Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 50-1 (2017) 13222–13227

Exponential Stability of LMS-Based Distributed Adaptive Filters ⋆

Siyu Xie *, **Lei Guo** *

* *Institute of Systems Science, AMSS, Chinese Academy of Sciences, Beijing, 100190, China, (e-mail: xiesiyu7@gmail.com, Lguo@amss.ac.cn).*

Abstract: In this work, we consider a class of distributed adaptive filters based on the standard least mean squares (LMS) algorithm, which is proposed to track an unknown signal process in sensor networks. We analyze the stability by introducing a stochastic cooperative information (SCI) condition, in the case of non-independent, non-stationary and possibly unbounded signals. Under the SCI condition, the distributed adaptive filters based on the standard LMS will be shown to be able to track a dynamic process of interest from noisy measurements by a set of sensors working collaboratively, in the natural scenario where any sensor cannot fulfil the estimation task individually.

Keywords: Least mean squares, distributed adaptive filters, L_p-exponentially stability, graph topology, stochastic averaging, stochastic cooperative information

1. INTRODUCTION

A. *Motivation*

Estimating or tracking an unknown signal process of interest based on the data gathered by a group of spatially distributed sensors has attracted a lot of research attention recently. There are basically two approaches for this problem, i.e., centralized and decentralized processing. In many practical situations, for centralized processing, collecting measurements from all other distributed sensors over the networks may not be feasible due to limited communication capabilities, energy consumptions, packet losses or privacy considerations. Alternatively, for the distributed processing each sensor acts as an individual adaptive filter, which estimates the signal by using its local observations and the estimation derived from its neighboring sensors. There are basically three types of decentralized strategies for distributed algorithms in the literature, namely, incremental, consensus and diffusion strategies. For example, Lopes and Sayed (2006) used incremental strategies to analyze distributed LMS algorithms. However, the incremental strategies suffer from several shortcomings and we can refer Sayed (2014) for a detailed account. Among all the three strategies, the last two are fully decentralized and do not suffer from the limitations of the incremental strategies, in which each sensor can communicate with its neighbors and no cyclic path is required. The consensus strategies have attracted much attention in diverse research areas, such as distributed estimation problems (see in Chen et al. (2014b); Kar and Moura (2011); Olfati-Saber and Shamma (2005); Schizas et al. (2009); Solo (2015)) and distributed optimization problems (see in Nedic and Ozdaglar (2009); Tsitsiklis et al. (1984)). As for the diffusion strategies, the Combine-Then-Adapt (CTA) diffusion strategy and the Adapt-Then-Combine (ATC) diffusion strategy can be found in Lopes and Sayed (2008); Piggott and Solo (2015).

However, due to the mathematical difficulty in analyzing the product of random matrices in a general setting, almost all of the existing literature on stability and convergence analysis require independency or stationarity assumptions for the system signals (see Lopes and Sayed (2008); Olfati-Saber and Shamma (2005); Sayed (2014); Schizas et al. (2009); Solo (2015)), which are quite stringent in many practical situations including feedback systems. Thus it is necessary to provide a stability analysis under more general, correlated and non-stationary conditions. To this end, Chen, Liu and Guo (see Chen et al. (2014a)) firstly analyzed the stability of a normalized LMS diffusion algorithm under a stochastic cooperative information (SCI) condition without independency or stationarity considerations. Their stability results and analysis were later improved in Chen et al. (2016) and with a detailed performance analysis given in Chen et al. (2015). In our recent work (Xie and Guo (2015)), we analyzed the stability of normalized consensus LMS algorithm under a more general SCI condition than that used in Chen et al. (2014a, 2015, 2016) without requiring independency nor stationarity conditions.

Note that the distributed algorithms in Chen et al. (2014a, 2015, 2016) and Xie and Guo (2015) are normalized ones, which may slow down the rate of adaptation, especially in the case of unbounded stochastic signals. This may also account for the fact that most of the literature study the standard LMS. In this paper, we will study consensus algorithm based on the standard LMS and will give a general SCI condition for the stability of this algorithm for possibly unbounded, non-independent and non-stationary signals. Despite of the similarity between normalized and standard situations, the analysis methods have essential

⋆ This work was supported by the National Natural Science Foundation of China under grants 61134008 and 61227902.

10.1016/j.ifacol.2017.08.1956

differences, because the methods used in the previous work (Chen et al. (2014a, 2015, 2016); Xie and Guo (2015)) fail to deal with possibly unbounded signals. Because of this, we need to use some stochastic averaging theorems to overcome this difficulty in the present work.

The rest of the paper is organized as follows. We will first introduce some preliminary notations and concepts in Subsection 1.B. The standard LMS based distributed adaptive filters will be presented in Section 2. In Section 3, we will give some definitions and conditions used in the paper. The main results together with the related proofs will be given in Sections 4 and 5, respectively. Section 6 concludes the paper with some remarks.

B. Preliminaries

The notations to be used in the paper are collected as follows: The set of $m \times n$ matrices with real entries is denoted as $\mathbb{R}^{m\times n}$. Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times n}$ be two symmetric matrices, then $A \geq B$ and $B \leq A$ both mean that $A - B$ is a positive semidefinite matrix. We denote by $\lambda_{max}(X)$ and $\lambda_{min}(X)$ the maximum and minimum eigenvalue of a matrix X. Recall that the Euclidean norm of a matrix $X \in \mathbb{R}^{m\times n}$ is the maximum singular value, i.e., $\|X\| \triangleq \{\lambda_{max}(XX^T)\}^{\frac{1}{2}}$ and the L_p-norm of a random matrix Y is defined as $\|Y\|_{L_p} \triangleq \{\mathbb{E} \parallel Y \parallel^p\}^{\frac{1}{p}}, p > 0$, where $\mathbb{E}$ is the mathematical expectation operator.

Graph theory. An undirected weighted graph $\mathcal{G}$ is a pair defined by $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, where $\mathcal{V} = \{1, 2, ..., n\}$ is the set of nodes, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges, the structure of the graph and the interaction strength among neighboring nodes are both described by matrix $\mathcal{A} = \{a_{ij}\}_{n\times n}$, which is called the weighted adjacency matrix with $a_{ij} \geq 0, \sum_{j=1}^{n} a_{ij} = 1, \forall i = 1, \dots, n$. Since the graph $\mathcal{G}$ is undirected, we have $a_{ij} = a_{ji}$, and so the matrix $\mathcal{A}$ is symmetric.

Let the node i denote the ith sensor and (i, j) denote the communication from sensor i to sensor j. Note that $(i, j) \in \mathcal{E} \Leftrightarrow a_{ij} > 0$. The set of neighbors of sensor i is denoted as

$$\mathcal{N}_i = \{l \in V | (l, i) \in \mathcal{E}\}.$$

The Laplacian matrix $\mathcal{L}$ of the graph $\mathcal{G}$ is defined as $\mathcal{L} = I - \mathcal{A}$ and from Godsil and Royle (2001), we know that its ordered eigenvalues are

$$0 = \lambda_1(\mathcal{L}) \leq \lambda_2(\mathcal{L}) \leq \cdots \leq \lambda_n(\mathcal{L}) \leq 2.$$

The smallest eigenvalue $\lambda_1(\mathcal{L})$ is always equal to zero, with $\frac{1}{\sqrt{n}}(1, \dots, 1)^T$ being the corresponding unit eigenvector. In addition, when the graph is connected (there exists a path between each pair of nodes), we have $\lambda_2(\mathcal{L}) > 0$.

2. PROBLEM FORMULATION

In the paper, we assume that the model at any sensor $i (i = 1, \dots, n)$ is described by a stochastic time-varying linear regression as follows

$$y_k^i = (\boldsymbol{\varphi}_k^i)^T \boldsymbol{\theta}_k + v_k^i, \quad k \geq 0, \tag{1}$$

where $(\cdot)^T$ denotes transpose operator, y_k^i is the scalar observation of sensor i at time k, v_k^i is the disturbance or unmodeled dynamics, $\boldsymbol{\varphi}_k^i$ is the $m \times 1$-dimensional stochastic regressor of sensor i, and $\boldsymbol{\theta}_k$ is an unknown $m \times 1$-dimensional time-varying parameter vector whose variation at time k is denoted by $\boldsymbol{\omega}_k$, i.e.,

$$\boldsymbol{\omega}_k \triangleq \boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1}, \quad k \geq 1. \tag{2}$$

The well-known standard LMS algorithm is defined recursively by

$$\widehat{\boldsymbol{\theta}}_{k+1}^i = \widehat{\boldsymbol{\theta}}_k^i + \mu \boldsymbol{\varphi}_k^i [y_k^i - (\boldsymbol{\varphi}_k^i)^T \widehat{\boldsymbol{\theta}}_k^i], \quad k \geq 0, \tag{3}$$

where $\mu > 0$ is the adaptation gain, and the increment of the algorithm is opposite to the stochastic gradient of the mean square prediction error

$$e_k^i(\boldsymbol{\theta}) = \mathbb{E}[y_k^i - (\boldsymbol{\varphi}_k^i)^T \boldsymbol{\theta}]^2, \quad k \geq 0.$$

Thus it is a type of steepest descent algorithm that aims at minimizing $e_k^i(\theta)$ recursively and this algorithm has been implemented in many adaptive signal processing applications (see in Widerow and Stearns (1985); Macchi (1995); Solo and Kong (1995); Haykin (1996)).

For most practical sensor networks, the sensor $i (i = 1, \dots, n)$ can only obtain the information from its neighbors $\{l \in \mathcal{N}_i\}$, and we can consider the following consensus-type algorithm based on the standard LMS (see Solo (2015); Sayed (2014))

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{k+1}^i = &\widehat{\boldsymbol{\theta}}_k^i + \mu \Big\{ \boldsymbol{\varphi}_k^i [y_k^i - (\boldsymbol{\varphi}_k^i)^T \widehat{\boldsymbol{\theta}}_k^i] \\ &- \nu \sum_{l \in \mathcal{N}_i} a_{li} (\widehat{\boldsymbol{\theta}}_k^i - \widehat{\boldsymbol{\theta}}_k^l) \Big\}, \quad k \geq 0, i = 1, \dots, n, \end{aligned} \tag{4}$$

where $\nu \in (0, 1)$ is a weighting constant, and $\mu > 0$ is the adaptation gain which is usually chosen to be small. Before the theoretical analysis, we may need the following notations

$$\begin{aligned}
&\boldsymbol{Y}_k \triangleq \mathrm{col}\{y_k^1, \dots, y_k^n\} && (n \times 1), \\
&\boldsymbol{\Phi}_k \triangleq \mathrm{diag}\{\boldsymbol{\varphi}_k^1, \dots, \boldsymbol{\varphi}_k^n\} && (mn \times n), \\
&\boldsymbol{V}_k \triangleq \mathrm{col}\{v_k^1, \dots, v_k^n\} && (n \times 1), \\
&\boldsymbol{\Omega}_k \triangleq \mathrm{col}\{\underbrace{\boldsymbol{\omega}_k, \dots, \boldsymbol{\omega}_k}_{n}\} && (mn \times 1), \\
&\boldsymbol{\Theta}_k \triangleq \mathrm{col}\{\underbrace{\boldsymbol{\theta}_k, \dots, \boldsymbol{\theta}_k}_{n}\} && (mn \times 1), \\
&\widehat{\boldsymbol{\Theta}}_k \triangleq \mathrm{col}\{\widehat{\boldsymbol{\theta}}_k^1, \dots, \widehat{\boldsymbol{\theta}}_k^n\} && (mn \times 1), \\
&\widetilde{\boldsymbol{\Theta}}_k \triangleq \mathrm{col}\{\widetilde{\boldsymbol{\theta}}_k^1, \dots, \widetilde{\boldsymbol{\theta}}_k^n\}, \text{ where } \widetilde{\boldsymbol{\theta}}_k^i = \widehat{\boldsymbol{\theta}}_k^i - \boldsymbol{\theta}_k && (mn \times 1), \\
&\boldsymbol{F}_k \triangleq \mathrm{diag}\{\boldsymbol{\varphi}_k^1 (\boldsymbol{\varphi}_k^1)^T, \dots, \boldsymbol{\varphi}_k^n (\boldsymbol{\varphi}_k^n)^T\} \\
&\quad \triangleq \mathrm{diag}\{\boldsymbol{F}_k^1, \dots, \boldsymbol{F}_k^n\} && (mn \times mn), \\
&\boldsymbol{G}_k \triangleq \boldsymbol{F}_k + \nu(\mathcal{L} \otimes I_m) && (mn \times mn).
\end{aligned} \tag{5}$$

where col$\{\cdots\}$ denotes a vector by stacking the specified vectors, diag$\{\cdots\}$ is used in a non-standard manner which means that $m \times 1$ column vectors are combined "in a diagonal manner" resulting in a $mn \times n$ matrix, and $\otimes$ denotes the Kronecker product. Note also that $\boldsymbol{\Omega}_k$ and $\boldsymbol{\Theta}_k$ mean just the n-times replication of identical vectors $\boldsymbol{\omega}_k$ and $\boldsymbol{\theta}_k$, respectively. Note that $\boldsymbol{\Omega}_k$ and $\boldsymbol{\Theta}_k$ just mean the n-times replication of identical vectors $\boldsymbol{\omega}_k$ and $\boldsymbol{\theta}_k$, respectively. By (1) and (2), we have

$$\boldsymbol{Y}_k = \boldsymbol{\Phi}_k^T \boldsymbol{\Theta}_k + \boldsymbol{V}_k, \quad k \geq 0, \tag{6}$$

and

$$\boldsymbol{\Omega}_{k+1} = \boldsymbol{\Theta}_{k+1} - \boldsymbol{\Theta}_k. \tag{7}$$

From (4), we obtain that

$$\widehat{\boldsymbol{\Theta}}_{k+1} = \widehat{\boldsymbol{\Theta}}_k + \mu \boldsymbol{\Phi}_k (\boldsymbol{Y}_k - \boldsymbol{\Phi}_k^T \widehat{\boldsymbol{\Theta}}_k) - \mu\nu(\mathcal{L} \otimes I_m)\widehat{\boldsymbol{\Theta}}_k, \tag{8}$$

where matrix $\mathcal{L}$ is the Laplacian matrix of graph $\mathcal{G}$. Let us denote $\widetilde{\boldsymbol{\Theta}}_k = \widehat{\boldsymbol{\Theta}}_k - \boldsymbol{\Theta}_k$ and notice (6) and (7), we can get

$$\begin{aligned}\widetilde{\boldsymbol{\Theta}}_{k+1} =& \widetilde{\boldsymbol{\Theta}}_k - \mu \boldsymbol{\Phi}_k \boldsymbol{\Phi}_k^T \widetilde{\boldsymbol{\Theta}}_k - \mu\nu(\mathcal{L} \otimes I_m)\widehat{\boldsymbol{\Theta}}_k \\ &+ \mu \boldsymbol{\Phi}_k \boldsymbol{V}_k - \boldsymbol{\Omega}_{k+1},\end{aligned}$$

and because $(\mathcal{L} \otimes I_m)\boldsymbol{\Theta}_k = 0$, we have

$$\widetilde{\boldsymbol{\Theta}}_{k+1} = (I_{mn} - \mu \boldsymbol{G}_k)\widetilde{\boldsymbol{\Theta}}_k + \mu \boldsymbol{\Phi}_k \boldsymbol{V}_k - \boldsymbol{\Omega}_{k+1}. \tag{9}$$

Obviously, its homogeneous equation is

$$\widetilde{\boldsymbol{X}}_{k+1} = (I_{mn} - \mu \boldsymbol{G}_k)\widetilde{\boldsymbol{X}}_k, \quad k \geq 0. \tag{10}$$

Note that by the internal-external stability results in Guo (1994) (see Propositions 2.1 and 2.2 there), we know that the stability of (9) is essentially determined by the exponential stability of the homogeneous equation (10). This motivates us to give some definitions and conditions on exponential stability at the beginning of next section.

3. DEFINITIONS AND CONDITIONS

3.1 Definitions

Before further discussions, we first present the following definitions introduced in Guo and Ljung (1995a,b).

Definition 1. For a sequence of $d \times d$ random matrices $A = \{A_k, k \geq 0\}$ and real numbers $p \geq 1, \mu^* \in (0,1)$, the L_p-exponentially stable family $S_p(\mu^*)$ is defined by

$$\begin{aligned}S_p(\mu^*) =& \Bigg\{ A : \Big\| \prod_{j=i+1}^{k} (I - \mu A_j) \Big\|_{L_p} \leq M(1-\mu\alpha)^{k-i}, \\ & \forall k \geq i, \forall i \geq 0, \forall \mu \in (0, \mu^*), \\ & \text{for some } M > 0 \textit{ and } \alpha \in (0,1) \Bigg\}.\end{aligned} \tag{11}$$

Likewise, we have the following definition about averaged deterministic exponentially stable family.

Definition 2. For a sequence of $d \times d$ random matrices $A = \{A_k, k \geq 0\}$ and real numbers $p \geq 1, \mu^* \in (0,1)$, the averaged deterministic exponentially stable family $S(\mu^*)$ is defined by

$$\begin{aligned}S(\mu^*) =& \Bigg\{ A : \Big\| \prod_{j=i+1}^{k} (I - \mu \mathbb{E}[A_j]) \Big\| \leq M(1-\mu\alpha)^{k-i}, \\ & \forall k \geq i, \forall i \geq 0, \forall \mu \in (0, \mu^*), \\ & \text{for some } M > 0 \textit{ and } \alpha \in (0,1) \Bigg\}.\end{aligned} \tag{12}$$

In addition, we denote

$$S_p \triangleq \bigcup_{\mu^* \in (0,1)} S_p(\mu^*), S \triangleq \bigcup_{\mu^* \in (0,1)} S(\mu^*). \tag{13}$$

Remark 1. In the next section, we will show that for a general class of $\{\boldsymbol{G}_k\}$, the L_p-exponential stability of (10) can be obtained from the averaged deterministic exponential stability under some general conditions, without boundedness, independency, stationarity properties of the system signals.

Definition 3. Let $p \geq 0, A = \{A_k, k \geq 0\}$. Set

$$S_j^{(T)} = \sum_{t=jT}^{(j+1)T-1} (A_t - \mathbb{E}[A_t]), \tag{14}$$

and define the following weakly dependent set

$$\mathcal{M}_p = \left\{ A : \sup_j \|S_j^{(T)}\|_{L_p} = o(T), \quad as \quad T \to \infty \right\}. \tag{15}$$

Remark 2. $S_j^{(T)}$ includes T zero-mean terms and $\mathcal{M}_p$ requires that its norm increases slower than T, which indicates that certain weakly dependent property is imposed on the A_k. As has been shown by Guo et al. (1993), martingale difference sequences, α-mixing and ϕ-mixing sequences, and linear random processes are all in the set $\mathcal{M}_p$.

Definition 4. Let $\{A_k\}$ be a matrix sequence and $b_k, \forall k \geq 0$ be a positive scalar sequence. Then by $A_k = O(b_k)$ we mean that there exists a constant $M > 0$ such that

$$\|A_k\| \leq M b_k, \quad \forall k \geq 0. \tag{16}$$

3.2 Conditions

Throughout the sequel, we may use the following conditions.

*Condition 1 (**Network topology**). The graph $\mathcal{G}$ is connected.*

Remark 3. By this condition, we know that the matrix $\mathcal{L}$ has only one zero eigenvalue with all other eigenvalues be positive and not more than 2.

*Condition 2 (**SCI condition**). There exist an integer $h > 0$ and a constant $\delta > 0$ such that*

$$\sum_{i=1}^{n} \sum_{j=k+1}^{k+h} \mathbb{E}[\boldsymbol{\varphi}_j^i (\boldsymbol{\varphi}_j^i)^T] \geq \delta I_m, \quad \forall k \geq 0. \tag{17}$$

Remark 4. When the regressor process $\boldsymbol{\varphi}_j^i$ is identically zero, it is clear that *Condition 2* is not satisfied, which is a trivial case where the system is not identifiable since the observations contain no information about the unknown signal $\{\boldsymbol{\theta}_k\}$. In order to estimate the unknown process, it is necessary to require some "persistent of excitation" or "full rank" properties on the regressors. *Condition 2* just requiring the regressor covariance matrices to add up to full rank over the space span and the time span of a given length, which is quite weak and is a natural extension of the classical case of a single sensor (Guo et al. (1997)). Notice that no conditions of independence and stationarity are made on the signals.

Condition 3. There exists $r > 2$ such that

$$\sigma \triangleq \sup_k \|\boldsymbol{\Phi}_k \boldsymbol{V}_k\|_{L_r} < \infty, \quad \gamma \triangleq \sup_k \|\boldsymbol{\Omega}_k\|_{L_r} < \infty.$$

Remark 5. Note that *Condition 3* only assumes the noises and parameter variations are bounded in an averaging sense, which can be used to obtain an upper bound of the tracking error. Here we do not need independency and Gaussian property for the noises, nor the signal $\boldsymbol{\theta}_k$ is assumed to be a random walk

4. THE MAIN RESULTS

In this section, we first present the L_p-exponential stability of (10) for a large class of non-independent, non-stationary and unbounded signals. Let us consider the following decomposition of $\{\boldsymbol{\varphi}_k^i\}(i=1,\ldots,n)$:

$$\begin{aligned}&\boldsymbol{\varphi}_k^i=\sum_{j=-\infty}^{\infty}A^i(k,j)\boldsymbol{\varepsilon}_{k-j}^i+\boldsymbol{\xi}_k^i,\\&\sum_{j=-\infty}^{\infty}\sup_k\|A^i(k,j)\|<\infty,\qquad i=1,\ldots,n,\end{aligned}\tag{18}$$

where $\{\boldsymbol{\xi}_k^i\}(i=1,\ldots,n)$ are all $m\times 1$-dimensional bounded deterministic processes and $\{\boldsymbol{\varepsilon}_k^i\}(i=1,\ldots,n)$ are general $d\times 1$-dimensional ϕ-mixing sequences. We assume that the weighting matrices $A^i(k,j)\in\mathbb{R}^{m\times d}(i=1,\ldots,n)$ are deterministic. It is obvious that $\{\boldsymbol{\varphi}_k^i\}(i=1,\ldots,n)$ have similar form as the well-known Wold decomposition for wide-sense stationary processes. However, here the process $\{\boldsymbol{\varphi}_k^i\}(i=1,\ldots,n)$ need not to be stationary processes nor Markov chains.

Recall that a random sequence $\boldsymbol{\varepsilon}_k$ is called ϕ-mixing if there exists a non-increasing function $\phi(m)$, which is 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

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

where $\mathcal{F}_i^j(-\infty\leq i\leq j\leq\infty)$ is the σ-algebra generated by $\{\boldsymbol{\varepsilon}_k,i\leq k\leq j\}$, where the definition of σ-algebra can be found in Chow and Teicher (1978). The ϕ-mixing property is a standard concept for describing weakly dependent random processes. Any M-dependence sequences and sequences generated from bounded white noises via a stable linear filter all satisfy ϕ-mixing property.

With these properties, we will present the main result of this section.

Theorem 1. Consider the distributed LMS error equation (10) and suppose that *Conditions 1* and *2* are satisfied. Let the signal processes $\{\boldsymbol{\varphi}_k^i\}(i=1,\ldots,n)$ be generated by (18) where $\{\boldsymbol{\xi}_k^i\}(i=1,\ldots,n)$ are all $m\times 1$-dimensional bounded deterministic processes and $\{\boldsymbol{\varepsilon}_k^i\}(i=1,\ldots,n)$ are $d\times 1$-dimensional ϕ-mixing sequences which satisfy that for any $s\geq 1$ and any integer sequence $j_1<j_2<\cdots<j_s$,

$$\mathbb{E}\left[exp\left(\alpha\sum_{t=1}^{s}\|\boldsymbol{\varepsilon}_{j_t}^i\|^2\right)\right]\leq Mexp(Ks),\quad i=1,\ldots,n,\tag{19}$$

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

$$\left[\mathbb{E}\left\|\prod_{j=k+1}^{t}(I_{mn}-\mu\boldsymbol{G}_j)\right\|^p\right]^{1/p}\leq M(1-\mu\alpha)^{t-k},\forall t\geq k\geq 0.$$

Remark 6. Taking $A^i(k,0)=I,A^i(k,j)=0,\xi_k^i=0,\forall k,\forall j\neq 0,\forall i=1,\ldots,n$ in (18), we know that $\{\boldsymbol{\varphi}_k^i\}$ coincides with $\{\boldsymbol{\varepsilon}_k^i\}$, which means that *Theorem 1* is applicable to any $\phi-$mixing sequence. Furthermore, if $\{\boldsymbol{\varepsilon}_k^i\}$ is bounded, then (19) is automatically satisfied, for example, signals from the normalized LMS algorithm. Note also that, because of the possible unboundedness of $\{\boldsymbol{\varepsilon}_k^i\}$, a linearly filtered $\phi-$mixing process like (18) may no longer be a $\phi-$mixing sequence.

Theorem 2. Consider the standard LMS-based distributed algorithm (9) and suppose that *Conditions 1-3* are satisfied. Let the signal processes $\{\boldsymbol{\varphi}_k^i\}(i=1,\ldots,n)$ be as in *Theorem 1* with (19) satisfied. Then for all $k>0,\nu\in(0,1)$ and all small $\mu>0$,

$$\|\widetilde{\boldsymbol{\Theta}}_k\|_{L_2}=O\left(\sigma+\frac{\gamma}{\mu}\right)+O([1-\mu\beta]^k),$$

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

Proof. Define $\boldsymbol{\Psi}(\cdot,\cdot)$ as follows

$$\begin{cases}\boldsymbol{\Psi}(k+1,j)=(I_{mn}-\mu\boldsymbol{G}_k)\boldsymbol{\Psi}(k,j),\\\boldsymbol{\Psi}(j,j)=I_{mn},\forall k\geq j\geq 0.\end{cases}$$

Then by Theorem 1, we have

$$\|\boldsymbol{\Psi}(k+1,j+1)\|_{L_s}\leq M(1-\mu\beta)^{k-j},\forall k\geq j\geq 0,$$

where M and $\beta\in(0,1)$ are constants and $\frac{1}{s}+\frac{1}{r}=\frac{1}{2}$. Using the definition of $\boldsymbol{\Psi}(k,i)$ and taking μ small enough, we can obtain

$$\begin{aligned}&\|\widetilde{\boldsymbol{\Theta}}_{k+1}\|_{L_2}\\&=\left\|\boldsymbol{\Psi}(k+1,0)\widetilde{\boldsymbol{\Theta}}_0+\sum_{j=0}^{k}\boldsymbol{\Psi}(k+1,j+1)[\mu\boldsymbol{\Phi}_j\boldsymbol{V}_j-\boldsymbol{\Omega}_{j+1}]\right\|_{L_2}\\&\leq O([1-\mu\beta]^{k+1})+\left\|\mu\sum_{j=0}^{k}\boldsymbol{\Psi}(k+1,j+1)\boldsymbol{\Phi}_j\boldsymbol{V}_j\right\|_{L_2}\\&\quad+\left\|\sum_{j=0}^{k}\boldsymbol{\Psi}(k+1,j+1)\boldsymbol{\Omega}_{j+1}\right\|_{L_2}\\&\leq O([1-\mu\beta]^{k+1})+\sigma\mu M\sum_{j=0}^{k}(1-\mu\beta)^{k-j}\\&\quad+\gamma M\sum_{j=0}^{k}(1-\mu\beta)^{k-j}\\&=O([1-\mu\beta]^{k+1})+\frac{M}{\beta}\left(\sigma+\frac{\gamma}{\mu}\right).\end{aligned}$$

This completes the proof.

Remark 7. If both the parameter variation and the disturbance are small, the tracking error will also be small. Here we only use a moment condition on noise in the result, instead of the rather stronger Gaussian noise property.

5. PROOF OF THEOREM 1

Before proving the theorem, we first give a critical lemma based on the stochastic averaging theorem of Guo et al. (1997).

Lemma 3. Let $\{\boldsymbol{F}_k^i\}(i=1,\ldots,n)$ be n random matrix processes and $\boldsymbol{F}_k=\text{diag}\{\boldsymbol{F}_k^1,\ldots,\boldsymbol{F}_k^n\},\boldsymbol{G}_k=\boldsymbol{F}_k+\nu(\mathcal{L}\otimes I_m)$. Then

$$\{\boldsymbol{G}_k\} \in S \Longrightarrow \{\boldsymbol{G}_k\} \in S_p, \quad \forall p \geq 1, \tag{20}$$

provided that the following two conditions are satisfied:

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

$$\mathbb{E}\left[exp\left(\epsilon \sum_{t=1}^{s} \|\boldsymbol{F}_{j_t}^i\|\right)\right] \leq M exp(Ks), \quad \forall i = 1, \ldots, n,$$

holds for any integer sequence $0 \leq j_1 < j_2 < \cdots < j_s$.

2) There exist 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 $s \geq t \geq 0, i = 1, \ldots, n$

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

where $S_j^{(T)(i)} = \sum_{t=jT}^{(j+1)T-1} (\boldsymbol{F}_t^i - \mathbb{E}[\boldsymbol{F}_t^i])$.

Proof. According to *Theorem 1* in Guo et al. (1997), we know that *Lemma 3* will be true if conditions 1) and 2) are satisfied for $\{\boldsymbol{G}_k\}$. By the Hölder inequality, it is easy to see that for any $s \geq 1$ and any integer sequence $0 \leq j_1 < j_2 < \cdots < j_s$, there exist $\epsilon' = \frac{\epsilon}{n} > 0$ and $K' = K + 2\epsilon'\nu > 0$ such that

$$\begin{aligned}
&\mathbb{E}\left[exp\left(\epsilon' \sum_{t=1}^{s} \|\boldsymbol{G}_{j_t}\|\right)\right] \\
\leq &\mathbb{E}\left[exp\left(\epsilon' \sum_{t=1}^{s} \|\boldsymbol{F}_{j_t}\|\right)\right] \cdot exp(2\epsilon' s\nu) \\
= &\mathbb{E}\left[exp\left(\epsilon' \sum_{t=1}^{s} \max_{i=1,\ldots,n} \|\boldsymbol{F}_{j_t}^i\|\right)\right] \cdot exp(2\epsilon' s\nu) \\
\leq &\mathbb{E}\left[exp\left(\epsilon' \sum_{i=1}^{n}\sum_{t=1}^{s} \|\boldsymbol{F}_{j_t}^i\|\right)\right] \cdot exp(2\epsilon' s\nu) \\
\leq &\left\{\prod_{i=1}^{n} \mathbb{E}\left[exp(n\epsilon' \sum_{t=1}^{s} \|\boldsymbol{F}_{j_t}^i\|)\right]\right\}^{\frac{1}{n}} \cdot exp(2\epsilon' s\nu) \\
\leq &M exp(Ks) \cdot exp(2\epsilon' s\nu) \\
= &M exp(K' s),
\end{aligned} \tag{21}$$

for some constants $M > 0, K > 0$. Then denote $P_j^{(T)} = \sum_{t=jT}^{(j+1)T-1} (\boldsymbol{G}_t - \mathbb{E}[\boldsymbol{G}_t])$, we have

$$\begin{aligned}
&P_j^{(T)} \\
= &\sum_{t=jT}^{(j+1)T-1} \left\{[\boldsymbol{F}_t + \nu(\mathcal{L} \otimes I_m)] - \mathbb{E}[\boldsymbol{F}_t + \nu(\mathcal{L} \otimes I_m)]\right\} \\
= &\sum_{t=jT}^{(j+1)T-1} (\boldsymbol{F}_t - \mathbb{E}[\boldsymbol{F}_t]).
\end{aligned}$$

Similarly, by the Hölder inequality, there exist $\mu' = \frac{\mu}{n} > 0$ and a nondecreasing function $g'(T)$ with $g'(T) = ng(T) = o(T), \quad as \quad T \to \infty$ such that

$$\begin{aligned}
&\mathbb{E}\left[exp\left(\mu' \sum_{j=t+1}^{s} \|P_j^{(T)}\|\right)\right] \\
\leq &\mathbb{E}\left[exp\left(\mu' \sum_{j=t+1}^{s} \max_{i=1,\ldots,n} \|S_j^{(T)(i)}\|\right)\right] \\
\leq &\mathbb{E}\left[exp\left(\mu' \sum_{i=1}^{n}\sum_{j=t+1}^{s} \|S_j^{(T)(i)}\|\right)\right] \\
\leq &\left\{\prod_{i=1}^{n} \mathbb{E}\left[exp(n\mu' \sum_{j=t+1}^{s} \|S_j^{(T)(i)}\|)\right]\right\}^{\frac{1}{n}} \\
\leq &M exp\{[\mu g(T) + o(\mu)](s-t)\} \\
= &M exp\{[\mu' g'(T) + o(\mu')](s-t)\}.
\end{aligned} \tag{22}$$

By Guo et al. (1997), this completes the proof.

Remark 8. Condition 1) is immediately satisfied if $\|\boldsymbol{F}_j^i\|, \forall j$ and $i = 1, \ldots, n$ are all bounded, for example, in the normalized LMS case in Guo (1994). Condition 1) generally implies that the distribution of the random processes $\{\boldsymbol{F}_k^i\}(i = 1, \ldots, n)$ have exponentially decaying tails.

Remark 9. Let $\{\boldsymbol{F}_k^i\}(i = 1, \ldots, n)$ be sequences in $\mathcal{M}_p$, then we know that $\sup_j \|S_j^{(T)(i)}\|_{L_p} = o(T)$ as $T \to \infty$, which gives an intuitive explanation for the upper bound in condition 2). Furthermore, condition 2) reflects the weak dependency property of the coefficient matrices and can not be removed in general.

Before proving *Theorem 1*, we also need the following lemmas.

Lemma 4. Let $\max_{i=1,\ldots,n} \sup_k \mathbb{E}\|\boldsymbol{\varphi}_k^i\|_{L_2} < \infty$ and *Conditions 1* and *2* hold. Then for the standard LMS-based distributed algorithm (9), we have $\{\boldsymbol{G}_k\} \in S$.

Proof. Assume that *Conditions 1* and *2* are true and take

$$\mu^* = \frac{1}{3 + \max_{i=1,\ldots,n} \sup_k \mathbb{E}\|\boldsymbol{\varphi}_k^i\|^2}.$$

Then for any $\mu \in (0, \mu^*]$, we have $0 \leq \mu\mathbb{E}[\boldsymbol{G}_k] \leq I_{mn}$. By applying Theorem 1 in Xie and Guo (2015) to the deterministic sequence $\mu\mathbb{E}[\boldsymbol{G}_k]$, it is easy to see that $\{\boldsymbol{G}_k\} \in S(\mu^*)$. This completes the proof.

Lemma 5. Let $\boldsymbol{F}_k^i = \boldsymbol{\varphi}_k^i(\boldsymbol{\varphi}_k^i)^T$, where $\{\boldsymbol{\varphi}_k^i\}$ is defined by (18) with (19) satisfied. Then $\{\boldsymbol{F}_k^i\}(i = 1, \ldots, n)$ satisfy Conditions 1) and 2) of *Lemma 3*.

Proof. By Lemma 4 in Guo et al. (1997), we know that there exist positive constants ϵ_i, M_i and K_i such that for any $s \geq 1$ and $i = 1, \ldots, n$

$$\mathbb{E}\left[exp\left(\epsilon_i \sum_{t=1}^{s} \|\boldsymbol{F}_{j_t}^i\|\right)\right] \leq M_i exp(K_i s), \quad \forall i = 1, \ldots, n,$$

holds for any integer sequence $0 \leq j_1 < j_2 < \cdots < j_s$. Let $\epsilon = \min_{i=1,\ldots,n} \epsilon_i, M = \max_{i=1,\ldots,n} M_i$ and $K = \max_{i=1,\ldots,n} K_i$, then Condition 1) holds. Similarly, by Lemma 6 in Guo et al. (1997), there exist constants M_i and nondecreasing functions $g_i(T)$ with $g_i(T) = o(T), \quad as \quad T \to \infty$, such that for any fixed T, all small $\mu > 0$ and any $s \geq t \geq 0, i = 1, \ldots, n$

$$\mathbb{E}\left[exp\left(\mu\sum_{j=t+1}^{s}\|S_j^{(T)(i)}\|\right)\right]$$
$$\leq M_i exp\{[\mu g_i(T)+o(\mu)](s-t)\},\quad \forall i=1,\ldots,n,$$

where $S_j^{(T)(i)} = \sum_{t=jT}^{(j+1)T-1}(\boldsymbol{F}_t^i - \mathbb{E}[\boldsymbol{F}_t^i])$. Let $M = \max_{i=1,\ldots,n} M_i$ and $g(T) = \max_{i=1,\ldots,n} g_i(T)$, then Condition 2) holds. This completes the proof of *Lemma 5*.

Finally, we complete the proof of *Theorem 1* as follows:

Under *Conditions 1* and *2*, we have $\{\boldsymbol{G}_k\} \in S$ by *Lemma 4*. By this and *Lemma 5*, we know that *Lemma 3* is applicable and consequently $\{\boldsymbol{G}_k\} \in S_p, \forall p \geq 1$. This completes the proof.

6. CONCLUDING REMARKS

We have studied the L_p-exponential stability of the standard LMS-based distributed adaptive filters for possibly unbounded correlated signals in this paper. It has been shown that for a large class of non-stationary weakly dependent signals satisfying *Condition 2*, the LMS-based distributed adaptive filters are exponentially stable. This is then used to obtain a preliminary tracking error bound and may be further used to obtain a refined performance bound in future investigations. There are still a number of problems remain to be solved, for examples,

- To extend the stability analysis to other distributed strategies, e.g. ATC strategy, and other adaptive filtering algorithms, e.g., Kalman filtering and RLS algorithms.
- To incorporate with more complicated network topologies, e.g. time-varying and directed topologies.

REFERENCES

Chen, C., Liu, Z., and Guo, L. (2014a). Stability of diffusion adaptive filters. In *Proc. 19th IFAC World Congress, Cape Town*, 10409–10414, August 2014.

Chen, C., Liu, Z., and Guo, L. (2015). Performance analysis of distributed adaptive filters. *Communications in Information and Systems*, 15(4), 453–476.

Chen, C., Liu, Z., and Guo, L. (2016). Performance bounds of distributed adaptive filters with cooperative correlated signals. *SCIENCE CHINA Information Science*, 59(11), 112202.

Chen, W., Wen, C., Hua, S., and Sun, C. (2014b). Distributed cooperative adaptive identification and control for a group of continuous-time systems with a cooperative pe condition via consensus. *IEEE Transactions on Automatic Control*, 59(1), 91–106.

Chow, Y.S. and Teicher, H. (1978). *Probability theory.* New York: Springer.

Godsil, C. and Royle, G. (2001). *Algebraic graph theory.* Springer-Verlag, New York.

Guo, L. (1994). Stability of recursive stochastic tracking algorithms. *SIAM Journal on Control and Optimization*, 32(5), 1195–1225.

Guo, L. and Ljung, L. (1995a). Exponential stability of general tracking algorithms. *IEEE Transactions on Automatic Control*, 40(8), 1376–1387.

Guo, L. and Ljung, L. (1995b). Performance analysis of general tracking algorithms. *IEEE Transactions on Automatic Control*, 40(8), 1388–1402.

Guo, L., Ljung, L., and Priouret, P. (1993). Performance analysis of the forgetting factor rls algorithm. *International journal of adaptive control and signal processing*, 7(6), 525–538.

Guo, L., Ljung, L., and Wang, G.J. (1997). Necessary and sufficient conditions for stability of lms. *IEEE Transactions on Automatic Control*, 42(6), 761–770.

Haykin, S. (1996). *Adaptive filter theory.* Prentice-Hall, Inc.

Kar, S. and Moura, J.M. (2011). Convergence rate analysis of distributed gossip (linear parameter) estimation: fundamental limits and tradeoffs. *IEEE Journal of Selected Topics in Signal Processing*, 5(4), 674–690.

Lopes, C.G. and Sayed, A.H. (2006). Distributed adaptive incremental strategies: formulation and performance analysis. In *2006 IEEE International Conference on Acoustics, Speech and Signal Processing*, 584–587. Toulouse, France.

Lopes, C.G. and Sayed, A.H. (2008). Diffusion least-mean squares over adaptive networks: formulation and performance analysis. *IEEE Transactions on Signal Processing*, 56(7), 3122–3136.

Macchi, O. (1995). *Adaptive processing: the least mean squares approach with applications in transmission.* New York: Wiley.

Nedic, A. and Ozdaglar, A. (2009). Cooperative distributed multi-agent optimization. *Convex Optimization in Signal Processing and Communications*, 340–386.

Olfati-Saber, R. and Shamma, J.S. (2005). Consensus filters for sensor networks and distributed sensor fusion. In *44th IEEE Conference on Decision and Control*, 6698–6703. Seville, Spain.

Piggott, M.J. and Solo, V. (2015). Stability of distributed adaptive algorithms ii: Diffusion algorithms. In *2015 54th IEEE Conference on Decision and Control (CDC)*, 7428–7433. IEEE.

Sayed, A.H. (2014). Adaptive networks. *Proceedings of the IEEE*, 102(4), 460–497.

Schizas, I.D., Mateos, G., and Giannakis, G.B. (2009). Distributed lms for consensus-based in-network adaptive processing. *IEEE Transactions on Signal Processing*, 57(6), 2365–2382.

Solo, V. (2015). Stability of distributed adaptive algorithms i: Consensus algorithms. In *2015 54th IEEE Conference on Decision and Control (CDC)*, 7422–7427. IEEE.

Solo, V. and Kong, X. (1995). *Adaptive signal processing algorithms.* Prentice-Hall, Inc.

Tsitsiklis, J.N., Bertsekas, D.P., and Athans, M. (1984). Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, 31(9), 803–812.

Widerow, B. and Stearns, S.D. (1985). *Adaptive Signal Processing.* Prentice-Hall Englewood Cliffs NJ.

Xie, S. and Guo, L. (2015). Stability of distributed lms under cooperative stochastic excitation. In *Proc. 34th Chinese Control Conference (CCC)*, 7445–7450. Hangzhou, China, July 28-30, 2015.