p. 278], we have

$$\|E[G_j G_k^T]\| \leq 2cd^2 \sqrt{\phi(|j-k|)}, \quad \forall j, k.$$

Consequently, by using the inequality

$$|\text{tr } F| \leq d \|F\|, \quad \forall F \in \mathcal{R}^{d \times d}.$$

We get

$$\begin{aligned} \|S_i^{(T)}\|_2^2 &= E \left\| \sum_{j,k=iT}^{(i+1)T-1} G_j G_k^T \right\| \\ &\leq \text{tr} \left\{ \sum_{j,k=iT}^{(i+1)T-1} EG_j G_k^T \right\} \\ &\leq d \sum_{j,k=iT}^{(i+1)T-1} \|EG_j G_k^T\| \\ &\leq 2c^2 d^2 \sum_{j,k=iT}^{(i+1)T-1} \sqrt{\phi(|j-k|)} \\ &\leq 4c^2 d^2 T \sum_{m=0}^{T-1} \sqrt{\phi(m)}. \end{aligned}$$

This gives the desired result. \square

Lemma 2: Let $F_k = \phi_k \phi_k^T$, where $\{\phi_k\}$ is defined by (14) with $\sup_k \|\varepsilon_k\|_4 < \infty$. Then $\{F_k\} \in \mathcal{M}_2$, where \mathcal{M}_2 is defined by (12).

Proof: First of all, we may assume that the process $\{\varepsilon_k\}$ is of zero mean (otherwise, the mean can be included in ξ_k). Then by (25)

$$\begin{aligned} & \|S_i^{(T)}\|_2 \\ &= \left\| \sum_{t=iT}^{(i+1)T-1} [\phi_t \phi_t^\tau - E\phi_t \phi_t^\tau] \right\|_2 \\ &\leq \sum_{k,j=-\infty}^{\infty} a_k a_j \left\| \sum_{t=iT}^{(i+1)T-1} [\varepsilon(t,k)\varepsilon(t,j)^\tau - E\varepsilon(t,k)\varepsilon(t,j)^\tau] \right\|_2 \\ &+ 2 \sum_{j=-\infty}^{\infty} a_j \left\| \sum_{t=iT}^{(i+1)T-1} \varepsilon(t,j)\xi_t^\tau \right\|_2. \end{aligned} \tag{27}$$

Note that for any fixed k and j , both the processes $\{\varepsilon(t,k)\varepsilon(t,j)^\tau\}$ and $\{\varepsilon(t,j)\}$ are ϕ -mixing with mixing rate $\phi(m - |k - j|)$ and $\phi(m)$, respectively (where by definition, $\phi(m) \triangleq 1, \forall m < 0$).

By Lemma 1, it is easy to see that the last term in (27) is of order $o(T)$. For dealing with the second last term, we denote

$$f_{kj}(T) = 2cd \left\{ T \sum_{m=0}^{T-1} \sqrt{\phi(m - |k - j|)} \right\}^{\frac{1}{2}} \tag{28}$$

where c is defined as in Lemma 1. Consequently, by $\phi(m) \leq 1, \forall m$, it is not difficult to see that

$$\sup_{k,j} f_{kj}(T) \leq 2cdT \tag{29}$$

and

$$\sup_{|k-j| < \sqrt{T}} f_{kj}(T) = o(T). \tag{30}$$

Now, by the summability of $\{a_j\}$

$$\sum_{|k-j| \geq \sqrt{T}} a_k a_j \rightarrow 0, \quad \text{as } T \rightarrow \infty.$$

Hence by (29)

$$\sum_{|k-j| \geq \sqrt{T}} a_k a_j f_{kj}(T) = o(T) \tag{31}$$

and by (30)

$$\sum_{|k-j| < \sqrt{T}} a_k a_j f_{kj}(T) = o(T). \tag{32}$$

Combining (31) and (32) gives

$$\sum_{k,j=-\infty}^{\infty} a_k a_j f_{kj}(T) = o(T). \tag{33}$$

By this and Lemma 1, we know that the second to the last term in (27) is also of the order $o(T)$, uniformly in i . Hence, $\{F_k\} \in \mathcal{M}_2$ by (12). \square

Lemma 3: Let $\sup_k E\|\phi_k\|_2 < \infty$. Then $\{\phi_k \phi_k^\tau\} \in \mathcal{S}$ if and only if (17) holds, where \mathcal{S} is defined in (11).

Proof: Let us first assume that (17) is true. Take $\mu^* = (1 + \sup_k E\|\phi_k\|_2^2)^{-1}$. Then applying [6, Th. 2.1] to the deterministic sequence $A_k = \mu E[\phi_k \phi_k^\tau]$ for any $\mu \in (0, \mu^*)$, it is easy to see that $\{\phi_k \phi_k^\tau\} \in \mathcal{S}(\mu^*)$.

Conversely, if $\{\phi_k \phi_k^\tau\} \in \mathcal{S}$, then there exists $\mu^* \in (0, (1 + \sup_k E\|\phi_k\|_2^2)^{-1})$ such that $\{\phi_k \phi_k^\tau\} \in \mathcal{S}(\mu^*)$. Now, applying [6, Th. 2.2] to the deterministic sequence $A_k = \mu^* E[\phi_k \phi_k^\tau]$, it is easy to see that (17) holds. This completes the proof. \square

Lemma 4: Let $F_k = \phi_k \phi_k^\tau$, where $\{\phi_k\}$ is defined by (14) with (15) satisfied. Then $\{F_k\}$ satisfies Condition 1) of Theorem 1.

Proof: Without loss of generality, assume that $\xi_k \equiv 0$. Let us denote

$$A \triangleq \sum_{j=-\infty}^{\infty} a_j \tag{34}$$

where $\{a_j\}$ is defined by (26). Then by the Schwarz inequality from (25) we have

$$\|\phi_k\|_2^2 \leq A \sum_{j=-\infty}^{\infty} a_j \|\varepsilon_{k-j}\|_2^2.$$

Consequently, by the Hölder inequality and (15) we have for $\varepsilon \leq \alpha A^{-2}$

$$\begin{aligned} & E \exp \left\{ \varepsilon \sum_{i=1}^n \|F_{j_i}\| \right\} \\ &\leq E \exp \left\{ \varepsilon A \sum_{j=-\infty}^{\infty} a_j \sum_{i=1}^n \|\varepsilon_{j_i-j}\| \right\} \\ &= E \prod_{j=-\infty}^{\infty} \exp \left\{ \varepsilon A a_j \sum_{i=1}^n \|\varepsilon_{j_i-j}\| \right\} \\ &\leq \prod_{j=-\infty}^{\infty} \left(E \exp \left\{ \varepsilon A^2 \sum_{i=1}^n \|\varepsilon_{j_i-j}\|^2 \right\} \right)^{\frac{a_j}{A}} \\ &\leq \prod_{j=-\infty}^{\infty} (M \exp\{Kn\})^{\frac{a_j}{A}} \\ &= M \exp\{Kn\}. \end{aligned}$$

This completes the proof. \square

The following lemma was originally proved in [5, p. 113].

Lemma 5: Let $\{z_k\}$ be a nonnegative random sequence such that for some $a > 0, b > 0$, and for all $i_1 < i_2 < \dots < i_n, \forall n \geq 1$

$$E \exp \left\{ \sum_{k=1}^n z_{i_k} \right\} \leq \exp\{an + b\}. \tag{35}$$

Then for any $L > 0$ and any $n \geq i \geq 0$

$$E \exp \left\{ \frac{1}{2} \sum_{j=i+1}^n z_j I(z_j \geq L) \right\} \leq \exp\{e^{a-\frac{L}{2}}(n-i) + b\}$$

where $I(\cdot)$ is the indicator function.

Proof: Denote

$$f_j = \exp\left(\frac{1}{2}z_j\right)I(z_j \geq L).$$

Then by first applying the simple inequality $I(x \geq L) \leq e^{\frac{x}{2}}/e^{\frac{L}{2}}$ and then using (35), we have for any subsequence $j_1 < j_2 \cdots < j_k$

$$\begin{aligned} E[f_{j_1} \cdots f_{j_k}] &= E \exp\left(\frac{1}{2} \sum_{i=1}^k z_{j_i}\right) \prod_{i=1}^k I(z_{j_i} \geq L) \\ &\leq E \exp\left(\sum_{i=1}^k z_{j_i}\right) / \exp\left(\frac{kL}{2}\right) \\ &\leq \exp\left\{\left(a - \frac{L}{2}\right)k + b\right\}. \end{aligned}$$

By this we have

$$\begin{aligned} E \exp\left\{\sum_{j=i+1}^n \frac{1}{2}z_j I(z_j \geq L)\right\} &= E \prod_{j=i+1}^n \exp\left\{\frac{1}{2}z_j I(z_j \geq L)\right\} \\ &\leq E \prod_{j=i+1}^n \left\{1 + \exp\left(\frac{1}{2}z_j\right)I(z_j \geq L)\right\} \\ &= E \prod_{j=i+1}^n \{1 + f_j\} \\ &= E \left\{ \sum_{k=0}^{n-i} \sum_{i+1 \leq j_1 < \cdots < j_k \leq n} f_{j_1} \cdots f_{j_k} \right\} \\ &\leq e^b \left\{ \sum_{k=0}^{n-i} \sum_{i+1 \leq j_1 < \cdots < j_k \leq n} \exp\left\{\left(a - \frac{L}{2}\right)k\right\} \right\} \\ &= e^b \prod_{j=i+1}^n \left\{1 + \exp\left(a - \frac{L}{2}\right)\right\} \\ &\leq \exp\left\{(n-i) \exp\left(a - \frac{L}{2}\right) + b\right\}. \end{aligned}$$

This completes the proof of Lemma 5. \square

Lemma 6: Let $F_k = \phi_k \phi_k^T$, where $\{\phi_k\}$ is defined by (14) with (15) satisfied. Then $\{F_k\}$ satisfies Condition 2) of Theorem 1.

Proof: Set for any fixed k and l

$$z_j \triangleq z_j(k, l) = \left\| \sum_{t=jT}^{(j+1)T-1} [\varepsilon(t, k)\varepsilon(t, l)^T - E\varepsilon(t, k)\varepsilon(t, l)^T] \right\|.$$

Then, similar to (27) from (25) we have

$$\begin{aligned} \sum_{j=i+1}^n \|S_j^{(T)}\| &\leq \sum_{k, l=-\infty}^{\infty} a_k a_l \sum_{j=i+1}^n z_j \\ &+ 2 \sum_{k=-\infty}^{\infty} a_k \sum_{j=i+1}^n \left\| \sum_{t=jT}^{(j+1)T-1} \varepsilon(t, k) \xi_t^T \right\|. \end{aligned} \quad (36)$$

We first consider the second to the last term in (36). By the Hölder inequality

$$\begin{aligned} E \exp\left\{\mu \sum_{k, l=-\infty}^{\infty} a_k a_l \sum_{j=i+1}^n z_j\right\} &= E \prod_{k, l=-\infty}^{\infty} \exp\left\{\mu a_k a_l \sum_{j=i+1}^n z_j\right\} \\ &\leq \prod_{k, l=-\infty}^{\infty} \left\{E \exp\left\{\mu A^2 \sum_{j=i+1}^n z_j\right\}\right\}^{\frac{a_k a_l}{A^2}} \end{aligned} \quad (37)$$

where A is defined by (34).

Now, let $c = \sup_k E\|\varepsilon_k\|^2$, and note that

$$\begin{aligned} \|\varepsilon(t, k)\varepsilon(t, l)^T\| &\leq \frac{1}{2}[\|\varepsilon(t, k)\|^2 + \|\varepsilon(t, l)\|^2] \\ &\leq \frac{1}{2}(\|\varepsilon_{t-k}\|^2 + \|\varepsilon_{t-l}\|^2) \end{aligned}$$

and we have

$$z_j \leq \frac{1}{2} \sum_{t=jT}^{(j+1)T-1} (\|\varepsilon_{t-k}\|^2 + \|\varepsilon_{t-l}\|^2) + cT.$$

By this and (15) it is easy to prove that the sequence $\{\alpha z_j\}$ satisfies (35) with $a = (K + c)T$ and $b = \log M$, where α is defined as in (15). Consequently, by Lemma 5 we have for any $L > 0$

$$\begin{aligned} E \exp\left\{\frac{\alpha}{2} \sum_{j=i+1}^n z_j I(z_j \geq LT)\right\} &\leq M \exp\left\{e^{(K+c-\frac{\alpha L}{2})T}(n-i)\right\}. \end{aligned} \quad (38)$$

Now, in view of (38), taking $\mu < \frac{\alpha A^2}{4}$ and $L > 2\alpha^{-1}(K + c)$, and applying the Hölder inequality, we have

$$\begin{aligned} E \exp\left\{2\mu A^2 \sum_{j=i+1}^n z_j I(z_j \geq LT)\right\} &\leq M \exp\{\mu\delta(T)(n-i)\} \end{aligned} \quad (39)$$

where $\delta(T) \rightarrow 0$ as $T \rightarrow \infty$, which is defined by

$$\delta(T) = 4\alpha^{-1}A^2 \exp\left\{\left(K + c - \frac{\alpha L}{2}\right)T\right\}.$$

Next, we consider the term $x_j \triangleq z_j I(z_j \leq LT)$.

By the inequality $e^x \leq 1 + 2x$, $0 \leq x \leq \log 2$, we have for small $\mu > 0$

$$\exp\left\{2\mu A^2 \sum_{j=i+1}^n x_j\right\} \leq \prod_{j=i+1}^n (1 + 4\mu A^2 x_j). \quad (40)$$

As noted before, for any fixed k and l , the process $\{\varepsilon(t, k)\varepsilon(t, l)^T\}$ is ϕ -mixing with mixing rate $\phi(m - |k - l|)$. Consequently, for any fixed k and l , both $\{z_j\}$ and $\{x_j\}$ are

also ϕ -mixing with mixing rate $\phi((m-1)T+1-|k-l|)$. Note also that by Lemma 1

$$Ex_j \leq Ez_j \leq \|z_j\|_2 \leq f_{kl}(T)$$

where $f_{kl}(T)$ is defined by (28).

Therefore, applying [7, Lemma 6.2, p. 1383], we have

$$\begin{aligned} E \prod_{j=i+1}^n (1 + 4\mu A^2 x_j) &\leq 2\{1 + 8\mu A^2 [f_{kl}(T) + 2LT\phi(T+1-|k-l|)]\}^{n-i} \\ &\leq 2\exp\{8\mu A^2 [f_{kl}(T) + 2LT\phi(T+1-|k-l|)](n-i)\}. \end{aligned} \quad (41)$$

Finally, combining (39)–(41) and using the Schwarz inequality we get

$$\begin{aligned} E \exp \left\{ \mu A^2 \sum_{j=i+1}^n z_j \right\} &\leq \left\{ E \exp \left\{ 2\mu A^2 \sum_{j=i+1}^n z_j I(z_j \geq LT) \right\} \right\}^{\frac{1}{2}} \\ &\quad \times \left\{ E \exp \left\{ 2\mu A^2 \sum_{j=i+1}^n x_j \right\} \right\}^{\frac{1}{2}} \\ &\leq \sqrt{2M} \exp\{\mu[\delta(T) + 8A^2 f_{kl}(T) + 16LTA^2\phi(T+1-|k-l|)](n-i)\}. \end{aligned}$$

Substituting this into (37) and noting (33), it is not difficult to see that there exists a function $g(T) = o(T)$ such that for all small $\mu > 0$

$$\begin{aligned} E \exp \left\{ \mu \sum_{k,l=-\infty}^{\infty} a_k a_l \sum_{j=i+1}^n z_j \right\} &\leq \sqrt{2M} \exp\{\mu g(T)(n-i)\}. \end{aligned}$$

Obviously, for the last term in (36), a similar bound can also be derived using a similar treatment. Hence it is easy to see that the lemma is true. \square

B. Proof of Theorem 2

Necessity: Let $\{\phi_k \phi_k^\tau\} \in \mathcal{S}_p$ for $p = 2$. Then by [7, Lemma 2 and Th. 3.1], we know that $\{\phi_k \phi_k^\tau\} \in \mathcal{S}$. Consequently, by Lemma 3 we know that (17) holds.

Sufficiency: If (17) holds, then by Lemma 3 we have $\{\phi_k \phi_k^\tau\} \in \mathcal{S}$. By this and Lemmas 4 and 6, we know that Theorem 1 is applicable, and consequently $\{\phi_k \phi_k^\tau\} \in \mathcal{S}_p, \forall p \geq 1$. This completes the proof.

C. Proof of Theorem 5

We need to verify all the conditions in [8, Th. 4.1]. However, by Theorem 2, Lemma 3, and the conditions of Theorem 5, it is not difficult to see that we need actually to verify the weak dependence condition in [8], p. 1392. In other words, we need

to show that for any $q \geq 3$, there is a bounded function $\delta(m)$ such that

$$\delta(m) \rightarrow 0, \quad \text{as } m \rightarrow \infty$$

and

$$\|E[\phi_k \phi_k^\tau | \mathcal{F}_{k-m}] - E[\phi_k \phi_k^\tau]\|_q \leq \delta(m), \quad \forall k \geq 0, m \geq 0. \quad (42)$$

First of all, since $\{\varepsilon_k, \Delta_k, v_{k-1}\}$ is ϕ -mixing, we can apply the mixing inequality in [17] to obtain

$$\begin{aligned} \|E[\varepsilon(k, i)\varepsilon(k, j)^\tau | \mathcal{F}_{k-m}] - E[\varepsilon(k, i)\varepsilon(k, j)^\tau]\|_q &\leq C_q [\phi(m - \max(i, j))]^{1-q^{-1}} \end{aligned} \quad (43)$$

for any nonnegative integers i, j, k, m and $q \geq 3$, where $\varepsilon(k, j)$ is defined by (26), C_q depends only on $\sup_{i,j} \|\varepsilon_i \varepsilon_j^\tau\|_q$, and where by definition $\phi(m) \triangleq 1$ for $m < 0$.

Without loss of generality, we may assume that $\{\varepsilon_k\}$ is of zero mean. Then, with some simple manipulations we get from (22) or (25) that

$$\begin{aligned} \sup_k \|E[\phi_k \phi_k^\tau | \mathcal{F}_{k-m}] - E[\phi_k \phi_k^\tau]\|_q &\leq \sum_{i,j=0}^{\infty} a_i a_j \sup_k \|E[\varepsilon(k, i)\varepsilon(k, j)^\tau | \mathcal{F}_{k-m}] - E[\varepsilon(k, i)\varepsilon(k, j)^\tau]\|_q \\ &\quad + 2 \sup_k \|\xi_k\| \sum_{i=0}^{\infty} a_i \|E[\varepsilon(k, i) | \mathcal{F}_{k-m}]\|_q. \end{aligned} \quad (44)$$

Now, by (43), the second last term in (44) can be bounded by

$$C_q \sum_{i,j=0}^{\infty} a_i a_j [\phi(m - \max(i, j))]^{1-q^{-1}}$$

which tends to zero as $m \rightarrow \infty$ by the dominated convergence theorem.

The last term in (44) can be treated similarly. Denote the right-hand side of (44) by $\delta(m)$. It thus tends to zero as $m \rightarrow \infty$. Hence, (42) is true and the proof of Theorem 5 is complete. To find the degree of approximation, define, analogously to [8, Th. 4.1]

$$\sigma(\mu) = \min_{m \geq 1} \{m\sqrt{\mu} + \delta(m)\}. \quad (45)$$

\square

V. CONCLUDING REMARKS

The LMS is a basic algorithm in the estimation of time-varying parameters of dynamical systems as well as in adaptive signal processing. There is an extensive and growing literature devoted to the study of its properties from various aspects, among which the exponential stability is the most fundamental. Despite the remarkably simple structure of LMS, characterizing its properties analytically has long been known very complicated in general. The main contributions of this paper are summarized as follows.

- 1) For a large class of nonstationary weakly dependent signals, (17) is shown to be *necessary and sufficient* for

the exponential stability of LMS, even in the case where the signals are unbounded and non- ϕ -mixing.

- 2) The main stability result—Theorem 2—has quite wide applicability. In particular, it is applicable to a typical situation where the signals are generated from, e.g., Gaussian white noises via a time-varying linear filter of infinite order (see Corollary 1).
- 3) A “three-stage procedure” for the tracking performance analysis is delineated (see Section III), according to different assumptions on parameter variations and disturbances. These assumptions include “worst case noises,” “colored noises,” and “white noises.” By doing so, we have also generalized and simplified the recent related results on LMS in [8]. The basis for this tracking performance analysis, in all its stages, is the exponential stability.

REFERENCES

- [1] R. R. Bitmead and B. D. O. Anderson, “Lyapunov techniques for the exponential stability of linear difference equations with random coefficients,” *IEEE Trans. Automat. Contr.*, vol. AC-25, no. 4, pp. 782–787, 1980.
- [2] P. M. Clarkson, *Optimal and Adaptive Signal Processing*. Boca Raton, FL: CRC, 1993.
- [3] J. L. Doob, *Stochastic Processes*. New York: Wiley, 1953.
- [4] E. Eweda and O. Macchi, “Tracking error bounds of adaptive nonstationary filtering,” *Automatica*, vol. 21, no. 3, pp. 293–302, 1985.
- [5] L. Guo, *Time-Varying Stochastic Systems: Stability, Estimation and Control*. Changchun, China: Jilin Science & Technology, 1993, in Chinese.
- [6] ———, “Stability of recursive stochastic tracking algorithms,” *SIAM J. Contr. Optimization*, vol. 32, pp. 1195–1225, 1994.
- [7] L. Guo and L. Ljung, “Exponential stability of general tracking algorithms,” *IEEE Trans. Automat. Contr.*, vol. 40, pp. 1376–1387, Aug. 1995.
- [8] ———, “Performance analysis of general tracking algorithms,” *IEEE Trans. Automat. Contr.*, vol. 40, pp. 1388–1402, Aug. 1995.
- [9] L. Guo, L. Ljung, and P. Priouret, “Performance analysis of the forgetting factor RLS algorithm,” *Int. J. Adaptive Contr. Signal Processing*, vol. 7, pp. 525–537, 1993.
- [10] P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Applications*. New York: Academic, 1980.
- [11] B. Hassibi, A. H. Sayed, and T. Kailath, “LMS is H^∞ optimal,” in *Proc. IEEE Conf. Decision Contr.*, San Antonio, TX, 1993, vol. 1, pp. 74–79.
- [12] H. J. Kushner and J. Yang, “Analysis of adaptive step-size SA algorithms for parameter tracking,” *IEEE Trans. Automat. Contr.*, vol. 40, no. 8, pp. 1403–1410, 1995.
- [13] L. Ljung, *System Identification—Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [14] L. Ljung and S. Gunnarsson, “Adaptive tracking in system identification—A survey,” *Automatica*, vol. 26, no. 1, pp. 7–22, 1990.
- [15] L. Ljung and P. Priouret, “A result of the mean square error obtained using general tracking algorithms,” *Int. J. Adaptive Contr. Signal Processing*, vol. 5, no. 4, pp. 231–250, 1991.
- [16] P. Priouret and A. Yu. Veretennikov, “A remark on the stability of the LMS tracking algorithm,” *Stochastic Analysis Appl.*, submitted.
- [17] R. J. Serfling, “Contributions to central limit theory for dependent variables,” *Annals Math. Stats.*, vol. 39, no. 4, pp. 1158–1175, 1968.
- [18] V. Solo, “Averaging analysis of the LMS algorithm,” in *Control and Dynamic Systems: Advances in Theory and Applications*, C. T. Leondes, Ed., vol. 65, Part 2. New York: Academic, 1994, pp. 379–397.
- [19] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms: Stability and Performance*. Englewood Cliffs, NJ: Prentice Hall, 1995.
- [20] B. Widrow and S. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.



Lei Guo (M’88-SM’96) was born in Shandong, China in 1961. He received the B.S. degree in mathematics from Shandong University in 1982, and the M.S. and the Ph.D. degrees in control theory from the Chinese Academy of Sciences in 1984 and 1987, respectively.

He was a Post-Doctoral Fellow at the Australian National University during 1987–1989. Since 1992, he has been a Professor with the Institute of Systems Science at the Chinese Academy of Sciences. He is the author or coauthor of over 60 journal papers and

of two books in the area of stochastic and adaptive systems.

The awards won by Dr. Guo include the National Natural Science Prize of China (1987), the Natural Science Prize of the Chinese Academy of Sciences (1994), and the Young Author Prize of the IFAC World Congress (1993). He has served on the editorial board of *SIAM Journal on Control and Optimization*, and is currently on the editorial boards of *Systems Science and Mathematical Sciences*, *Science in China*, *Chinese Science Bulletin*, *Control and Computers*, and *Information and Control*.



Lennart 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 a Associate Editor of several journals.



Guan-Jun Wang was born in Zhejiang, China, in 1967. She received the M.A. degree in probability theory from Hangzhou University in 1992 and the Ph.D. degree in control theory from the Chinese Academy of Sciences in 1996.

She is currently a Lecturer at the Department of Mathematics, Central University for Nationalities, Beijing, China. Her research interests include stochastic stability, estimation, and approximation.