Automatica 93 (2018) 12–19

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

A necessary and sufficient condition for stability of LMS-based consensus adaptive filters ☆

Siyu Xie, Lei Guo *

Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
School of Mathematical Science, University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:
Received 4 January 2017
Received in revised form 10 September 2017
Accepted 28 December 2017
Available online 28 March 2018

Keywords:
Least mean squares
Adaptive filters
Consensus strategies
Exponentially stable
Stochastic averaging
Switching networks
Performance analysis

ABSTRACT

This paper investigates the stability and performance of the standard least mean squares (LMS)-based consensus adaptive filters under a changing network topology. We first analyze the stability for possibly unbounded, non-independent and non-stationary signals, by introducing an information condition that can be shown to be not only sufficient but also necessary for the global stability. We also demonstrate that the distributed adaptive filters can estimate a dynamic process of interest from noisy measurements by a set of sensors working in a collaborative manner, in the natural scenario where none of the sensors can fulfill the estimation task individually. Furthermore, we give an analysis of the filtering error under various assumptions without stationarity and independency constraints on the system signals, and thus do not exclude applications to stochastic systems with feedback. In contrast to the analyses of the normalized LMS-based distributed adaptive filters, we need to use stochastic averaging theorems in the stability analysis due to possible unboundedness of the system signals.

1. Introduction

When filtering or tracking an unknown signal or parameter process in a distributed sensor network, each sensor can produce a local estimate based on its own noisy measurements and on the information gathered from other sensors. There are essentially three scenarios for this problem, i.e., centralized processing, distributed processing and the combination of both. In the first scenario, all the sensors transmit information to a fusion center, while in the second scenario, no fusion center is required and any sensor will conduct the estimation task through communicating with its neighbors. For centralized processing, collecting measurements from all other distributed sensors over the network may not be feasible in many practical situations due to limited communication capabilities, energy consumptions, packet losses or privacy considerations. These are the main motivations for the development of the distributed algorithms, in which any sensor only needs to exchange information with its neighbors, which will be more robust, need fewer communications and allow parallel information processing.

There are basically three types of strategies for distributed algorithms in the literature, namely, incremental strategies (Cattivelli & Sayed, 2011; Lopes & Sayed, 2007), consensus strategies (Braca, Marano, & Matta, 2008; Chen, Wen, Hua, & Sun, 2014; Kar & Moura, 2011; Solo, 2015) and diffusion strategies (Cattivelli & Sayed, 2010; Chen & Sayed, 2015a, b; Nosrati, Shamsi, Taheri, & Sedaaghi, 2015; Piggott & Solo, 2016, 2017; Sayed, 2014a, b). Despite of extensive researches on distributed adaptive filtering algorithms in recent years, most of the existing literature on stability and performance analyses require statistical independency or stationarity assumptions for the system signals (e.g., Cattivelli & Sayed, 2010, 2011; Kar & Moura, 2011; Lopes & Sayed, 2007; Nosrati et al., 2015; Sayed, 2014a, b), which cannot be satisfied in many practical situations, for example, signals from feedback systems. Some other interesting works that do not require the temporal independency of the regressors can be found when the signal to be estimated is a constant (Chen & Sayed, 2015a; Piggott & Solo, 2016, 2017; Solo, 2015). In particular, Piggott and Solo (2016) appear to be the first to study the almost sure convergence of the LMS-based distributed algorithms, and Piggott and Solo (2017) establish the second order performance results under temporally strictly stationary and strong mixing assumptions.

To provide stability and performance analyses under more general correlated and non-stationary situations, Chen, Liu, and

☆ This work was supported by the National Natural Science Foundation of China under grants 11688101 and 61227902. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Er-Wei Bai under the direction of Editor Torsten Söderström.

* Corresponding author at: Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China.
E-mail addresses: xiesiyu@amss.ac.cn (S. Xie), lguo@amss.ac.cn (L. Guo).

https://doi.org/10.1016/j.automatica.2018.03.027
0005-1098/

Guo (2014) considered a normalized diffusion LMS algorithm under a cooperative stochastic information condition with a fixed topology, but without independency or stationarity considerations. Their stability results and analyses were later improved (Chen, Liu, & Guo, 2016). In our recent work (Xie & Guo, 2015), we analyzed the stability of normalized consensus LMS algorithm with a fixed topology, under a more general information condition than that used in Chen et al. (2014, 2016), which has also been shown to be necessary for the stability of the consensus algorithm in a certain sense. However, the random matrix product methods (Chen et al., 2014, 2016; Xie & Guo, 2015) for the normalized LMS fail to be applicable to the standard LMS consensus adaptive filtering algorithms, because of the possible unboundedness of the system signals. This is one of the key issues that we have to deal with for the standard LMS filtering algorithms, and a preliminary step was recently made on stability analysis with a fixed network topology (Xie & Guo, 2017).

The main contributions of the paper contain the following three aspects: (i) We will present a weakest possible information condition for the stability of the standard LMS-based consensus adaptive filters under possibly unbounded, non-independent and non-stationary assumptions, which does not exclude applications to stochastic systems with feedback. Stochastic averaging theorems will be used in the stability analysis. (ii) We will show that the whole sensor network can accomplish the estimation task cooperatively, even if none of the sensors can do it individually due to lack of sufficient information, and we will also give a performance analysis for the mean square tracking error matrix under some mild assumptions. (iii) We allow the network topology to change over time and be jointly connected, which is applicable to situations where communication interruptions may happen between sensors.

In the rest of the paper, we will present the consensus adaptive filters based on the standard LMS in Section 2. Some necessary notations, concepts and mathematical definitions are stated in Section 3. The main results and proofs are given in Sections 4 and 5, respectively. Section 6 gives a simulation result and Section 7 concludes the paper and discusses related future problems.

2. Problem formulation

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

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

where y_k^i is the scalar observation of the sensor i at time k, v_k^i is the disturbance or un-modeled dynamics, $\boldsymbol{\varphi}_k^i$ is the $m \times 1$-dimensional stochastic regressor of the sensor i, and $\boldsymbol{\theta}_k$ is an unknown $m \times 1$-dimensional stochastic signals whose variation at time k is denoted by $\boldsymbol{\omega}_k$, i.e.,

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

To estimate the unknown $\{\boldsymbol{\theta}_k, k \ge 0\}$, we consider the standard LMS-based consensus adaptive filter, which is recursively defined at each sensor $i = 1, \dots, n$ as follows

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{k+1}^i = \widehat{\boldsymbol{\theta}}_k^i + \mu \Bigg\{ & \boldsymbol{\varphi}_k^i [y_k^i - (\boldsymbol{\varphi}_k^i)^T \widehat{\boldsymbol{\theta}}_k^i] \\ & - \nu \sum_{l \in \mathcal{N}_{i,k}} a_{li,k} (\widehat{\boldsymbol{\theta}}_k^i - \widehat{\boldsymbol{\theta}}_k^l) \Bigg\}, \quad k \ge 0, \end{aligned} \tag{3}$$

where $\nu \in (0, 1)$ is a weighting constant, $\mu \in (0, 1)$ is the step-size, $\{a_{li,k}\}$ is the adjacency matrix of the sensor network, and $\mathcal{N}_{i,k}$ is the set of neighbors of the sensor i at time k (see the next section for details). We remark that if $\nu = 0$, then the above algorithm reduces to n independent LMS filters, which have been extensively studied in the literature (see e.g., Guo, Ljung, & Wang, 1997; Solo & Kong, 1995; Widrow & Stearns, 1985).

To write the above distributed adaptive filters into a compact form, we introduce the following notations:

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

where $\mathrm{col}\{\cdots\}$ denotes a vector by stacking the specified vectors, $\mathrm{diag}\{\cdots\}$ is used in a non-standard manner which means that $m \times 1$ column vectors are combined "in a diagonal manner" resulting in an $mn \times n$ matrix, $\mathcal{L}_k$ is the Laplacian matrix of the graph at time k (see the next section for details), $\otimes$ is the Kronecker product, and I_m denotes the m-dimensional identity matrix. By (1) and (2), we have

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

and

$$\boldsymbol{\Omega}_{k+1} = \boldsymbol{\Theta}_{k+1} - \boldsymbol{\Theta}_k, \quad k \ge 0.$$

From (3), we obtain that $\forall k \ge 0$,

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

where $\mathcal{L}_k$ is the Laplacian matrix of graph $\mathcal{G}_k$ (see the next section). Let us denote $\widetilde{\boldsymbol{\Theta}}_k = \widehat{\boldsymbol{\Theta}}_k - \boldsymbol{\Theta}_k$ and because $(\mathcal{L}_k \otimes I_m)\boldsymbol{\Theta}_k = 0$, we can get $\forall k \ge 0$,

$$\widetilde{\boldsymbol{\Theta}}_{k+1} = (I_{mn} - \mu \boldsymbol{G}_k)\widetilde{\boldsymbol{\Theta}}_k + \mu \boldsymbol{\Phi}_k \boldsymbol{V}_k - \boldsymbol{\Omega}_{k+1}. \tag{4}$$

Note that by the stochastic internal–external stability results (see Propositions 2.1 and 2.2 in Guo, 1994), we know that the stability of (4) essentially hinges on the exponential stability of the homogeneous part:

$$\widetilde{\boldsymbol{\Theta}}_{k+1} = (I_{mn} - \mu \boldsymbol{G}_k)\widetilde{\boldsymbol{\Theta}}_k. \tag{5}$$

This motivates us to give some definitions on exponential stability in the next section.

3. Notations and definitions

Notations. Let $\mathbb{R}^{m\times n}$ denote the set of $m \times n$ matrices with real entries. For any random matrix $X \in \mathbb{R}^{m\times n}$, its Euclidean norm is defined as its maximum singular value, i.e., $\|X\| = \{\lambda_{max}(XX^T)\}^{\frac{1}{2}}$, where $\lambda_{max}(\cdot)$ denotes the largest eigenvalue of matrix $(\cdot)$ and $(\cdot)^T$ denotes the transpose operator, and its L_p-norm is defined as $\|X\|_{L_p} = \{\mathbb{E}[\|X\|^p]\}^{\frac{1}{p}}$, where $\mathbb{E}[\cdot]$ denotes the expectation operator.

Network topology. Consider a set of n sensors and model it as an undirected weighted graph. Since the relationship between neighbors may change over time, so does the graph describing it. Then we have a class of graphs $\mathcal{G}_k$ on the n vertexes at time $k\,(k \ge 0)$ composed of $\{\mathcal{V}, \mathcal{E}_k, \mathcal{A}_k\}$, where $\mathcal{V} = \{1, 2, \dots, n\}$ is the vertex set, $\mathcal{E}_k \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set at time k, and $\mathcal{A}_k$ is a matrix reflects the interaction strength among neighboring vertexes at time k. Define $\mathcal{A}_k = \{a_{ij,k}\}_{n\times n}, i, j = 1, \dots, n, k \ge 0$, which is called the weighted adjacency matrix with $a_{ij,k} \ge 0$, $\sum_{j=1}^n a_{ij,k} = 1, \forall i = 1, \dots, n, k \ge 0$. Since the graph $\mathcal{G}_k$ is undirected, we have $a_{ij,k} = a_{ji,k}$. Vertex i denotes the ith sensor and (i, j) denotes the connection from sensor

i to sensor j. Note that $(i,j) \in \mathcal{E}_k \Leftrightarrow a_{ij,k} > 0$. The set of neighbors of the sensor i at time k is denoted as $\mathcal{N}_{i,k} = \{l \in \mathcal{V} | (l,i) \in \mathcal{E}_k\}$. The Laplacian matrix $\mathcal{L}_k$ of the graph $\mathcal{G}_k$ at time k is defined as $\mathcal{L}_k = I_n - \mathcal{A}_k$. By the union of a collection of graphs $\{\mathcal{G}_{k+1}, \ldots, \mathcal{G}_{k+q}\}$, is meant the graph with the vertex set $\mathcal{V}$ and the edge set equaling to the union of the edge sets of all the graphs in the collection. We say that such a collection of the graphs is jointly connected if the union of the members is a connected graph.

To proceed with further discussions, we also need the following definitions (Guo & 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

$$S_p(\mu^*) = \Bigg\{ A : \Bigg\| \prod_{j=i+1}^{k} (I - \mu A_j) \Bigg\|_{L_p} \leq M(1-\mu\alpha)^{k-i}, \quad \forall k \geq i+1, \forall i \geq 0, \forall \mu \in (0,\mu^*), \text{ for some } M > 0 \text{ and } \alpha \in (0,1) \Bigg\}. \tag{6}$$

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

$$S(\mu^*) = \Bigg\{ A : \Bigg\| \prod_{j=i+1}^{k} (I - \mu \mathbb{E}[A_j]) \Bigg\| \leq M(1-\mu\alpha)^{k-i}, \quad \forall k \geq i+1, \forall i \geq 0, \forall \mu \in (0,\mu^*), \text{ for some } M > 0 \text{ and } \alpha \in (0,1) \Bigg\}. \tag{7}$$

Denote $S_p \triangleq \bigcup_{\mu^* \in (0,1)} S_p(\mu^*)$ and $S \triangleq \bigcup_{\mu^* \in (0,1)} S(\mu^*)$. In the following two sections, we will show that for a quite general class of $\{\boldsymbol{G}_k, k \geq 0\}$, the L_p-exponential stability of the homogeneous part of Eq. (4) is equivalent to the averaged deterministic exponential stability under some general conditions.

Definition 3. A random sequence $x = \{x_k, k \geq 0\}$ is called an element of the weakly dependent class $\mathcal{D}_p (p \geq 1)$, if there exists a constant C_p^x only depending on p and the distribution of $\{x_k, k \geq 0\}$ such that

$$\sup_k \Bigg\| \sum_{i=k+1}^{k+h} x_i \Bigg\|_{L_p} \leq C_p^x h^{\frac{1}{2}}, \quad \forall h \geq 1. \tag{8}$$

It is known that the martingale difference, zero mean ϕ- and α-mixing sequences, and the linear process driven by white noises are all in set $\mathcal{D}_p$.

Definition 4. A random process $\{\boldsymbol{\varepsilon}_k, k \geq 0\}$ is called an element of $\mathcal{B}(\alpha)$ class with $\alpha > 0$, if there exist constants $M > 0$ and $K > 0$ such that for any integer $s \geq 1$ and any integer sequence $j_1 < j_2 < \cdots < j_s$,

$$\mathbb{E}\Bigg[\exp\Bigg(\alpha \sum_{t=1}^{s} \|\boldsymbol{\varepsilon}_{j_t}\|^2\Bigg)\Bigg] \leq M \exp(Ks). \tag{9}$$

Note that the inequality (9) can easily be verified for any bounded signals or independent Gaussian signals.

Definition 5. Let $\{A_k, k \geq 0\}$ be a matrix sequence and $\{b_k, k \geq 0\}$ be a positive scalar sequence. Then by $A_k = O(b_k)$ we mean that there exists a constant $M > 0$ such that $\|A_k\| \leq M b_k$, $\forall k \geq 0$.

4. The main results

4.1. Stability results

In this section, we first study the L_p-exponential stability of the homogeneous part of Eq. (4), then give the upper bound of the tracking error $\widetilde{\Theta}_{k+1}$ in the next section. Let us consider the following decomposition of the regressor processes $\{\boldsymbol{\varphi}_k^i, k \geq 0\} (i = 1, \ldots, n)$:

$$\boldsymbol{\varphi}_k^i = \sum_{j=-\infty}^{\infty} A^i(k,j) \boldsymbol{\varepsilon}_{k-j}^i + \boldsymbol{\xi}_k^i,$$

$$\sum_{j=-\infty}^{\infty} \sup_k \|A^i(k,j)\| < \infty, \quad i = 1, \ldots, n, \tag{10}$$

where $\{\boldsymbol{\xi}_k^i, k \geq 0\} (i = 1, \ldots, n)$ are $m \times 1$-dimensional bounded deterministic processes and $\{\boldsymbol{\varepsilon}_k^i, k \geq 0\} (i = 1, \ldots, n)$ are general $d \times 1$-dimensional ϕ-mixing sequences satisfying $\{\boldsymbol{\varepsilon}_k^i, k \geq 0\} \in \mathcal{B}(\alpha)$ for some $\alpha > 0$ and for any $i = 1, \ldots, n$. The weighting matrices $A^i(k,j) \in \mathbb{R}^{m \times d} (i = 1, \ldots, n)$ are assumed to be deterministic. We remark that (10) has a similar form with the well-known Wold decomposition and includes time-varying linear state space models. The signal process generated by (10) may neither be a stationary process nor a ϕ-mixing process in general.

We need the following assumptions in the main results.

Assumption 6 (*Network Topology*)**.** There exists an integer $q > 0$ such that for all $k \geq 0$, the union of collection of graphs $\{\mathcal{G}_{k+1}, \ldots, \mathcal{G}_{k+q}\}$ is jointly connected. The non-zero entries of the adjacency matrices $\mathcal{A}_k = \{a_{ij,k}\}_{n \times n} (\forall k \geq 0, \forall i,j = 1, \ldots, n)$ have the following uniform lower bound a:

$$\min_{(i,j): a_{ij,k} > 0} a_{ij,k} \geq a > 0, \quad \forall k \geq 0.$$

Assumption 7 (*Information Condition*)**.** There exist an integer $h > 0$ and a constant $\delta > 0$ such that

$$\sum_{i=1}^{n} \sum_{j=k+1}^{k+h} \mathbb{E}[\boldsymbol{\varphi}_j^i (\boldsymbol{\varphi}_j^i)^T] \geq \delta I_m, \quad \forall k \geq 0. \tag{11}$$

Remark 8. When $n = 1$, the sensor network degenerates to a single sensor and the information condition becomes the weakest possible information condition (Guo et al., 1997). This "persistence of excitation" or "full rank" condition on $\boldsymbol{\varphi}_j^i$ is the information condition required to track unknown signals that may change constantly. When $n > 1$, Assumption 7 is a kind of cooperative information condition, which enables the distributed filters to track the unknown signals, even in the case where any individual filter cannot.

We are now in a position to state the first main result of this paper, and the detailed proof of Theorem 9 is given in Section 5.

Theorem 9. *Consider the error* Eq. (4) *of the distributed adaptive filter. Suppose that the regressor processes* $\{\boldsymbol{\varphi}_k^i, k \geq 0\} (i = 1, \ldots, n)$ *are generated by* (10) *and that* Assumption 6 *is satisfied. Then for any* $p \geq 1$, *there exists a constant* $\mu^* \in (0,1)$ *such that for all* $\nu \in (0,1)$, *the homogeneous equation* (5) *of the filter is* L_p*-exponentially stable (i.e.,* $\{\boldsymbol{G}_k, k \geq 0\} \in S_p(\mu^*)$*), if and only if* Assumption 7 *holds.*

As we explained at the end of Section 2, the stability of the filtering equation (4) hinges on the exponential stability of the homogeneous equation (5). The above theorem naturally gives conditions on stability of the distributed adaptive filters, based on which we will further examine the tracking error bounds in the sequel.

4.2. Performance results

We will first establish the tracking error bound under a "worst case" situation, i.e., both the noises and the parameter variations are only assumed to be L_r-bounded.

Assumption 10. There exists $r > 2$ such that $\sigma \triangleq \sup_k \|\Phi_k V_k\|_{L_r} < \infty$, and $\gamma \triangleq \sup_k \|\Omega_k\|_{L_r} < \infty$.

Theorem 11. *Consider the standard LMS-based distributed algorithm (4) and suppose that Assumptions 6, 7 and 10 are satisfied. Let the signal processes $\{\varphi_k^i, k \geq 0\}(i = 1, \ldots, n)$ be defined by (10). Then for all $k > 0$, $\nu \in (0, 1)$ and all small $\mu > 0$,*

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

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

By Theorem 9 and since the proof of Theorem 11 is similar to Theorem 2 in Xie and Guo (2017), here we omit it. We remark that when both the observation noises v_k^i and parameter variation ω_k are small in the "L_r sense", the tracking error will also be small in the L_2 sense. With additional assumptions, we can further establish a better upper bound for the distributed LMS error equation. This assumption simply implies that both the noises and parameter variations are weakly dependent.

Assumption 12. For some $r > 2$, $\{\Phi_k V_k, k \geq 0\} \in \mathcal{D}_r$ and $\{\Omega_k, k \geq 0\} \in \mathcal{D}_r$, where $\mathcal{D}_r$ is defined by Definition 3.

Theorem 13. *Consider the standard LMS-based distributed algorithm (4) and suppose that Assumptions 6, 7 and 12 are satisfied. Let the signal processes $\{\varphi_k^i, k \geq 0\}(i = 1, \ldots, n)$ be defined by (10). Then for all $k > 0$, $\nu \in (0, 1)$ and all small $\mu > 0$,*

$$\|\widetilde{\Theta}_k\|_{L_2} = O\left(C_r^{\Phi V}\sqrt{\mu} + \frac{C_r^{\Omega}}{\sqrt{\mu}}\right) + O([1 - \mu\beta]^k),$$

where $\beta \in (0, 1)$ is a constant and $C_r^{\Phi V}$, C_r^{Ω} are two constants defined as in Definition 3.

By Theorem 9 and since the proof of Theorem 13 is similar to Theorem 4 in Guo et al. (1997), here we omit it. Note that the upper bound in Theorem 13 significantly improves the "crude" bound given in Theorem 11 and it shows the tradeoff between noise sensitivity and tracking ability.

Based on Theorem 9, we can further establish the performance of the mean square tracking error matrix in the following part. Let $\widehat{\Pi}_k$ be defined by the following linear deterministic difference equation

$$\widehat{\Pi}_{k+1} = (I_{mn} - \mu\mathbb{E}[G_k])\widehat{\Pi}_k(I_{mn} - \mu\mathbb{E}[G_k])^T + \mu^2\mathbb{E}[S_k] + Q_\omega(k+1), \quad k \geq 0, \tag{12}$$

where $S_k = \Phi_k V_k V_k^T \Phi_k^T$, $Q_\omega(k+1) = \mathbb{E}[\Omega_{k+1}\Omega_{k+1}^T]$ and $\widehat{\Pi}_0 = \mathbb{E}[\widetilde{\Theta}_0\widetilde{\Theta}_0^T]$. Next we will show that the deterministic matrix (12) provides a good approximation of the true mean square tracking error matrix $\Pi_k = \mathbb{E}[\widetilde{\Theta}_k\widetilde{\Theta}_k^T]$ when the step-size is small enough. Note that $\widehat{\Pi}_k$ obeys a simple linear difference equation and can easily be calculated and examined. We then have the performance result under the following assumption and $\mathcal{F}_k = \sigma\{\cdot\}$ denotes the σ-algebra, where the definition of σ-algebra can be found in Chow and Teicher (1978).

Assumption 14. Denote $\mathcal{F}_k = \sigma\{\varphi_j^i, \omega_j, v_{j-1}^i, i = 1, \ldots, n, j \leq k\}$, then for all $k \geq 1$, $i = 1, \ldots, n$,

$$\mathbb{E}[V_k|\mathcal{F}_k] = 0, \quad \mathbb{E}[\Omega_{k+1}|\mathcal{F}_k] = 0, \quad \mathbb{E}[\Omega_{k+1}V_k|\mathcal{F}_k] = 0,$$
$$\mathbb{E}[V_k V_k^T|\mathcal{F}_k] = P_v(k) \geq 0, \quad \mathbb{E}[\Omega_k\Omega_k^T] = Q_\omega(k) \geq 0,$$
$$\sup_k \mathbb{E}[\|V_k\|^r|\mathcal{F}_k] \leq M, \quad \gamma \triangleq \sup_k \|\Omega_k\|_{L_r} < \infty, \tag{13}$$

where $r > 2$ and $M > 0$ are constants.

Assumption 14 means that the measurement noise V_k and the parameter variation Ω_{k+1} have white noise characters, which is a worst case analysis since the future behavior of the model is unpredictable. This assumption also means that the observation noises and the parameter variations are uncorrelated given the past signals, but spatial correlations of the noises are allowed.

Theorem 15. *Consider the standard LMS-based distributed algorithm (4) and suppose that Assumptions 6, 7 and 14 are satisfied. Let the signal processes $\{\varphi_k^i, k \geq 0\}(i = 1, \ldots, n)$ be defined by (10), and $\{E_k, \Omega_k, V_{k-1}, k \geq 1\}$ be a ϕ-mixing process where $E_k = col\{\varepsilon_k^1, \ldots, \varepsilon_k^n\}$. Then, the tracking error covariance matrix has the following expansion for all $k > 0$, $\nu \in (0, 1)$ and all small $\mu > 0$:*

$$\Pi_k = \widehat{\Pi}_k + O\left(\bar{\delta}(\mu)\left[\mu + \frac{\gamma^2}{\mu} + (1 - \beta\mu)^k\right]\right), \tag{14}$$

where $\beta \in (0, 1)$ is a constant and the function $\bar{\delta}(\mu) \to 0$ as $\mu \to 0$.

Since the proof of Theorem 15 is similar to Theorem 5 in Guo et al. (1997), here we omit it. Theorem 15 provides a good approximation of the true mean square tracking error matrix Π_k by $\widehat{\Pi}_k$, and $\bar{\delta}(\mu)$ plays an important role, since the faster it tends to 0, the better approximation will be. To simplify $\widehat{\Pi}_k$, we consider the wide-sense stationary case in the next theorem.

Theorem 16. *Let*

$$F = \mathbb{E}[F_k] = diag\{F^1, \ldots, F^n\},$$
$$\mathcal{L} = \mathcal{L}_k, G = F + \nu(\mathcal{L} \otimes I_m),$$
$$S = \mathbb{E}[\Phi_k V_k V_k^T \Phi_k^T], Q \equiv Q_\omega(k).$$

Under the conditions of Theorem 15, there exists a constant $\mu^* \in (0, 1)$ such that for all $\mu \in (0, \mu^*]$, $\nu \in (0, 1)$ and $k > 0$,

$$\Pi_k = \mu\bar{R}_v + \frac{\bar{R}_\omega}{\mu} + O\left(\bar{\delta}(\mu)\left[\mu + \frac{\gamma^2}{\mu}\right]\right) + o(1),$$

where

$$\bar{R}_v = \int_0^\infty e^{-Gt}Se^{-Gt}dt, \quad \bar{R}_\omega = \int_0^\infty e^{-Gt}Qe^{-Gt}dt.$$

Remark 17. Since the proof of Theorem 16 is similar to Corollary 4.2 in Guo and Ljung (1995b), here we omit it. Note that $\lim_{\mu\to 0}\bar{\delta}(\mu) = 0$. As a result, we have

$$\Pi_k \sim \mu\bar{R}_v + \frac{\bar{R}_\omega}{\mu}, \quad (k \to \infty, \mu \to 0).$$

To minimize the tracking error, we may take $\mu^0 = \sqrt{Tr(\bar{R}_\omega)/Tr(\bar{R}_v)}$, where $Tr(\cdot)$ denotes the trace operator, which gives

$$\sum_{i=1}^n \mathbb{E}[\|\widetilde{\theta}_k^i\|^2] \sim 2\sqrt{Tr(\bar{R}_\omega) \cdot Tr(\bar{R}_v)}, k \to \infty, \mu \to 0.$$

5. Proof of Theorem 9

The proof consists of a series of lemmas.

Lemma 18 (Wang, Liu, & Guo, 2007). *Suppose that Assumption 6 is satisfied, then we have*

$$\inf_{k\geq 0}\lambda_2\Big(\sum_{j=k+1}^{k+q}\mathcal{L}_j\Big) \neq 0,$$

where λ_2 is the second smallest eigenvalue of the matrix.

Lemma 19. *Let $\{\varphi_k^i, k \geq 0\}(i = 1, \ldots, n)$ belong to $\mathcal{B}(\epsilon)$ class, and denote $F_k^i = \varphi_k^i(\varphi_k^i)^T$, $F_k = diag\{F_k^1, \ldots, F_k^n\}$, $G_k = F_k + \nu(\mathcal{L}_k \otimes I_m)$. Then*

$$\{G_k, k \geq 0\} \in S \Longrightarrow \{G_k, k \geq 0\} \in S_p, \forall p \geq 1,$$

provided that the following condition holds: 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 $s \geq t \geq 0$,

$$\mathbb{E}\left[\exp\left(\mu \sum_{j=t+1}^{s} \|S_j^{(T)(i)}\|\right)\right] \leq M \cdot \exp\{[\mu g(T) + o(\mu)](s-t)\}, \quad i = 1, \ldots, n, \tag{15}$$

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

Proof. 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

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\epsilon' \sum_{t=1}^{s} \|G_{j_t}\|\right)\right] \\
&\leq \mathbb{E}\left[\exp\left(\epsilon' \sum_{t=1}^{s} \|F_{j_t}\|\right)\right] \cdot \exp(2\epsilon' s\nu) \\
&= \mathbb{E}\left[\exp\left(\epsilon' \sum_{t=1}^{s} \max_{i=1,\ldots,n} \|F_{j_t}^i\|\right)\right] \cdot \exp(2\epsilon' s\nu) \\
&\leq \mathbb{E}\left[\exp\left(\epsilon' \sum_{i=1}^{n}\sum_{t=1}^{s} \|F_{j_t}^i\|\right)\right] \cdot \exp(2\epsilon' s\nu) \\
&\leq \left\{\prod_{i=1}^{n} \mathbb{E}\left[\exp(n\epsilon' \sum_{t=1}^{s} \|F_{j_t}^i\|)\right]\right\}^{\frac{1}{n}} \cdot \exp(2\epsilon' s\nu) \\
&\leq M \exp(Ks) \cdot \exp(2\epsilon' s\nu) \\
&= M \exp(K's),
\end{aligned} \tag{16}$$

for some constants $M > 0, K > 0$, where we have used the assumption that $\{\varphi_k^i, k \geq 0\} \in \mathcal{B}(\epsilon)$. Next, denote $P_j^{(T)} = \sum_{t=jT}^{(j+1)T-1}(G_t - \mathbb{E}[G_t])$, we then have

$$\begin{aligned}
\|P_j^{(T)}\|_{L_2} &= \left\| \sum_{t=jT}^{(j+1)T-1} (G_t - \mathbb{E}[G_t]) \right\|_{L_2} \\
&= \left\| \sum_{t=jT}^{(j+1)T-1} (F_t - \mathbb{E}[F_t]) \right\|_{L_2} \\
&= \left\| \begin{matrix} S_j^{(T)(1)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & S_j^{(T)(n)} \end{matrix} \right\|_{L_2} \\
&= \max_{i=1,\ldots,n} \|S_j^{(T)(i)}\|_{L_2}.
\end{aligned} \tag{17}$$

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

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\mu' \sum_{j=t+1}^{s} \|P_j^{(T)}\|\right)\right] \\
&\leq \mathbb{E}\left[\exp\left(\mu' \sum_{j=t+1}^{s} \max_{i=1,\ldots,n} \|S_j^{(T)(i)}\|\right)\right] \\
&\leq \mathbb{E}\left[\exp\left(\mu' \sum_{i=1}^{n}\sum_{j=t+1}^{s} \|S_j^{(T)(i)}\|\right)\right] \\
&\leq \left\{\prod_{i=1}^{n} \mathbb{E}\left[\exp(n\mu' \sum_{j=t+1}^{s} \|S_j^{(T)(i)}\|)\right]\right\}^{\frac{1}{n}} \\
&\leq M \exp\{[\mu g(T) + o(\mu)](s-t)\} \\
&= M \exp\{[\mu' g'(T) + o(\mu')](s-t)\},
\end{aligned} \tag{18}$$

where we have used (15). Here, by Theorem 1 in Guo et al. (1997) we know that Lemma 19 is true.

Lemma 20. *Let $\max_{i=1,\ldots,n}\sup_k\|\varphi_k^i\|_{L_2} < \infty$ and Assumptions 6 and 7 hold. Then we have*

$$\{G_k, k \geq 0\} \in S.$$

Proof. Take

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

then we have $0 \leq \mu\mathbb{E}[G_k] \leq I_{mn}$, for $\mu \in (0, \mu^*)$. Denote $s = \max\{h, q\}$ and by Assumption 7, we have $\sum_{i=1}^{n}\sum_{j=k+1}^{k+s}\mathbb{E}[\varphi_j^i(\varphi_j^i)^T] \geq \delta I_m$.

By Lemma 18, we know that matrix $\sum_{j=k+1}^{k+s}\mathcal{L}_j$ has only one zero eigenvalue. Then we apply Theorem 1 in Xie and Guo (2015) to the deterministic sequence $\mu\mathbb{E}[G_k]$ by replacing matrix $\mathcal{L}$ with $\frac{1}{s}\sum_{j=k+1}^{k+s}\mathcal{L}_j$, it is easy to see that $\{G_k, k \geq 0\} \in S(\mu^*)$ holds.

We then present the "converse lemma" of Lemma 19. Before that, we define the following set. Let $p \geq 0, A = \{A_k, k \geq 0\}$. Set $S_j^{(T)} = \sum_{t=jT}^{(j+1)T-1}(A_t - \mathbb{E}[A_t])$, and define the following weakly dependent class called $\mathcal{M}_p$-class

$$\mathcal{M}_p = \left\{A : \sup_j \|S_j^{(T)}\|_{L_p} = o(T), \quad as \ T \to \infty\right\}. \tag{19}$$

Lemma 21. *Let $\{F_k^i, k \geq 0\}(i = 1, \ldots, n)$ be n random matrix processes all belonging to $\mathcal{M}_2$. Then*

$$\{G_k, k \geq 0\} \in S_2 \Longrightarrow \{G_k, k \geq 0\} \in S.$$

Proof. By Theorem 3.1 in Guo and Ljung (1995a), we only need to prove that $\{G_k, k \geq 0\} \in \mathcal{M}_2$. Since $\{F_k^i, k \geq 0\} \in \mathcal{M}_2(i = 1, \ldots, n)$, we have $\forall i = 1, \ldots, n$,

$$\begin{aligned}
\sup_j \|S_j^{(T)(i)}\|_{L_2} &= \sup_j \left\| \sum_{t=jT}^{(j+1)T-1} (F_t^i - \mathbb{E}[F_t^i]) \right\|_{L_2} \\
&= o(T), \quad as \ T \to \infty,
\end{aligned}$$

and consequently by (17)

$$\begin{aligned}
\sup_j \|P_j^{(T)}\|_{L_2} &= \sup_j \max_{i=1,\ldots,n} \|S_j^{(T)(i)}\|_{L_2} \\
&= o(T), \quad as \ T \to \infty,
\end{aligned}$$

which implies that $\{G_k, k \geq 0\} \in \mathcal{M}_2$. This completes the proof.

Lemma 22. *Let $\max_{i=1,\ldots,n}\sup_k\|\varphi_k^i\|_{L_2} < \infty$ and Assumption 6 be satisfied. Then $\{G_k, k \geq 0\} \in S(\mu^*)$ implies that there exists an integer $h > 0$ such that*

$$\inf_k \lambda_{min}\left\{\sum_{i=1}^{n}\sum_{j=k+1}^{k+h} \mathbb{E}[\varphi_j^i(\varphi_j^i)^T]\right\} > 0. \tag{20}$$

Proof. The detailed proof is in Appendix.

Now, we are able to complete the proof of Theorem 9.

Sufficiency: If Assumptions 6 and 7 hold, then by Lemma 20, we have $\{G_k, k \geq 0\} \in S$. By Lemmas 4 and 6 in Guo et al. (1997), we know that for any $i = 1, \ldots, n$, $\{F_k^i, k \geq 0\}$ satisfies the conditions of Lemma 19. Then, Lemma 19 is applicable, and consequently $\{\boldsymbol{G}_k, k \geq 0\} \in S_p, \forall p \geq 1$ holds, which implies that the homogeneous equation (5) is L_p-exponentially stable.

Necessity: By Lemma 2 in Guo et al. (1997), we know that $\{\boldsymbol{F}_k^i, k \geq 0\} \in \mathcal{M}_2 (i = 1, \ldots, n)$. Let $\{\boldsymbol{G}_k, k \geq 0\} \in S_2$, then by applying Lemma 21, it is obvious that $\{\boldsymbol{G}_k, k \geq 0\} \in S$. Consequently, by Lemma 22 we know that Assumption 7 holds. This completes the proof of Theorem 9.

6. Simulation results

To illustrate the effectiveness of the cooperative information condition, let us take $n = 3$ with the following adjacency matrices, for $k \geq 0$,

$$\mathcal{A}_{2k} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathcal{A}_{2k+1} = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix},$$

then the corresponding graphs are jointly connected with $q = 2$. We will estimate or track an unknown 3-dimensional signal $\boldsymbol{\theta}_k$, and consider $\gamma = 0$ ($\boldsymbol{\theta}_k$ is time-invariant) and $\gamma = 0.1$ ($\boldsymbol{\theta}_k$ is time-varying) respectively, and the parameter variation is assigned to be $\boldsymbol{\omega}_k \sim N(0, 0.3, 3, 1)$ (Gaussian distribution) in (2). The observation noises $\{v_k^i, k \geq 1, i = 1, 2, 3\}$ are temporally and spatially independently and identically distributed with $v_k^i \sim N(0, 0.3)$ in (1). Moreover, we assume that $\boldsymbol{\varphi}_k^i (i = 1, 2, 3)$ are generated by a state space model

$$\begin{cases} \boldsymbol{x}_k^i = & A_i \boldsymbol{x}_{k-1}^i + B_i \xi_k^i, \\ \boldsymbol{\varphi}_k^i = & C_i \boldsymbol{x}_k^i, \end{cases}$$

where $\{\xi_k^i, k \geq 1, i = 1, 2, 3\}$ are temporally and spatially independently and identically distributed with $\xi_k^i \sim N(0, 0.3)$, and

$$A_1 = A_2 = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/5 \end{pmatrix}, A_3 = \begin{pmatrix} 4/5 & 0 & 0 \\ 4/5 & 0 & 0 \\ 4/5 & 0 & 0 \end{pmatrix},$$

$$B_1 = B_2 = B_3 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}^T,$$

$$C_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, C_2 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, C_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It can be verified that Assumptions 6 and 7 are satisfied with $q = 2$ and $h = 2$. However, the necessary information assumption in Guo et al. (1997) is not satisfied for any individual sensor. Let $\boldsymbol{x}_0^1 = \boldsymbol{x}_0^2 = \boldsymbol{x}_0^3 = (1, 1, 1)^T, \boldsymbol{\theta}_0 = (1, 1, 1)^T, \hat{\boldsymbol{\theta}}_0^i = (0, 0, 0)^T (i = 1, 2, 3), \mu = 0.3, \nu = 0.5$. Here we repeat the simulation for $m = 500$ times with the same initial states. Then for sensor $i (i = 1, 2, 3)$, we can get m sequences $\{\|\hat{\boldsymbol{\theta}}_k^{i,j} - \boldsymbol{\theta}_k^j\|^2, k = 1, 100, 200, \ldots, 2000\} (j = 1, \ldots, m)$, where the superscript j denotes the jth simulation result. We use $\frac{1}{m}\sum_{j=1}^m \|\hat{\boldsymbol{\theta}}_k^{i,j} - \boldsymbol{\theta}_k^j\|^2 (i = 1, 2, 3, k = 1, 100, 200, \ldots, 2000)$ to approximate the estimation or tracking errors with $\gamma = 0$ and $\gamma = 0.1$.

When $\boldsymbol{\theta}_k$ is time-invariant, the upper one of Fig. 1 is the standard non-cooperative LMS algorithm in which the estimation errors of the three sensors are all very large, because no sensor satisfies the information condition in Guo et al. (1997) and only partial information of the unknown parameter is observed. The lower one is the standard LMS-based distributed algorithm in which all the estimation errors converge to a small neighborhood of zero as k increases, since the whole system satisfies Assumption 7. In Fig. 2, $\boldsymbol{\theta}_k$ is time-varying. The upper one is the individual situation in which the tracking errors of the three sensors diverge as k increases, while the lower one is the standard LMS-based distributed algorithm in which all the tracking errors have a nice upper bound.

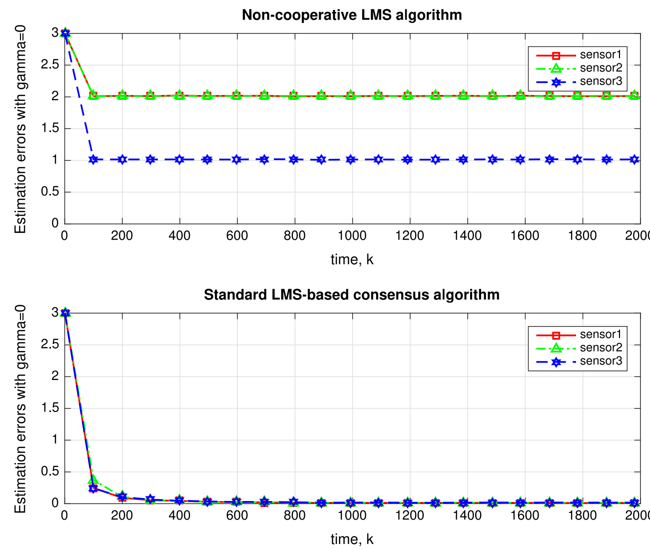

Fig. 1. Estimation errors of the three sensors with $\gamma = 0$.

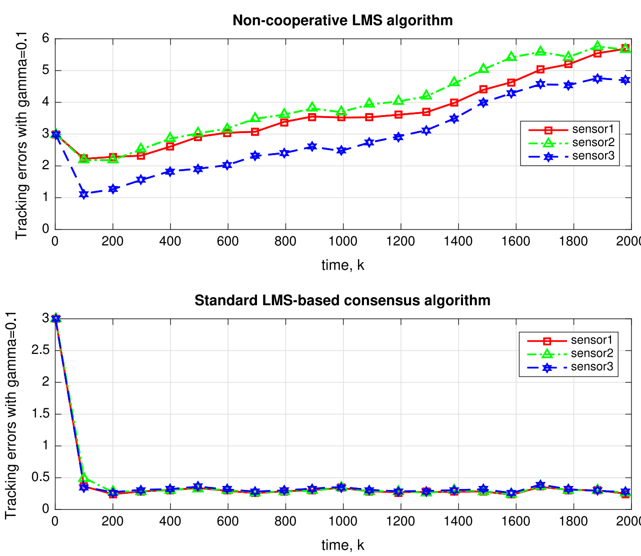

Fig. 2. Tracking errors of the three sensors with $\gamma = 0.1$.

7. Concluding remarks

In this paper we have provided a theory on the standard LMS-based consensus adaptive filters, by resorting to stochastic averaging theorems instead of using the methods based on product of random matrices in the normalized LMS situations. The main result Theorem 9 shows that Assumption 7 is the weakest possible information condition for the exponential stability of the consensus adaptive filters for a large class of non-stationary weakly dependent and possibly unbounded signals including possible feedback signals, whenever the step-size is suitably small. In addition, our theory shows that such consensus algorithms can fulfill the estimation task by working in a collaborative manner, even when any individual sensor cannot. Furthermore, we have also presented a performance analysis for the tracking error covariance matrix, which can be well approximated by a simple deterministic difference matrix equation. There are many other interesting problems for further research. It will be interesting to study the stability analysis of the RLS-based and Kalman filtering-based distributed algorithms and to combine distributed adaptive filtering with distributed control problems.

Appendix

Proof of Lemma 22. Take $\mu^* = \frac{1}{1+\max_{i=1,\dots,n}\sup_k \mathbb{E}\|\varphi_k^i\|^2}$, then we have $0 < \mu\mathbb{E}[\boldsymbol{G}_k] < I_{mn}$ for any $\mu \in (0, \mu^*]$. By Theorem 2.2 in Guo (1994), there exists an integer $h \geq q > 0$ such that $\inf_k \rho_k \triangleq \inf_k \lambda_{min}\{\sum_{j=k+1}^{k+h} \mu\mathbb{E}[\boldsymbol{G}_k]\} > 0$, where $\lambda_{min}(\cdot)$ denotes the smallest eigenvalue of matrix $(\cdot)$. We proceed to show that there exists a positive constant δ such that

$$\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=k+1}^{k+h} \boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T\right] \geq \delta I_m,$$

for all $k \geq 0$. We first prove that for any $k \geq 0$, the eigenvalues of matrices $\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=k+1}^{k+h}\{\boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T\}\right]$ are all positive. This can be done through contradiction by assuming that there exists a time instant k^* such that the smallest eigenvalue of matrix $\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=k+1}^{k+h}\{\boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T\}\right]$ is 0. Denote the unit eigenvector of 0 is $\boldsymbol{y}_{k^*} \in \mathbb{R}^{m\times 1}$, then we have

$$\boldsymbol{y}_{k^*}^T\left(\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=k^*+1}^{k^*+h} \boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T\right]\right)\boldsymbol{y}_{k^*} = 0. \tag{21}$$

Now, let $\boldsymbol{H}_{k^*}^i = \sum_{j=k^*+1}^{k^*+h}[\boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T]$, and $\boldsymbol{H}_{k^*} = \mathrm{diag}\{\boldsymbol{H}_{k^*}^1, \dots, \boldsymbol{H}_{k^*}^n\}$. By the definition of $\boldsymbol{G}_j$, we have

$$\boldsymbol{\Delta}_{k^*} \triangleq \sum_{j=k^*+1}^{k^*+h}\mathbb{E}[\boldsymbol{G}_j] = \mathbb{E}[\boldsymbol{H}_{k^*}] + \nu\left(\sum_{j=k^*+1}^{k^*+h}\mathcal{L}_j\right)\otimes I_m. \tag{22}$$

Note also that

$$\boldsymbol{\Gamma}_{k^*} \triangleq \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=k^*+1}^{k^*+h}\boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T\right] = \mathbb{E}\left[\sum_{i=1}^{n}\boldsymbol{H}_{k^*}^i\right]. \tag{23}$$

Denote $\mathscr{L}_{k^*} \triangleq \frac{1}{h}\sum_{j=k^*+1}^{k^*+h}\mathcal{L}_j$, we have $\boldsymbol{\Delta}_{k^*} = \mathbb{E}[\boldsymbol{H}_{k^*}] + h\nu(\mathscr{L}_{k^*}\otimes I_m)$. Since $h \geq q > 0$ and by Lemma 18, we know that graphs $\{\mathcal{G}_{k^*+1}, \dots, \mathcal{G}_{k^*+h}\}$ are jointly connected. Then the matrix $\mathscr{L}_{k^*}$ has only one zero eigenvalue whose unit eigenvector is $\frac{1}{\sqrt{n}}(1,\dots,1)^T$, i.e., $\frac{1}{\sqrt{n}}\mathbf{1}$ where $\mathbf{1} = (1,\dots,1)_{n\times 1}^T$. Correspondingly, $\mathscr{L}_{k^*}\otimes I_m$ has m zero eigenvalues whose orthogonal unit eigenvectors are $\boldsymbol{\xi}_1 = \frac{1}{\sqrt{n}}\mathbf{1}\otimes\boldsymbol{e}_1, \dots, \boldsymbol{\xi}_m = \frac{1}{\sqrt{n}}\mathbf{1}\otimes\boldsymbol{e}_m$, where $\boldsymbol{e}_i$ is a unit column vector with the ith element is 1.

The other eigenvalues of $\mathscr{L}_{k^*}\otimes I_m$ are $l_{m+1}, \dots, l_{mn}$ arranged in a non-decreasing order whose orthogonal unit eigenvectors are denoted as $\boldsymbol{\xi}_{m+1}, \dots, \boldsymbol{\xi}_{mn}$ correspondingly. Here, for an arbitrary unit vector $\boldsymbol{\eta} \in \mathbb{R}^{mn\times 1}$, it can be expressed as $\boldsymbol{\eta} = \sum_{j=1}^{m}x_j\boldsymbol{\xi}_j + \sum_{j=m+1}^{mn}x_j\boldsymbol{\xi}_j \triangleq \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2$, where $\sum_{j=1}^{mn}x_j^2 = 1$.

By (22), let us consider the following quadratic form

$$\begin{aligned}
&\boldsymbol{\eta}^T\boldsymbol{\Delta}_{k^*}\boldsymbol{\eta}\\
&= (\boldsymbol{\eta}_1+\boldsymbol{\eta}_2)^T\boldsymbol{\Delta}_{k^*}(\boldsymbol{\eta}_1+\boldsymbol{\eta}_2)\\
&= \boldsymbol{\eta}_1^T\mathbb{E}[\boldsymbol{H}_{k^*}]\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2^T\mathbb{E}[\boldsymbol{H}_{k^*}]\boldsymbol{\eta}_2 + 2\boldsymbol{\eta}_1^T\mathbb{E}[\boldsymbol{H}_{k^*}]\boldsymbol{\eta}_2\\
&\quad + \boldsymbol{\eta}_1^T\left[h\nu(\mathscr{L}_{k^*}\otimes I_m)\right]\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2^T\left[h\nu(\mathscr{L}_{k^*}\otimes I_m)\right]\boldsymbol{\eta}_2\\
&\quad + 2\boldsymbol{\eta}_1^T\left[h\nu(\mathscr{L}_{k^*}\otimes I_m)\right]\boldsymbol{\eta}_2\\
&\triangleq c_1^{k^*} + c_2^{k^*} + c_3^{k^*} + c_4^{k^*} + c_5^{k^*} + c_6^{k^*}.
\end{aligned} \tag{24}$$

Next, following a similar proof as that in Appendix B of Xie and Guo (2018), we estimate the right hand side of (24) term by term.

Firstly, since $\boldsymbol{\eta}_1 = \sum_{j=1}^{m}x_j\boldsymbol{\xi}_j$ and $\boldsymbol{\xi}_j (1 \leq j \leq m)$ is the eigenvector corresponding to the zero eigenvalue of $\mathscr{L}_{k^*}\otimes I_m$, we have $c_4^{k^*} = c_6^{k^*} = 0$.

Secondly, we know that for any $\delta > 0$,

$$\begin{aligned}
c_3^{k^*} &= 2\boldsymbol{\eta}_1^T\mathbb{E}[\boldsymbol{H}_{k^*}]\boldsymbol{\eta}_2 \leq \delta\boldsymbol{\eta}_1^T\mathbb{E}[\boldsymbol{H}_{k^*}]\boldsymbol{\eta}_1 + \frac{1}{\delta}\boldsymbol{\eta}_1^T\mathbb{E}[\boldsymbol{H}_{k^*}]\boldsymbol{\eta}_1\\
&= \delta c_1^{k^*} + \frac{1}{\delta}c_2^{k^*}.
\end{aligned} \tag{25}$$

Moreover, by simple manipulations we have

$$c_1^{k^*} = \boldsymbol{\eta}_1^T\mathbb{E}[\boldsymbol{H}_{k^*}]\boldsymbol{\eta}_1 = \frac{1}{n}\boldsymbol{X}^T\boldsymbol{\Gamma}_{k^*}\boldsymbol{X}, \tag{26}$$

where $\boldsymbol{X} = [x_1, \dots, x_m]^T \in \mathbb{R}^{m\times 1}$. Notice that by the definition of $\boldsymbol{H}_{k^*}$ and μ^*, we have

$$|c_2^{k^*}| \leq \frac{h}{\mu^*}\|\boldsymbol{\eta}_2\|^2 = \frac{h}{\mu^*}\sum_{j=m+1}^{mn}x_j^2. \tag{27}$$

Lastly, for $c_5^{k^*}$, we know that

$$c_5^{k^*} = h\nu\sum_{j=m+1}^{mn}l_jx_j^2 \leq h\nu l_{mn}\sum_{j=m+1}^{mn}x_j^2. \tag{28}$$

From (24)–(28) and the fact that $\rho_{k^*} = \mu\lambda_{min}(\boldsymbol{\Delta}_{k^*})$, it is easy to know that by taking $\boldsymbol{X} = \boldsymbol{y}_{k^*}$ in (24),

$$\rho_{k^*} \leq (1+\delta)\frac{\mu}{n}\cdot\boldsymbol{y}_{k^*}^T\boldsymbol{\Gamma}_{k^*}\boldsymbol{y}_{k^*} = 0, \tag{29}$$

which contradicts with our assumption $\rho_k > 0, \forall k$. Hence, for any $k \geq 0$, the eigenvalues of matrices $\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=k+1}^{k+h}\boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T\right]$ are positive.

Furthermore, we prove that all of the eigenvalues of the above matrix must have a uniform lower bound $\delta > 0$ with respect to $k \geq 0$. This can be similarly completed by the contradiction argument by assuming that there exists a subsequence $\{k_s\}_{s=1}^{\infty}$ such that

$$\begin{aligned}
&\lim_{s\to\infty}\boldsymbol{y}_{k_s}^T\left(\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=k_s+1}^{k_s+h}\boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T\right]\right)\boldsymbol{y}_{k_s}\\
&= \lim_{s\to\infty}\boldsymbol{y}_{k_s}^T\beta_{k_s}\boldsymbol{y}_{k_s} = 0,
\end{aligned}$$

where β_{k_s} is the eigenvalue corresponding to the eigenvector $\boldsymbol{y}_{k_s}$. Similar to the proof of (29), we can obtain

$$\lim_{s\to\infty}\rho_{k_s} \leq (1+\delta)\frac{\mu}{n}\cdot\lim_{s\to\infty}\boldsymbol{y}_{k_s}^T\boldsymbol{\Gamma}_{k_s}\boldsymbol{y}_{k_s} = 0,$$

which contradicts with $\inf_k\rho_k > 0$. Therefore, Lemma 22 holds.

References

Braca, P., Marano, S., & Matta, V. (2008). Running consensus in wireless sensor networks. In *11th international conference on information fusion* (pp. 1–6), Cologne, Germany.

Cattivelli, F. S., & Sayed, A. H. (2010). Diffusion LMS strategies for distributed estimation. *IEEE Transactions on Signal Processing*, *58*(3), 1035–1048.

Cattivelli, F. S., & Sayed, A. H. (2011). Analysis of spatial and incremental lms processing for distributed estimation. *IEEE Transactions on Signal Processing*, *59*(4), 1465–1480.

Chen, C., Liu, Z. X., & Guo, L. (2014). Stability of diffusion adaptive filters. In *Proc. 19th IFAC world congress* (pp 10409–10414), Cape Town.

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

Chen, J. S., & Sayed, A. H. (2015a). On the learning behavior of adaptive networks part i: transient analysis. *IEEE Transactions on Information Theory*, *61*(6), 3487–3517.

Chen, J. S., & Sayed, A. H. (2015b). On the learning behavior of adaptive networks part ii: performance analysis. *IEEE Transactions on Information Theory*, *61*(6), 3518–3548.

Chen, W. S., Wen, C. Y., Hua, S. Y., & Sun, C. Y. (2014). 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., & Teicher, H. (1978). *Probability theory*. New York: Springer.
Guo, L. (1994). Stability of recursive stochastic tracking algorithms. *SIAM Journal on Control and Optimization*, *32*(5), 1195–1225.
Guo, L., & Ljung, L. (1995a). Exponential stability of general tracking algorithms. *IEEE Transactions on Automatic Control*, *40*(8), 1376–1387.
Guo, L., & Ljung, L. (1995b). Performance analysis of general tracking algorithms. *IEEE Transactions on Automatic Control*, *40*(8), 1388–1402.
Guo, L., Ljung, L., & Wang, G. J. (1997). Necessary and sufficient conditions for stability of LMS. *IEEE Transactions on Automatic Control*, *42*(6), 761–770.
Kar, S., & Moura, J. MF. (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., & Sayed, A. H. (2007). Incremental adaptive strategies over distributed networks. *IEEE Transactions on Signal Processing*, *55*(8), 4064–4077.
Nosrati, H., Shamsi, M., Taheri, S. M., & Sedaaghi, M. Ho. (2015). Adaptive networks under non-stationary conditions: formulation, performance analysis, and application. *IEEE Transactions on Signal Processing*, *63*(16), 4300–4314.
Piggott, M. J., & Solo, V. (2016). Diffusion LMS with correlated regressors i: realization-wise stability. *IEEE Transactions on Signal Processing*, *64*(21), 5473–5484.
Piggott, M. J., & Solo, V. (2017). Diffusion LMS with correlated regressors ii: performance. *IEEE Transactions on Signal Processing*, *65*(15), 3934–3947.
Sayed, A. H. (2014a). Adaptation, learning, and optimization over networks. *Foundations and Trends in Machine Learning*, *7*(4–5), 311–801.
Sayed, A. H. (2014b). Adaptive networks. *Proceddings of the IEEE*, *102*(4), 460–497.
Solo, V. (2015). Stability of distributed adaptive algorithms i: consensus algorithms. In *Proc. 54th IEEE conference on decision and control* (pp 7422–7427), Osaka, Japan.
Solo, V., & Kong, X. (1995). *Adaptive signal processing algorithms*. Prentice-Hall, Inc.
Wang, L., Liu, Z. X., & Guo, L. (2007) Robust consensus of multi-agent systems with noise. In *Proc. 26th Chinese Control Conference* (pp 737–741), Zhangjiajie, China.
Widrow, B., & Stearns, S. D. (1985). *Adaptive signal processing*. Englewood Cliffs NJ: Prentice-Hall.
Xie, S. Y., & Guo, L. (2015) Stability of distributed LMS under cooperative stochastic excitation. In *Proc. 34th Chinese control conference* (pp 7445–7450), Hangzhou, China.
Xie, S. Y., & Guo, L. (2017). Exponential stability of LMS-based distributed adaptive filters. In *Proc.of the 20th world congress of the international federation of automatic control* (pp. 13764–13769), Toulouse, France.
Xie, S. Y., & Guo, L. (2018). Analysis of distributed adaptive filters based on diffusion strategies over sensor networks. *IEEE Transactions on Automatic Control*. http://dx.doi.org/10.1109/TAC.2018.2799567.

Siyu Xie received her B.S. degree in information and computing science (systems control) from Beihang University, Beijing, China. She is currently a Ph.D. candidate majoring in control theory at Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China.

Her research interests include networked systems, distributed adaptive filters, machine learning, compressive sensing and distributed control. She received the IEEE CSS Beijing Chapter Young Author Prize at Chinese Control Conference in 2015, and China National Scholarship for graduate students in 2017.

Lei Guo was born in 1961. He received his B.S. degree in mathematics from Shandong University in 1982, and Ph.D. degree in control theory from the Chinese Academy of Sciences(CAS) in 1987. Since 1992, he has been a Professor at Institute of Systems Science, CAS. He is a Fellow of IEEE, Member of Chinese Academy of Sciences, Fellow of Academy of Sciences for the Developing World (TWAS), Foreign Member of Royal Swedish Academy of Engineering Sciences, and Fellow of International Federation of Automatic Control (IFAC). He was awarded an honorary doctorate by Royal Institute of Technology (KTH), Sweden, in 2014, Outstanding Service Award of IFAC in 2011, National Natural Science Prizes of China in 1987 and 1997. He has served as a Council Member of IFAC, President of China Society for Industrial and Applied Mathematics (CSIAM), Vice-President of Chinese Mathematical Society, and Vice-President of Chinese Association of Automation.

He has worked on problems in stochastic adaptive control, system identification, adaptive signal processing, quantum control systems, and the maximum capability of feedback. His current research interests include distributed adaptive filtering, nonlinear uncertain systems control, multi-agent systems, game-based control systems, and PID control theory, among others.