Necessary and Sufficient Conditions for Stability of LMS*

Lei Guo† Lennart Ljung‡ and Guan-Jun Wang§

First Version: November 9, 1995
Revised Version: August 28, 1996

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 tracking performance of LMS are derived under various assumptions. Neither stationarity nor Markov chain assumptions are necessarily required in the paper.

* This work was supported by the National Natural Science Foundation of China and the Swedish Research Council for Engineering Sciences (TFR).
† Institute of Systems Science, Chinese Academy of Sciences, Beijing, 100080, P.R. China. Email: Lguo@iss03.iss.ac.cn.
‡ Department of Electrical Engineering, Linköping University, S-581 83 Linköping, Sweden. Email: Ljung@isy.liu.se.
§ Department of Mathematics, The Central University for Nationalities, Beijing 100081, P. R. China.
1 Introduction

1.1 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

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

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, the books [20], [19] and [2], the survey [14], and the references therein). Also, it has been found recently that the LMS is \( H^\infty \)-optimal in the sense that it minimizes the energy gain from the disturbances to the predicted errors, and it is also 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 to calculate and to 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:

\[
x_{k+1} = (I - \mu \phi_k \phi_k^T) x_k
\]

with fundamental matrix

\[
\beta_\mu(t, k) = \prod_{j=k}^t (I - \mu \phi_j \phi_j^T)
\]
The expression for tracking errors will then be of the form

\[ \sum_{k=1}^{t} \beta_{\mu}(t, k)v(k) \]  

(4)

where \( \{v(k)\} \) describes the error sources (measurement noise, parameter variations etc). As elaborated in, e.g., [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 \( M, \alpha \) and \( \mu \) such that

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

(5)

The expectation \( \mathbb{E} \) here is with respect to the sequence \( \{\phi_{k}\} \).

Clearly, the property (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 \( \phi_{k} \):** This takes the form that \( \phi_{k} \) is formed by possibly time varying, but uniformly stable filtering of a noise source \( \varepsilon_{j} \) which is mixing and obeys an additional condition on the rate of decay of dependence.

- **Restrictions on the tail of the distribution of \( \phi_{k} \).** This takes the form that

\[ \mathbb{E}[\exp(\alpha\|\varepsilon_{k}\|^{2})] < C, \quad \forall k, \]  

(6)

for some \( \alpha > 0 \) and some constant \( C \). Here \( \varepsilon_{k} \) is the “source” from which \( \phi_{k} \) was formed.

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

\[ \sum_{i=k+1}^{k+h} \mathbb{E}[\phi_{i}\phi_{i}^{\top}] \geq \delta I, \quad \forall k \geq 0, \]  

(7)

for some \( h > 0 \) and \( \delta > 0 \). This is the “persistence of excitation” or “full rank” condition on \( \phi_{k} \).

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].

3
1.2 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],[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 [13]), 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 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], and 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:

\[
\phi_k = \sum_{j=-\infty}^{\infty} A_j \varepsilon_{k-j} + \xi_k, \quad \sum_{j=-\infty}^{\infty} \|A_j\| < \infty,
\]

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

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

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

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

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

The motivation of this paper has thus been to relax the moment condition (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 Theorem 3.2 of [7].

2 The Main results

2.1 Notations

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

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.,

\[ \|X\| \triangleq \{\lambda_{\max}(XX^*)\}^{\frac{1}{2}}, \]

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

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

b). 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 \( S_p \) is defined by

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

Likewise, the averaged exponentially stable family \( S \) is defined by

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

In what follows, it will be convenient to set

\[
S_p \triangleq \bigcup_{\mu^* \in (0, 1]} S_p(\mu^*), \quad S \triangleq \bigcup_{\mu^* \in (0, 1)} S(\mu^*), \quad (11)
\]
c) Let $p \geq 1$, $F \triangleq \{F_i\}$. Set

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

(12)

where

$$S^{(T)}_i = \sum_{j=i}^{(i+1)T-1} (F_j - E[F_j])$$

(13)

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

### 2.2 The Main Results

We first present a preliminary theorem.

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

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

provided that the following **two conditions** are satisfied:

(i). There exist positive constants $\varepsilon, M$ 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 $0 \leq j_1 < j_2 \ldots < j_n$.

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

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

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

The proof is given in Section 4.

**Remark 1.** The form of Theorem 1 is similar to that of Theorem 3.2 in [7]. The key difference lies in the condition (i). This condition was introduced in [3], p.112 and is, in a certain sense, a relaxation of the corresponding condition used in Theorem 3.2 of [7]. 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 non-stationary signals including (8), the condition (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):

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

where \(\{\xi_k\}\) is a \(d\)-dimensional bounded deterministic process, and \(\{\xi\}\) is now a general \(m\)-dimensional \(\phi\)-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 \(\{\xi_k\}\) is called \(\phi\)-mixing if there exists a non-increasing function \(\phi(m)\) (called the mixing rate) with \(\phi(m) \in [0, 1], \forall m \geq 0\) and \(\phi(m) \to 0\) as \(m \to \infty\) such that

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

where by definition \(\mathcal{F}_i^j, -\infty \leq i \leq j \leq \infty\), is the \(\sigma\)-algebra generated by \(\{\xi_k, i \leq k \leq j\}\).

The \(\phi\)-mixing concept is a standard one in the literature for describing weakly dependent random processes. As is well-known, the \(\phi\)-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 the random linear equation (2). Let the signal process \(\{\phi_k\}\) be generated by (14) where \(\{\xi_k\}\) is a bounded deterministic sequence, and \(\{\xi\}\) is a \(\phi\)-mixing process which satisfies for any \(n \geq 1\) and any integer sequence \(j_1 < j_2 < \ldots < j_n\)

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

where \(\alpha, M\) 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 \(\mu \in (0, \mu^*)\)

$$
\left[ E \| \prod_{j=k+1}^{t} (I - \mu_{j} \phi_j \phi_j^* ) \|^p \right]^{1/p} \leq M (1 - \mu \alpha)^{t-k}, \quad \forall t \geq k \geq 0 \tag{16}
$$
if and only if there exists an integer $h > 0$ and a constant $\delta > 0$ such that

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

(17)

The proof is also given in Section 4.

**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 $\phi$-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 $\phi$-mixing process like (14) will no longer be a $\phi$-mixing sequence in general (because of the possible unboundedness of $\{\varepsilon_k\}$). In fact, Theorem 2 is applicable also to a quite large class of processes other than $\phi$-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 condition (10). Then $\{\phi_k \phi_k^*\} \in \mathcal{S}_p$ 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 need only to show that condition (15) is true. This is obvious since $\{\varepsilon_k\}$ is an independent sequence satisfying (10). Q.E.D.

**Remark 3.** Corollary 1 continues to hold if the independence assumption of $\{\varepsilon_k\}$ is weakened to M-dependence. Moreover, the moment condition (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]).

### 3 Performance of Adaptive tracking

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

$$
y_k = \phi_k^* x_k^* + v_k
$$

(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

\[
\Delta_k = x^*_k - x^*_{k-1}
\]  

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

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

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

Obviously, the quality of tracking will essentially depend on properties of \( \{\phi_k, \Delta_k, v_k\} \). The homogeneous part of (20) is exactly the equation (2), and can be dealt with by Theorem 2. Hence, we need only to consider the non-homogeneous 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.

3.1 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 \( \mu > 0 \)

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

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., RLS and KF) may be found in [6]. The main implication of Theorem 3 is that the tracking error will be small if both the parameter variation \((\gamma)\) and the disturbance \((\sigma)\) are small.

### 3.2 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 \(r \geq 1\),

\[
\mathcal{N}_r = \left\{ w : \sup_k \left\| \sum_{i=k+1}^{k+n} w_i \right\|_r \leq c^w_r \sqrt{n}, \; \forall n \geq 1 \right\}
\]  

(21)

where \(c^w_r\) 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 martingale difference, zero mean \(\phi-\) and \(\alpha-\) mixing sequences can all be included in \(\mathcal{N}_r\). Also, from the proof of Lemma 3 in [9], it is known that the constant \(c^w_r\) 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:

\[
\left\| \sum_{i=k+1}^{k+n} \phi_i \right\|_r = \left\| \sum_{j=-\infty}^{\infty} A_j \sum_{i=k+1}^{k+n} \varepsilon_{i-j} \right\|_r \\
\leq \sum_{j=-\infty}^{\infty} \left\| A_j \right\| \cdot \left\| \sum_{i=k+1}^{k+n} \varepsilon_{i-j} \right\|_r
\]

Thus, random processes generated from martingale differences, or \(\phi-\) or \(\alpha-\) 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 the model (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 $\mu > 0$,

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

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

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

Note that the upper bound in Theorem 4 significantly improves the “crude” bound given in Theorem 3 for small $\mu$, and it roughly indicates the familiar trade-off 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 least mean squares solution

$$x^* = [E(\phi_k^2)]^{-1}E(\phi_k y_k)$$

of

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

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

Now, define $\{v_k\}$ by

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

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 $t \to \infty$ and $\mu \to 0$.

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

**3.3 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

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

where \( \{\xi_k\} \) is a bounded deterministic sequence, and \( \{\varepsilon_k, \Delta_k, v_{k-1}\} \) is a \( \phi \)-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:

\[
\begin{align*}
(i) \quad E[v_k|\mathcal{F}_k] &= 0, \quad E[\Delta_{k+1}|\mathcal{F}_k] = E[\Delta_{k+1}v_k|\mathcal{F}_k] = 0; \\
(ii) \quad E[v_k^2|\mathcal{F}_k] &= R_v(k), \quad E[\Delta_k \Delta_k'] = Q(k); \\
(iii) \quad \sup_k E[|v_k|^2|\mathcal{F}_k] \leq M, \quad \gamma \overset{\triangle}{=} \sup_k \|\Delta_k\|_r < \infty,
\end{align*}
\]

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

**Theorem 5.** Consider the LMS algorithm (1) applied to the model (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 \( \mu > 0 \)

\[
E[\tilde{x}_t \tilde{x}_t'] = \Pi_t + O \left( \sigma(\mu)|\mu + \frac{\gamma^2}{\mu} + (1 - \beta\mu)^t \right)
\]

where the function \( \sigma(\mu) \to 0 \) as \( \mu \to 0 \), and \( \Pi_t \) is recursively defined by

\[
\Pi_{t+1} = (I - \mu S_t)\Pi_t(I - \mu S_t)^\top + \mu^2 R_v(t)S_t + Q(t + 1)
\]

with \( S_t = E[\phi_t \phi_t'] \) and \( R_v(t) \) and \( Q(t) \) being defined as in condition A4).

This theorem relaxes and unifies the conditions used in Theorem 5.1 of [8]. The proof is given in Section 4. The expression for the function \( \sigma(\mu) \) may be found from the proof, and from the related formula in Theorem 4.1 of [8]. (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 \( \Pi_t \) will converge to a matrix \( \Pi \) 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 $\Pi$, which represents the dominating part of the tracking error $E\|\tilde{x}_d\|^2$ for small $\mu$ and large $t$, can be expressed as

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

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

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

4 Proof of Theorems 1, 2 and 5

Proof of Theorem 1.

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

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

(23)

where $M > 0$ is a constant and

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

with

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

Now, let us set

$$f_j = \exp\left\{\mu^\frac{1}{2} \sum_{s=jT}^{(j+1)T-1} \|F_s\|\right\}$$

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

$$\mu^k \|F_{j_1} \cdots F_{j_k}\|$$

$$\leq \mu^\frac{3}{2}(\mu^\frac{1}{2}\|F_{j_k}\|) \cdots (\mu^\frac{1}{2}\|F_{j_1}\|)$$

$$\leq \mu^\frac{3}{2} \exp\{\mu^\frac{1}{2}(\|F_{j_1}\| + \cdots + \|F_{j_k}\|)\}$$

$$\leq \mu^\frac{3}{2} f_j$$

13
Consequently,

\[
(1 + \mu^2 c \|H_j\|) \\
\leq \prod_{k=2}^{T} (1 + \mu^k c \|H_j(k)\|)(1 + O(\mu^2)) \\
\leq \prod_{k=2}^{T} \prod_{iT \leq j_1 < j_2 \leq (i+1)T-1} (1 + \mu^k c \|F_{j_k} \cdots F_{j_1}\|)(1 + O(\mu^2)) \\
\leq (1 + \mu^\frac{3}{2} c f_j)^{2^T} (1 + O(\mu^2))
\]

(24)

Note that

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

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

\[
\| \prod_{j=i+1}^{n} (1 + \mu^\frac{3}{2} c f_j) \|_{2^T t} \\
\leq \sum_{k=0}^{n-i} (\mu^\frac{3}{2} c)^k \sum_{i+1 \leq j_1 < \ldots < j_k \leq n} M^{\frac{1}{2^T t}} \exp\{\left(\frac{K T}{2^T t}\right)k\} \\
\leq M^{\frac{1}{2^T t}} \left(1 + c\mu^\frac{3}{2} \exp\left(\frac{K T}{2^T t}\right)\right)^{n-i}
\]

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

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

which is (23). This completes the proof of Theorem 1.
The proof of Theorem 2 is rather involved, and so it is divided (prefaced) with several lemmas.

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

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

(25)

where by definition

\[ a_j \triangleq \sup_k \|A(k, j)\|, \quad \varepsilon(k, j) \triangleq a_j^{-1} A(k, j) \varepsilon_{k-j} \]

(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:

(i). For any \( k \) and \( j \), \( \|\varepsilon(k, j)\| \leq \|\varepsilon_{k-j}\| \);

(ii). For any fixed \( j \), the process \( \{\varepsilon(k, j)\} \) is \( \phi \)-mixing with the same mixing rate as \( \{\varepsilon_k\} \);

(iii). 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 \( \phi \)-mixing \( d \times d \) dimensional matrix process with mixing rate \( \{\phi(m)\} \). Then

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

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 Theorem A.6 in [10] (p.278) we have

\[ \|E[G_j G_k^*]\| \leq 2c d^2 \sqrt{\phi(|j-k|)}, \quad \forall j, k \]

Consequently, by using the inequality

\[ \|tr F\| \leq d\|F\|, \quad \forall F \in \mathcal{R}^{d \times d} \]

We get

\[ \|S_i^{(T)}\|_2^2 = E\| \sum_{j,k=iT}^{(i+1)T-1} G_j G_k^* \| \]
This gives the desired result. \hfill \Box

**Lemma 2.** Let $F_k = \phi_k \phi_k^\top$, 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 $\xi_k$). Then by (25),

$$
\|S^\top_k\|_2 = \| \sum_{t=T}^{(i+1)T-1} [\phi_t \phi_t^\top - E\phi_t \phi_t^\top] \|_2
\leq \sum_{k,j=-\infty}^{\infty} a_k a_j \| \sum_{t=T}^{(i+1)T-1} [\varepsilon(t,k)\varepsilon(t,j)^\top - E\varepsilon(t,k)\varepsilon(t,j)^\top] \|_2
+ 2 \sum_{j=-\infty}^{\infty} a_j \| \sum_{t=T}^{(i+1)T-1} \varepsilon(t,j)\xi_t^\top \|_2
$$

(27)

Note that for any fixed $k$ and $j$, both the processes $\{\varepsilon(t,k)\varepsilon(t,j)^\top\}$ and $\{\varepsilon(t,j)\}$ are $\phi$-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

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

(28)

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

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

(29)

and

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

(30)
Now, by the summability of \( \{a_j\} \),
\[
\sum_{|k-j| \geq \sqrt{T}} a_k a_j \to 0, \quad \text{as } T \to \infty
\]
Hence by (29)
\[
\sum_{|k-j| \geq \sqrt{T}} a_k a_j f_{k,j}(T) = o(T)
\] (31)
and by (30)
\[
\sum_{|k-j| < \sqrt{T}} a_k a_j f_{k,j}(T) = o(T).
\] (32)
Combining (31) and (32) gives
\[
\sum_{k,j=-\infty}^{\infty} a_k a_j f_{k,j}(T) = o(T).
\] (33)
By this and Lemma 1, we know that the second last term in (27) is also of the order \( o(T) \) uniformly in \( i \). Hence, \( \{F_k\} \in \mathcal{M}_2 \) by the definition (12).

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

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

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

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

Proof. Without loss of generality assume that \( \xi_k \equiv 0 \). Let us denote
\[
A \triangleq \sum_{j=-\infty}^{\infty} a_j
\] (34)
where \( \{a_j\} \) is defined by (26). Then by the Schwarz inequality from (25) we have
\[
\|\phi_k\|^2 \leq A \sum_{j=-\infty}^{\infty} a_j \|\xi_{k-j}\|^2
\]
Consequently, by the Hölder inequality and (15) we have for \( \varepsilon \leq \alpha A^{-2} \)

\[
E \exp \{ \varepsilon \sum_{i=1}^{n} \| F_{ji} \| \} \\
\leq E \exp \{ \varepsilon A \sum_{j=-\infty}^{\infty} a_j \sum_{i=1}^{n} \| \varepsilon_{j,-j} \|^2 \} \\
= E \prod_{j=-\infty}^{\infty} \exp \{ \varepsilon A a_j \sum_{i=1}^{n} \| \varepsilon_{j,-j} \|^2 \} \\
\leq \prod_{j=-\infty}^{\infty} \left( E \exp \{ \varepsilon A^2 \sum_{i=1}^{n} \| \varepsilon_{j,-j} \|^2 \} \right)^{a_j} \\
\leq \prod_{j=-\infty}^{\infty} \left( M \exp \{ Kn \} \right)^{a_j} \\
= M \exp \{ Kn \}.
\]

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 \( i_1 < i_2 < \ldots < i_n, \forall n \geq 1 \),

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

Then for any \( L > 0 \) and any \( n \geq i \geq 0 \),

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

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 \( I(x \geq L) \leq e^{\frac{x}{L}} / e^{\frac{L}{2}} \) and then using (35), we have for any subsequence \( j_1 < j_2 < \ldots < j_k \)

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

18
By this we have

\[
E \exp\left\{ \sum_{j=i+1}^{n} \frac{1}{2} z_j I(z_j \geq L) \right\}
\]

\[= E \prod_{j=i+1}^{n} \exp\left\{ \frac{1}{2} z_j I(z_j \geq L) \right\}
\]

\[\leq E \prod_{j=i+1}^{n} \{1 + \exp\left( \frac{1}{2} z_j \right) I(z_j \geq L) \}
\]

\[= E \prod_{j=i+1}^{n} \{1 + f_j \}
\]

\[= E \left\{ \sum_{k=0}^{n-i} \sum_{i+1 \leq j_1 < \ldots < j_k \leq n} f_{j_1} \cdots f_{j_k} \right\}
\]

\[\leq e^b \left\{ \sum_{k=0}^{n-i} \sum_{i+1 \leq j_1 < \ldots < j_k \leq n} \exp\left\{ (a - \frac{L}{2}) k \right\} \right\}
\]

\[= e^b \prod_{j=i+1}^{n} \{1 + \exp(a - \frac{L}{2}) \}
\]

\[\leq \exp\left\{ (n - i) \exp(a - \frac{L}{2}) + b \right\}
\]

This completes the proof of Lemma 5. \(\square\)

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

**Proof.** Set for any fixed \(k\) and \(l\),

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

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

\[\sum_{j=i+1}^{n} \|S_j^{(T)}\| \leq \sum_{k,l = -\infty}^{\infty} a_k a_l \sum_{j=i+1}^{n} z_j + \]

\[+ 2 \sum_{k=-\infty}^{\infty} a_k \sum_{j=i+1}^{n} \| (j+1)T-1 \sum_{t=jT}^{(j+1)T-1} \varepsilon(t, k) \xi_t^\top \|.
\]

(36)

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

\[E \exp \left\{ \mu \sum_{k,l = -\infty}^{\infty} a_k a_l \sum_{j=i+1}^{n} z_j \right\}
\]
\begin{align}
&= 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{\alpha A}{A}}
\end{align}

where $A$ is defined by (34).

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

\[
\|\varepsilon(t, k)\varepsilon(t, l)\| \leq \frac{1}{2} (\|\varepsilon(t, k)\|^2 + \|\varepsilon(t, l)\|^2) \\
\leq \frac{1}{2} (\|\varepsilon_{t-k}\|^2 + \|\varepsilon_{t-l}\|^2)
\]

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 condition (35) with $a = (K + c)T$ and $b = \log M$, where $a$ is defined as in (15). Consequently, by Lemma 5 we have for any $L > 0$

\[
E \exp \left\{ \frac{\alpha}{2} \sum_{j=i+1}^{n} z_j I(z_j \geq LT) \right\} \\
\leq M \exp \left\{ e^{(K + c - \frac{\alpha L}{2})T (n - i)} \right\}
\]

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

\[
E \exp \left\{ 2\mu A^2 \sum_{j=i+1}^{n} z_j I(z_j \geq LT) \right\} \leq M \exp \left\{ \mu \delta(T) (n - i) \right\}
\]

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

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

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

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

\[
\exp \left\{ 2\mu A^2 \sum_{j=i+1}^{n} x_j \right\} \leq \prod_{j=i+1}^{n} (1 + 4\mu A^2 x_j)
\]

20
As noted before, for any fixed $k$ and $l$, the process $\{\varepsilon(t, k)\varepsilon(t, l)^{\tau}\}$ is $\phi$-mixing with mixing rate $\phi(m - |k - l|)$. Consequently, for any fixed $k$ and $l$, both $\{z_j\}$ and $\{x_j\}$ are also $\phi$-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 Lemma 6.2 in [7] (p.1383), we have

$$E \prod_{j=i+1}^{n} (1 + 4\mu A^2 x_j)$$

$$\leq 2 \left\{ 1 + 8\mu A^2 [f_{kl}(T) + 2LT\phi(T + 1 - |k - l|)] \right\}^{n-i}$$

$$\leq 2 \exp \left\{ 8\mu A^2 [f_{kl}(T) + 2LT\phi(T + 1 - |k - l|)] (n - i) \right\}.$$  \hfill (41)

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

$$E \exp \{\mu A^2 \sum_{j=i+1}^{n} z_j \}$$

$$\leq \left\{ E \exp \left\{ 2\mu A^2 \sum_{j=i+1}^{n} z_j I(z_j \geq LT) \right\} \right\}^{1/2} \left\{ E \exp \left\{ 2\mu A^2 \sum_{j=i+1}^{n} x_j \right\} \right\}^{1/2}$$

$$\leq \sqrt{2M} \exp \left\{ \mu [\delta(T) + 8A^2 f_{kl}(T) + 16LT A^2 \phi(T + 1 - |k - l|)] (n - i) \right\}.$$

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 $\mu > 0$,

$$E \exp \left\{ \mu \sum_{k,l=-\infty}^{\infty} a_k a_l \sum_{j=i+1}^{n} z_j \right\}$$

$$\leq \sqrt{2M} \exp \{\mu g(T)(n - i)\}.$$ 

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. \hfill \Box

**Proof of Theorem 2.**

**Necessity:** Let $\{\phi_k \phi_k^\tau\} \in \mathcal{S}_p$ for $p = 2$. Then by Lemma 2 and Theorem 3.1 in [7], we know that $\{\phi_k \phi_k^\tau\} \in \mathcal{S}$. Consequently, by Lemma 3 we know that (17) holds.
Sufficiency: If condition (17) holds, then by Lemma 3 we have \( \{\phi_k \phi_k^*\} \in S \). By this and Lemmas 4 and 6, we know that Theorem 1 is applicable, and consequently \( \{\phi_k \phi_k^*\} \in S_p, \forall p \geq 1 \). This completes the proof.

\[ \square \]

**Proof of Theorem 5.**

We need to verify all the conditions in Theorem 4.1 of [8]. 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

\[
\delta(m) \to 0, \quad \text{as} \quad m \to \infty
\]

and

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

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

\[
\|E[\varepsilon(k,i)\varepsilon(k,j)^* | \mathcal{F}_{k-m}] - E[\varepsilon(k,i)\varepsilon(k,j)^*] \|_q \leq C_q[\phi(m - \max(i,j))]^{1-q^{-1}} \tag{43}
\]

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 \( \sup_{i,j}\|\varepsilon \varepsilon_i^* \|_q \), and where by definition \( \phi(m) \overset{\Delta}{=} 1 \) 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

\[
\sup_k \|E[\phi_k \phi_k^* | \mathcal{F}_{k-m}] - E[\phi_k \phi_k^*] \|_q \\
\leq \sum_{i,j=0}^\infty a_i a_j \sup_k \|E[\varepsilon(k,i)\varepsilon(k,j)^* | \mathcal{F}_{k-m}] - E[\varepsilon(k,i)\varepsilon(k,j)^*] \|_q \\
+ 2 \sup_k \|\xi_k\| \sum_{i=0}^\infty a_i \|E[\varepsilon(k,i) | \mathcal{F}_{k-m}] \|_q \tag{44}
\]

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 \( m \to \infty \) 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 \( m \to \infty \). Hence, (42) is true and the
proof of Theorem 5 is complete. To find the degree of approximation, define, analogously to Theorem 4.1 in [8],
\[ \sigma(\mu) = \min_{m \geq 1} \{m\sqrt{\mu} + \delta(m)\}. \] (45)

5 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:

(i). For a large class of nonstationary weakly dependent signals, the condition (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-i.i.d.-mixing.

(ii). 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);

(iii). A “three stage procedure” for the tracking performance analysis is delineated (see Section 3), 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


