SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

November 2016, Vol. 59 112202:1–112202:10
doi: 10.1007/s11432-016-0050-9

# Performance bounds of distributed adaptive filters with cooperative correlated signals

Chen CHEN[1], Zhixin LIU[2] & Lei GUO[2]*

[1]*Shannon Cognitive Computing Laboratory, 2012 Lab, Huawei Technologies Co. Ltd., Beijing* 100085*, China;*
[2] *Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing* 100190*, China*

Received January 18, 2016; accepted February 25, 2016; published online October 17, 2016

**Abstract** In this paper, we studied the least mean-square-based distributed adaptive filters, aiming at collectively estimating a sequence of unknown signals (or time-varying parameters) from a set of noisy measurements obtained through distributed sensors. The main contribution of this paper to relevant literature is that under a general stochastic cooperative signal condition, stability and performance bounds are established for distributed filters with general connected networks without stationarity or independency assumptions imposed on the regression signals.

**Keywords** distributed adaptive filters, LMS, random process, stochastic stability, graph connectivity

**Citation** Chen C, Liu Z X, Guo L. Performance bounds of distributed adaptive filters with cooperative correlated signals. Sci China Inf Sci, 2016, 59(11): 112202, doi: 10.1007/s11432-016-0050-9

## 1 Introduction

Filtering or parameter tracking is a basic problem in several areas including communication, statistics, signal processing, system identification, and control systems. Some classical algorithms and theoretical results are presented in [1–3]. However, to perform filtering task in more complex systems, a single sensor may only sense or observe partial information of unknown signals. Moreover, due to limited communication ability, a sensor may exchange information with its neighbors only. Thus, scientists design distributed filters according to partial measurements and local communications among neighbor sensors, and utilize the network topology of the spatially distributed sensors to enhance the observability of the network. Owing to the increasing practical demands, the design and theoretical analysis of such distributed filters are of great significance in, for instance, sensor networks [4]. Compared to the centralized method in which all measurements must be transmitted to a fusion sensor for processing, the distributed filtering scheme may have the advantages of reducing the communication and computation costs and enhancing the robustness with respect to partial node failures [5].

Recently, a variety of consensus-based distributed filtering algorithms have been proposed, which contain certain consensus schemes to reflect the cooperation among sensors of networks. Stability and performance analyses of various algorithms have been conducted, and most results concern with the estimation

* Corresponding author (email: lguo@amss.ac.cn)

 

49

of constant unknown parameters. Further, the existing theoretical literature can be roughly classified into two categories depending on the assumptions of the regression vectors (or observation matrices): One is the case in which the regression vectors are deterministic, so that the desired cooperation properties of the distributed filters can be established under a joint deterministic excitation condition [5–8]. It is worth remarking that the distributed algorithm investigated in [5] may still be classified into this category because the mathematical expectation of the observation matrices rather than stochastic matrices is used in the implementation of the algorithms. The other category is the stochastic case in which the regression vectors are usually assumed to be stationary and independent with the positive covariance matrices [9–11]. More recent results are presented in [12, 13], which show that with optimal choices for the combined weights, the distributed strategy can be designed to match or exceed the performance of the conventional unweighted, centralized strategy. Note that the deterministic case can also be regarded as a degenerated independent stochastic case.

However, the stationarity and independency are simplifications and idealizations of complicated practical situations. In fact, in almost all control systems described by stochastic regression models, the regression vectors usually contain input and output signals of the closed-loop systems, which are obviously correlated. A typical case is the stochastic adaptive control systems, in which the closed loop signals are usually determined through a set of highly complicated, nonlinear, and stochastic equations, which cannot be simplified to stationary and independent cases [14].

Through this paper, we aim to relax the usually used stationarity and independency assumptions of the stochastic regression vectors in the existing relevant literature, and to realize the stability of distributed filters in a general stochastic case. We investigate the diffusion least mean-square algorithm (DLMS) introduced in [9]. The output signals are generated using a linear time-varying stochastic regression model. We will establish the stability and performance bounds of the distributed filters under a general stochastic cooperative signal condition, without imposing the stationarity or independency assumptions on the regression signals. To this end, we need to investigate the product of random matrices, in which the major problem is that the random matrices are temporally non-commutative, non-independent and spatially coupled. Furthermore, we will also prove that general connected networks of sensors can cooperate to guarantee the stability of the filtering algorithm, even though some individual filters may not have such a capability.

The remainder of this paper is organized as follows. In Section 2, we present the DLMS algorithm and the main results concerning stability and performance analyses. In Sections 3 and 4, we prove a key technical lemma, and the main theorems, respectively. Finally, the concluding remarks are made in Section 5.

## 2 The main results

### 2.1 Problem formulation

To present the DLMS algorithm, we first introduce some notations. $I_n$ represents an identity matrix with order $n$. Operators $(\cdot)^\tau$, $\mathrm{diag}(\cdot)$, $\lambda_{\max}(\cdot)$, and $\lambda_{min}(\cdot)$ denote matrix transpose, diagonal matrix, the largest eigenvalue, and the smallest eigenvalue of the corresponding matrix, respectively. The notation $\mathrm{col}(\cdots)$ stands for a vector in a stack of specified vectors. $\otimes$ denotes the Kronecker product of two matrices. A matrix $A \geqslant 0$ indicates that $A$ is semi-positive definite, and $A \geqslant B$ implies $A - B \geqslant 0$. The matrix $A$ is stochastic if each element of $A$ is non-negative and the sum of each row is 1. Furthermore, a matrix is called doubly stochastic if it is a stochastic matrix and the sum of each column is also 1. A stochastic matrix $A$ is called ergodic if $\lim_{t\to\infty} A^t$ exists and all its rows are the same. For a vector $x$, $\|x\|$ represents the Euclidean norm of $x$, and for a matrix $A$, its norm is defined as $\|A\| = \{\lambda_{\max}(AA^\tau)\}^{\frac{1}{2}}$. We refer to $\|A\|_{L_p}$ defined by $\|A\|_{L_p} \triangleq \{\mathbb{E}\|A\|^p\}^{\frac{1}{p}}$ as the $L_p$-norm of $A$.

Consider a network comprising of $n$ sensors in which only single-hop communication is allowed, that is, sensor $i$ can only communicate with the sensors in its neighborhood $\mathcal{N}_i \subset \{1, \ldots, n\}$. We use graph $\mathcal{G} = \{V, E\}$ to describe the relationship between sensors, where the vertex $V$ is the set of sensors and the edge set $E$ is defined as follows: $(i,j) \in E$ if and only if the sensor $j$ is a neighbor of sensor $i$. For convenience of analysis, we assume that the graph $\mathcal{G}$ is undirected and contains a self-loop at each vertex, that is, $i \in \mathcal{N}_i$ for any $i \in V$. An $n \times n$ matrix $A = \{a_{ij}\}$ is introduced to represent the weights of the edges with $a_{ij} = a_{ji}$ and $\sum_{j=1}^{n} a_{ij} = 1$, where $a_{ij} > 0$ if and only if $(i,j) \in E$.

The task of the sensor network is to estimate a sequence of $m$-dimensional time-varying signal (or parameter) vectors $\{\theta_k,\ k = 1,2,\ldots\}$. We may write the evolution of the $\theta_k$ as follows:

$$\theta_k = \theta_{k-1} + \gamma\omega_k, \tag{1}$$

where $\gamma$ represnts a scaling constant and $\omega_k$ represents a vector describing the direction of variation. We assume that the unknown signal $\{\theta_k\}$ is to be estimated by $n$ distributed sensors, and that the signal $\{y_k^i, \varphi_k^i\}$ received conforms to the following time-varying stochastic linear regression model:

$$y_k^i = (\varphi_k^i)^\tau \theta_k + v_k^i, \tag{2}$$

where $y_k^i$ and $v_k^i$ are scalar observation and noise signal at sensor $i$ respectively, and $\varphi_k^i$ is the $m$-dimensional stochastic regression vector. Obviously, the classical linear regression model corresponds to the single sensor case $n = 1$.

In this paper, we use the combine-then-adapt (CTA) diffusion method, which was first introduced in [9] to estimate $\{\theta_k\}$. First, at each time instant $k$ $(k \geqslant 0)$, sensor $i$ has access to the estimates of its neighbors $j$ $(j \in \mathcal{N}_i)$ and obtains the aggregate estimate $\vartheta_k^i$ for $\theta_k$ by minimizing the following objective function:

$$\vartheta_k^i = \mathrm{argmin}_{\theta \in R^m} \sum_{j=1}^{n} a_{ij}(\theta - \hat{\theta}_k^j)^2 = \sum_{j=1}^{n} a_{ij}\hat{\theta}_k^j,$$

where $\hat{\theta}_k^j$ is the estimate of sensor $i$ for the unknown parameter $\theta_k$ at step $k$. We then use the normalized LMS algorithm to estimate the unknown parameters in which the estimate of sensor $i$ at step $k$ is replaced by the aggregate estimate $\vartheta_k^i$, that is, for $i = 1,2,\ldots,n$,

$$\hat{\theta}_{k+1}^i = \vartheta_k^i + \mu_i \frac{\varphi_k^i}{1 + \|\varphi_k^i\|^2}(y_k^i - (\varphi_k^i)^\tau \vartheta_k^i), \tag{3}$$

where $\mu_i \in (0,1)$ is a step-size, and the initial estimates $\{\hat{\theta}_0^i\}$ can be chosen arbitrarily.

**Remark 1.** The aforementioned CTA Diffusion LMS algorithm appears to be first introduced in [15], and then further investigated and reviewed in [12, 13]. It allows all the sensors to estimate the unknown signals simultaneously and to be able to respond to changes in the environment. In this form, the combination is performed before adaptation. There is another similar form in which the order is switched, that is, adaptation performed before combination, and is referred to as adapt-then-combine (ATC) DLMS, which shares almost the same properties with CTA form. Superior performance properties of DLMS implementations are established in [12, 13].

Our main goal is to establish the global stability and performance analysis of the algorithm with more practical and general regression signals. To proceed with our analysis, we introduce the following global quantities:

$$\begin{aligned}
&\Theta_k \triangleq \mathrm{col}\{\underbrace{\theta_k,\ldots,\theta_k}_{n}\},\quad W_k \triangleq \mathrm{col}\{\underbrace{\omega_k,\ldots,\omega_k}_{n}\},\\
&Y_k \triangleq \mathrm{col}\{y_k^1,\ldots,y_k^n\},\quad \vartheta_k \triangleq \mathrm{col}\{\vartheta_k^1,\ldots,\vartheta_k^n\},\\
&\Psi_k \triangleq \mathrm{diag}\{\varphi_k^1,\ldots,\varphi_k^n\},\quad \hat{\Theta}_k \triangleq \mathrm{col}\{\hat{\theta}_k^1,\ldots,\hat{\theta}_k^n\},\\
&\widetilde{\Theta}_k \triangleq \mathrm{col}\{\widetilde{\theta}_k^1,\ldots,\widetilde{\theta}_k^n\} \ \text{with}\ \widetilde{\theta}_k^i = \widehat{\theta}_k^i - \theta_k^i,\\
&V_k = \mathrm{col}\{v_k^1,\ldots,v_k^n\},\\
&L_k \triangleq \mathrm{diag}\left\{\frac{\varphi_k^1}{1+\|\varphi_k^1\|^2},\ldots,\frac{\varphi_k^n}{1+\|\varphi_k^n\|^2}\right\},
\end{aligned}$$

$$\Lambda \triangleq \mathrm{diag}\{\mu_1 I_m, \dots, \mu_n I_m\}, \quad \boldsymbol{A} = A \otimes I_m.$$

The linear regression model (2) can then be written as

$$Y_k = \Psi_k^\tau \Theta_k + V_k. \tag{4}$$

Correspondingly, by using the Kronecker product $\otimes$, Eq. (3) can be written as

$$\hat{\Theta}_{k+1} = \boldsymbol{A}\hat{\Theta}_k + \Lambda L_k (Y_k - \Psi_k^\tau \boldsymbol{A}\hat{\Theta}_k). \tag{5}$$

By subtracting $\Theta_k$ from both sides of (5) and considering $(A \otimes I_m)\Theta_k = \Theta_k$, we obtain

$$\begin{aligned} \widetilde{\Theta}_{k+1} &= (I_{mn} - \Lambda F_k)\boldsymbol{A}\widetilde{\Theta}_k + \Lambda L_k V_k - \gamma W_{k+1}, \\ F_k &= L_k \Psi_k^\tau. \end{aligned} \tag{6}$$

### 2.2 Definitions and assumptions

To analyze the properties of $\widetilde{\Theta}_k$, some conditions are required on the regressors and noises. To this end, we need to introduce some definitions from [16–18].

**Definition 1.** A random matrix (or vector) sequence $\{A_k, k \geqslant 0\}$ defined on the basic probability space $(\Omega, \mathcal{F}, P)$ is called $L_p$-stable ($p > 0$) if $\sup_{k \geqslant 0} \mathbb{E}\|A_k\|_{L_p} < \infty$.

**Definition 2.** A sequence of $d \times d$ random matrices $A = \{A_k, k \geqslant 0\}$ is called $L_p$-exponentially stable ($p > 0$) with parameter $\lambda \in [0, 1)$ if it belongs to the set $S_p(\lambda) = \{A : \| \prod_{j=i+1}^k A_j \|_{L_p} \leqslant M\lambda^{k-i}, \ \forall k \geqslant i \geqslant 0, \text{ for some M} > 0\}$.

**Definition 3.** For a scalar random sequence $a = \{a_i, i \geqslant 0\}$, we set $S^o(\lambda) = \big\{a_i \in [0,1] : \mathbb{E}\prod_{j=i+1}^k (1 - a_j) \leqslant M\lambda^{k-i}, \forall k \geqslant i \geqslant 0, \text{ for some M} > 0\big\}$.

Denote $S^o \triangleq \cup_{\lambda \in (0,1)} S^o(\lambda)$. To proceed further, we must introduce a class of weakly dependent random sequences as follows.

**Definition 4.** We say that a random sequence $x \triangleq \{x_k\} \in \mathcal{M}_p$ ($p \geqslant 1$) if there exists a constant $C_p(x)$ only depending on $p$ and $\{x_k\}$ such that for $j \geqslant 0$,

$$\left\| \sum_{i=j+1}^{j+h} x_i \right\|_{L_p} \leqslant C_p(x) h^{\frac{1}{2}}, \quad \forall h \geqslant 1. \tag{7}$$

**Remark 2.** As is known, the martingale difference sequence, $\phi$-mixing and $\alpha$-mixing sequences, and the linear process driven by white noises are all in the set $\mathcal{M}_p$ (see, [16]).

**Definition 5.** Let $\{A_k\}$ be a matrix sequence, and $b_k$ be a positive scalar sequence. Then, by $A_k = O(b_k)$, we indicate that there exists a constant $M < \infty$ such that for any $k \geqslant 0$, we have $\|A_k\| \leqslant M b_k$.

Throughout the sequel, we use $\mathcal{F}_k$ to denote the $\sigma$-algebra generated by $\{\varphi_i^j, \ \omega_i, \ v_{i-1}^j, \ j = 1, \dots, n, \ i \leqslant k\}$.

**Assumption 1** (Connectivity). The graph $\mathcal{G}$ is connected.

**Remark 3.** By Assumption 1 and the symmetry and self-loop properties of $\mathcal{G}$, we see that the weighted stochastic matrix $A$ is ergodic. Furthermore, we know that $A$ has $n$ real eigenvalues which can be arranged in a non-decreasing order: $-1 < \lambda_n(A) \leqslant \lambda_{n-1}(A) \leqslant \cdots \leqslant \lambda_2(A) < \lambda_1(A) = 1$. Then, $\lambda_{\text{gap}} \triangleq \max\{|\lambda_2(A)|, |\lambda_n(A)|\} < 1$.

**Assumption 2** (Cooperative excitation condition). Let $\{\varphi_k^i, \ \mathcal{F}_k, \ k \geqslant 0, \ i = 1, \dots, n\}$ be $n$ adapted sequences and $\{\lambda_k, \ k \geqslant 0\} \in S^o(\lambda)$ for some $\lambda \in (0, 1)$, where

$$\lambda_k \triangleq \lambda_{\min}\left\{ \mathbb{E}\left[ \frac{1 - \lambda_{\text{gap}}}{n} \sum_{i=1}^n \frac{\varphi_{k+1}^i (\varphi_{k+1}^i)^\tau}{1 + \|\varphi_{k+1}^i\|^2} \ \Bigg| \ \mathcal{F}_k \right] \right\}. \tag{8}$$

**Remark 4.** To estimate the unknown signals, it is necessary to require some excitation properties on the regressors. The aforementioned is called a cooperative excitation condition because it is imposed on the collected regressors. Furthermore, if the regressors are stationary and independent with positive definite covariance matrices and if the network has a connected topology, then $\{\lambda_k\}$ defined by (8) will have a positive lower bound and $\lambda_{\text{gap}} < 1$. Therefore, Assumption 2 will be typically true. This implies that the commonly used assumptions in literature are a special case of Assumption 2. The advantages of Assumption 2 in the stability analysis of the diffused adaptive filters will be discussed in Remark 5. In addition, more discussions on similar conditions for a single filter can be found in [17], in which it has been shown that some non-stationary and correlated signals of practical interests can be included.

**Assumption 3.** For some $p \geqslant 1$, let the initial estimation error be bounded, that is, $\|\widetilde{\Theta}_0\|_{L_{2p}} < \infty$. Furthermore, let $\{L_k V_k\} \in \mathcal{M}_{2p}$ and $\{W_k\} \in \mathcal{M}_{2p}$.

We remark that by Definition 4, this assumption simply implies that both the noises and parameter variations are weakly dependent with certain bounded moments.

### 2.3 The main theorems

**Lemma 1.** If Assumptions 1 and 2 are satisfied, then for any $0 \leqslant \Lambda < I_{mn}$ and $p \geqslant 1$, the coefficient matrix sequence of the homogeneous part of (6) $\{(I_{mn} - \Lambda F_k)\boldsymbol{A}, k \geqslant 1\}$ is $L_p$-exponentially stable.

**Remark 5.** For a single sensor with signals generated by the regression model

$$y_k = \varphi_k^\tau \theta_k + v_k,$$

where $y_k$, $\varphi_k$, $\theta_k$ and $v_k$ are observations, regression vectors, unknown parameters, and noises, respectively. If we employ the conventional single LMS to estimate $\theta_k$, then the corresponding homogeneous part of the estimation error equation will be $L_2$-exponentially stable whenever the following conditional excitation condition [17] is true:

$$\lambda_{\min}\left\{\mathbb{E}\left[\frac{1}{1+h}\sum_{j=kh+1}^{(k+1)h}\frac{\varphi_j(\varphi_j)^\tau}{1+\|\varphi_j\|^2}\,\middle|\,\mathcal{F}_{kh}\right]\right\} \in S^o, \tag{9}$$

where $S^o$ is defined in Definition 3. In addition, this condition is considered necessary for stability when the signals are weakly correlated in a certain sense [17]. For a network of sensors, even though none of the sensors satisfy condition (9), all the sensors as a whole can still be able to satisfy Assumption 2. Thus, the DLMS algorithm can fulfill the estimation task in a cooperative manner, even when any sensor does not have sufficient excitation. For example, in [19], the regressor vectors of each sensor are the outputs of a linear stochastic system, which are neither independent nor stationary. It can be verified that each single sensor does not satisfy (9), but Assumption 2 can be satisfied jointly .

According to Lemma 1, we can establish a preliminary result on the upper bound of the moments of the tracking error.

**Theorem 1.** Suppose that Assumptions 1 and 2 are satisfied and that for some $p \geqslant 1$, $\beta > 1$, the initial estimation error is bounded, that is, $\|\widetilde{\Theta}_0\|_{L_p} < \infty$, and $\sigma_p \triangleq \sup_k \|\xi_k \log^\beta(e + \xi_k)\|_{L_p} < \infty$, where $\xi_k = \|V_k\| + \|W_{k+1}\|$. Then, the filtering error sequence $\{\widetilde{\Theta}_k, k \geqslant 1\}$ generated by (6) is $L_p$-stable and

$$\limsup_{k\to\infty} \|\widetilde{\Theta}_k\|_{L_p} \leqslant c[\sigma_p \log(e + \sigma_p^{-1})],$$

where $e$ is the base of natural logarithms and $c$ is a positive constant.

*Proof.* Let $c_{ki} = \|\prod_{j=i+1}^k ((I_{mn} - \Lambda F_j)\boldsymbol{A})\|$. By considering Assumption 2, Lemma 2.3 in [17], and Lemma 1, $\{c_{ki}\}$ satisfies the conditions in Lemma 4.1 in [17]. Note that $\|\Lambda L_k\| \leqslant \max_{1\leqslant i\leqslant n} \mu_i$. Thus, according to (6),

$$\|\widetilde{\Theta}_{k+1}\|_{L_p} \leqslant \|c_{k,-1}\widetilde{\Theta}_0\|_{L_p} + \sum_{i=0}^{k}\|c_{ki}\xi_k\|_{L_p}.$$

The desired result of the theorem can be obtained by applying Lemma 4.1 in [17].

Under more specific conditions on signals and measurement noises, we can obtain a better bound for the tracking error.

**Theorem 2.** Assume that $\Lambda = \mu I_{mn}$, where $\mu \in (0, 1/e)$, and Assumptions 1–3 are satisfied. Then, we have

$$\|\widetilde{\Theta}_{k+1}\|_{L_p} = O\bigg(\bigg[\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}}\bigg]\log\frac{1}{\mu} + (1-\mu\alpha)^k\|\widetilde{\Theta}_0\|_{L_{2p}}\bigg),$$

where the positive constant $\alpha \in (0, 1)$ depends on $\{F_k, k \geqslant 0\}$ and $p$.

The proof of Theorem 2 is provided in Section 4.

According to this theorem and following similar ideas in [16], we can further obtain the "dominant term" of the tracking errors because of the connectivity of the filtering network, which is not presented in this study because of space limitations, see [20].

## 3 Proof of Lemma 1

To prove Lemma 1, we consider the following lemma.

**Lemma 2.** Suppose that Assumption 1 holds true and that $\{\Phi_k = (\Phi_k^{ij}) \in \mathbb{R}^{m\times m},\ k = 1, \dots, n\}$ is a sequence of symmetric matrices satisfying $0 \leqslant \Phi_k \leqslant I_m,\ k = 1, \dots, n$. Then, $\lambda_{\max}[\boldsymbol{A} \cdot \mathrm{diag}(I_m - \Phi_1, \dots, I_m - \Phi_n) \cdot \boldsymbol{A}] \leqslant 1 - (1 - \lambda_{\text{gap}})\delta$, where $\delta = \frac{1}{n}\lambda_{\min}(\Phi_1 + \cdots + \Phi_n)$ and $\lambda_{\text{gap}}$ is defined in Remark 3.

**Remark 6.** The aforementioned lemma is an improvement of the related result in [19] and clearly demonstrates the dependence of the upper bound of the random product on both the spectrum gap of the underlying network and the joint excitation of random matrices.

*Proof.* Let $\{\alpha_1, \dots, \alpha_n\}$ be a set of orthogonal basis of $\mathbb{R}^n$ composed of unit eigenvectors corresponding to $\{\lambda_1, \dots, \lambda_n\}$, with $\alpha_1 = (\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})^\tau$. Then, by the properties of Kronecker product in [21], the eigenvalues of $\boldsymbol{A}$ are $\{\lambda_i,\ i = 1, \dots, n\}$, and the eigenvectors corresponding to $\lambda_i$ are $\{\alpha_i \otimes e_j,\ j = 1, \dots, m\}$, where $e_i$ is the $i$th row of $I_m$. For convenience, we use $\beta_k$ $(1 \leqslant k \leqslant mn)$ to denote $\alpha_{\lceil k/m\rceil} \otimes e_{k-m\lfloor k/m\rfloor}$. Then, $\{\beta_k,\ k = 1, \dots, mn\}$ is a set of orthogonal eigenvectors of $A \otimes I_m$.

For any two = vectors $x \in \mathbb{R}^{mn}$ and $y \in \mathbb{R}^{mn}$, we denote

$$x^\tau * y \triangleq x^\tau\big[\boldsymbol{A}\cdot \mathrm{diag}(I_m - \Phi_1, \dots, I_m - \Phi_n)\cdot \boldsymbol{A}\big]y, \qquad x^\tau \diamond y \triangleq x^\tau \mathrm{diag}(\Phi_1, \dots, \Phi_n)y.$$

Therefore, for $i = 1, \dots, mn$, we have $\beta_i^\tau * \beta_i = (\lambda_{\lceil i/m\rceil})^2\beta_i^\tau \mathrm{diag}(I_m - \Phi_1, I_m - \Phi_2, \dots, I_m - \Phi_n)\beta_i = (\lambda_{\lceil i/m\rceil})^2(1 - \beta_i^\tau \diamond \beta_i)$, and for $i \neq j$, $\beta_i^\tau * \beta_j = (\lambda_{\lceil i/m\rceil})(\lambda_{\lceil j/m\rceil})\beta_i^\tau \mathrm{diag}(I_m - \Phi_1, I_m - \Phi_2, \dots, I_m - \Phi_n)\beta_j = -(\lambda_{\lceil i/m\rceil})(\lambda_{\lceil j/m\rceil})\beta_i^\tau \diamond \beta_j$. Further, for any arbitrary unit vector $x \in \mathbb{R}^{mn}$, $x$ can be written as $x = \sum_{i=1}^{mn} x_i\beta_i$, with $\sum_{i=1}^{mn} x_i^2 = 1$. Thus, it is not difficult to see that

$$x^\tau * x = S_1 - S_2 \triangleq \sum_{i=1}^{mn} x_i^2(\lambda_{\lceil i/m\rceil})^2 - \left(\sum_{i=1}^{mn} x_i\lambda_{\lceil i/m\rceil}\beta_i\right)^\tau \diamond \left(\sum_{i=1}^{mn} x_i\lambda_{\lceil i/m\rceil}\beta_i\right). \tag{10}$$

We will consider $S_1$ and $S_2$ separately. Note that $\lambda_{\lceil i/m\rceil} = 1$ for $1 \leqslant i \leqslant m$, and $\lambda_{\lceil i/m\rceil} \leqslant \lambda_{\text{gap}}$ for $m + 1 \leqslant i \leqslant mn$. Therefore, we have

$$S_1 \leqslant \sum_{i=1}^{m} x_i^2 + \lambda_{\text{gap}}^2\left(1 - \sum_{i=1}^{m} x_i^2\right). \tag{11}$$

Set $S_2^1 \triangleq \sum_{i=1}^{m} x_i\lambda_{\lceil i/m\rceil}\beta_i$ and $S_2^2 \triangleq \sum_{i=m+1}^{mn} x_i\lambda_{\lceil i/m\rceil}\beta_i$. Then,

$$S_2 = (S_2^1 + S_2^2)^\tau \diamond (S_2^1 + S_2^2) = (S_2^1)^\tau \diamond S_2^1 + (S_2^1)^\tau \diamond S_2^2 + (S_2^2)^\tau \diamond S_2^1 + (S_2^2)^\tau \diamond S_2^2. \tag{12}$$

The two cross terms are equal and can be managed by Schwarz inequality as follows: for any $\epsilon \in (0, 1)$, we have $2(S_2^1)^\tau \diamond S_2^2 \leqslant \epsilon(S_2^1)^\tau \diamond S_2^1 + \epsilon^{-1}(S_2^2)^\tau \diamond S_2^2$. Consequently, substituting this into (12), we have $S_2 \geqslant (1 - \epsilon)(S_2^1)^\tau \diamond S_2^1 + (1 - \epsilon^{-1})(S_2^2)^\tau \diamond S_2^2$.

Note that $S_2^1 = \frac{1}{\sqrt{n}}\mathrm{col}\{\underbrace{(x_1,\ldots,x_m)^\tau}_{1},\ldots,\underbrace{(x_1,\ldots,x_m)^\tau}_{n}\}$. Then,

$$|(S_2^1)^\tau \diamond S_2^1| = \frac{1}{n}(x_1,\ldots,x_m)\left(\sum_{i=1}^{n}\Phi_i\right)(x_1,\ldots,x_m)^\tau \geqslant \delta\left(\sum_{i=1}^{m}x_i^2\right). \tag{13}$$

Next, the term $(S_2^2)^\tau \diamond S_2^2$ is relatively easy to be managed. As $0 \leqslant \mathrm{diag}(\Phi_1,\ldots,\Phi_n) \leqslant I_{mn}$, we know that

$$|(S_2^2)^\tau \diamond S_2^2| \leqslant \|S_2^2\|^2 \leqslant \sum_{i=m+1}^{mn} x_i^2(\lambda_{\lceil i/m\rceil})^2 \leqslant \lambda_{\mathrm{gap}}^2(1-\textstyle\sum_{i=1}^m x_i^2). \tag{14}$$

Let us now denote $y=\sum_{i=1}^m x_i^2$. By substituting (11)–(14) into (10), we finally obtain $x^\tau * x \leqslant [1-(1-\epsilon)\delta]y+\epsilon^{-1}\lambda_{\mathrm{gap}}^2(1-y)$. Taking $\epsilon=\lambda_{\mathrm{gap}}<1$, we get $x^\tau * x \leqslant [1-(1-\lambda_{\mathrm{gap}})\delta]y+\lambda_{\mathrm{gap}}(1-y) \leqslant [1-(1-\lambda_{\mathrm{gap}})\delta]$, where the monotonicity of $[1-(1-\lambda_{\mathrm{gap}})\delta]y+\lambda_{\mathrm{gap}}(1-y)$ for $y\in(0,1)$ is used in the last inequality. This completes the proof.

**Lemma 3.** If Assumptions 1 and 2 are satisfied, then $\lambda_{\max}\{\mathbb{E}[\Phi^\tau(k+1,k)\Phi(k+1,k) \mid \mathcal{F}_{k-1}]\} \leqslant 1-\mu_{\min}\lambda_{k-1}$, where $\mu_{\min}=\min\{\mu_1,\ldots,\mu_n\}$ and $\Phi(\cdot,\cdot)$ is defined as

$$\begin{aligned}&\Phi(n+1,m)=(I_{mn}-\Lambda F_n)\boldsymbol{A}\Phi(n,m),\\ &\Phi(m,m)=I_{mn}, \forall n\geqslant m.\end{aligned} \tag{15}$$

*Proof.* By considering the notation $\boldsymbol{A}=A\otimes I_m$, and Lemma 2, we have

$$\begin{aligned}&\lambda_{\max}\{\mathbb{E}[\Phi^\tau(k+1,k)\Phi(k+1,k) \mid \mathcal{F}_{k-1}\}\\ &=\lambda_{\max}\{\boldsymbol{A}\mathbb{E}[(I_{mn}-\Lambda F_k)^2 \mid \mathcal{F}_{k-1}]\boldsymbol{A}\}\\ &\leqslant\lambda_{\max}\{\boldsymbol{A}\mathbb{E}[(I_{mn}-\Lambda F_k) \mid \mathcal{F}_{k-1}]\boldsymbol{A}\}\\ &=\lambda_{\max}\left\{\boldsymbol{A}\mathrm{diag}\left\{\left(I_m-\mu_1\mathbb{E}\left[\frac{\varphi_k^1(\varphi_k^1)^\tau}{1+\|\varphi_k^1\|^2}\middle|\mathcal{F}_{k-1}\right]\right),\ldots,\left(I_m-\mu_n\mathbb{E}\left[\frac{\varphi_k^n(\varphi_k^n)^\tau}{1+\|\varphi_k^n\|^2}\middle|\mathcal{F}_{k-1}\right]\right)\right\}\boldsymbol{A}\right\}\\ &\leqslant 1-\mu_{\min}\lambda_{k-1}.\end{aligned}$$

**Lemma 4.** For any $k_0\geqslant 0$ and $k\geqslant k_0+1$, consider the equation $x_k=\Phi(k+1,k)x_{k-1}$, where $x_{k_0}$ is deterministic and satisfies $\|x_{k_0}\|=1$. Then, under the same conditions and notations of Lemma 3, there exists a sequence $\{\alpha_k\in[0,1], k\geqslant k_0+1\}$ such that $\alpha_k\in\mathcal{F}_k$ and

$$\|x_k\|\leqslant(1-\alpha_k)\|x_{k-1}\|, \quad k\geqslant k_0+1, \tag{16}$$

and $E[\alpha_{k+1}|\mathcal{F}_k]\geqslant\frac{1}{2}\mu_{\min}\lambda_k,\ k\geqslant k_0+1$.

The proof of Lemma 4 is similar to that of Theorem 2.1 in [17], and we omitted it because of space limitation.

Proof of Lemma 1. As $\lambda_k\leqslant 1-\lambda_{\mathrm{gap}}$, then according to Lemma 2.3 in [17], we have $\{\frac{1}{2}\mu_{\min}\lambda_k\}\in S^o(\lambda^{\frac{1}{2}\lambda_{\mathrm{gap}}\mu_{\min}})$.

By using (4) and Lemma 2.1 in [17], we obtain $\{\alpha_k\}\in S^o(\lambda^{\frac{1}{4}\lambda_{\mathrm{gap}}\mu_{\min}})$. In addition, by Lemma 4,

$$\begin{aligned}\left\{\mathbb{E}\left\|\prod_{j=k_0+1}^{k}(I_{mn}-\Lambda F_j)\boldsymbol{A}\right\|^2\right\}^{1/2}&=\left\{\sup_{x_k\in\mathbb{R}^{mn}}\mathbb{E}\|x_k\|^2\right\}^{1/2}\\ &\leqslant\{\mathbb{E}(1-\alpha_k)^2\cdots(1-\alpha_{k_0+1})^2\}^{1/2}\\ &\leqslant\{\mathbb{E}(1-\alpha_k)\cdots(1-\alpha_{k_0+1})\}^{1/2}\\ &\leqslant M\lambda^{\frac{1}{8}\lambda_{\mathrm{gap}}\mu_{\min}(k-k_0)},\end{aligned}$$

where $M$ is a positive constant independent of $k$ and $k_0$.

By using Lyapunov inequality for $1 \leqslant p \leqslant 2$ and applying the inequality $\|(I_{mn} - \Lambda F_k)\boldsymbol{A}\| \leqslant 1$ for $p \geqslant 2$,

$$\left\| \prod_{j=k_0+1}^{k} (I_{mn} - \Lambda F_j)\boldsymbol{A} \right\|_{L_p} \leqslant \begin{cases} \left\| \prod_{j=k_0+1}^{k} (I_{mn} - \Lambda F_j)\boldsymbol{A} \right\|_{L_2}, & 1 \leqslant p \leqslant 2; \\ \left\| \prod_{j=k_0+1}^{k} (I_{mn} - \Lambda F_j)\boldsymbol{A} \right\|_{L_2}^{2/p}, & p > 2. \end{cases} \tag{17}$$

This completes the proof of Lemma 1.

## 4 Proof of Theorem 2

In this section, we set $\mu_i = \mu \in (0,1)$ $(i = 1, \ldots, n)$. The error iteration equation (6) can be expressed as

$$\widetilde{\Theta}_{k+1} = (I_{mn} - \mu F_k)\boldsymbol{A}\widetilde{\Theta}_k + \mu L_k V_k - \gamma W_{k+1}. \tag{18}$$

Next, define the matrix $\Phi(k+1, i)$ as follows

$$\Phi(k+1, i) = (I_{mn} - \mu F_k)\boldsymbol{A}\Phi(k, i), \quad \Phi(i, i) = I_{mn}, \quad \forall k \geqslant i \geqslant 0.$$

To prove Theorem 2, we must first establish some lemmas .

**Lemma 5.** Under Assumptions 1 and 2, we have for $p \geqslant 1$,

$$\left\|\Phi(k+1, i+1)\right\|_{L_p} \leqslant M_p(1 - \mu\alpha_p)^{k-i}, \tag{19}$$

where $M_p$ and $\alpha_p$ are positive constants depending on $\{F_j, j \geqslant 0\}$ and $p$.

*Proof.* By using (17), we can obtain $\|\Phi(k+1, i+1)\|_{L_p} \leqslant M_p\{\beta_p\}^{\mu(k-i)}$, where $\beta_p = \lambda^{\frac{1}{8}\lambda_{\text{gap}}}$ for $1 \leqslant p \leqslant 2$ and $\beta_p = \lambda^{\frac{1}{4p}\lambda_{\text{gap}}}$ for $p > 2$.

Note that $\beta_p \in (0,1)$. Then, there exists a constant $\alpha_p \in (0,1)$ such that $\beta_p = (1 - \alpha_p)$. According to Bernoulli inequality, for $\mu \in (0,1)$, we have $\beta_p^{\mu} = (1 - \alpha_p)^{\mu} < 1 - \mu\alpha_p$. This completes the proof of the lemma.

In the sequel, we use $\alpha$ to represent $\alpha_p$ regardless of the value of $p$.

We first present the following lemma, which describes the consensus property of the product of a sequence of ergodic stochastic matrices.

**Lemma 6** ([22]). Assume that $\{A_t \in \mathbb{R}^{N \times N},\ t = 1, 2, \ldots, n\}$ is a sequence of stochastic matrices, with a common stationary distribution $\pi = (\pi_1, \pi_2, \ldots, \pi_N)$. Set $A \triangleq A_n A_{n-1} \cdots A_1$. Then, for $i = 1, 2, \ldots, N$, we have $\sum_{j=1}^{N} \frac{1}{\pi_j}(A_{ij} - \pi_j)^2 \leqslant (\frac{1}{\pi_i} - 1) \prod_{t=1}^{n} \sigma_2^2(A_t)$, where $\sigma_2(A_t)$ is the second largest singular value of the matrix $A_t$, and $\sigma_2(A_t) < 1$ if $A_t$ is ergodic.

**Lemma 7.** Let Assumptions 1 and 2 hold true, and $\{e_k\} \in \mathcal{M}_r$ with $r \geqslant 1$. Then, for $s = (r^{-1} + p^{-1})^{-1}$ with $p \geqslant 1$ and $\mu \in (0, 1/e)$, we have $\| \sum_{i=0}^{k} \Phi(k+1, i+1)e_i \|_{L_s} = O(\mu^{-1/2} \log \frac{1}{\mu})$.

*Proof.* Set $S(k, i) = \sum_{j=i}^{k} e_j$, then

$$\sum_{i=0}^{k} \Phi(k+1, i+1)e_i = \Phi(k+1, 1)S(k, 0) + \sum_{i=1}^{k} [\Phi(k+1, i+1) - \Phi(k+1, i)]S(k, i). \tag{20}$$

By considering Hölder inequality for the first term on the right hand side of (20), we have

$$\| \Phi(k+1, 1)S(k, 0) \|_{L_s} \leqslant \| \Phi(k+1, 1) \|_{L_p} \| S(k, 0) \|_{L_r} = O\big([1 - \mu\alpha]^k \sqrt{k}\big) = O\Big(\mu^{-\frac{1}{2}}\Big), \tag{21}$$

where the last equality holds true because of i) of Lemma A.1 in [16].

Next, we analyze the second term on the right hand side of (20). For arbitrarily large $k$,

$$\sum_{i=1}^{k} [\Phi(k+1, i+1) - \Phi(k+1, i)]S(k, i) = \sum_{i=1}^{k-L} \Phi(k+1, i+1)\big[I_{mn} - (I_{mn} - \mu F_i)\boldsymbol{A}\big] S(k, i)$$

$$+\sum_{i=k-L+1}^{k}\Phi(k+1,i+1)\big[I_{mn}-(I_{mn}-\mu F_i)\boldsymbol{A}\big]S(k,i)$$
$$\triangleq S_1+S_2, \tag{22}$$

where $L$ is an integer satisfying $L=\lceil\frac{\log\mu}{\log\sigma_2(A)}\rceil+1$. Thus, by Lemma 6, we have

$$\max_{i,j}\left|(A^L)_{ij}-\frac{1}{n}\right|\leqslant\left\{\sum_{j=1}^{n}\left((A^L)_{ij}-\frac{1}{n}\right)^2\right\}^{1/2}\leqslant(\sigma_2(A))^L<\mu. \tag{23}$$

Note that $\|F_i\|\leqslant 1$. From i) of Lemma A.1 in [16], we have

$$\|S_2\|_{L_s}=O\left(\sum_{i=k-L+1}^{k}\|\Phi(k+1,i+1)\|_{L_p}\|S(k,i)\|_{L_r}\right)=O\left(\sum_{i=k-L+1}^{k}(1-\mu\alpha)^{k-i}\cdot\sqrt{k-i}\right)$$
$$=O\Big(\mu^{-\frac{1}{2}}\log\mu^{-1}\Big), \tag{24}$$

where the boundedness of $\|I_{mn}-(I_{mn}-\mu F_i)\boldsymbol{A}\|$ is used.

To estimate $\|S_1\|_{L_s}$, we can write it as $S_1=\sum_{i=1}^{k-L}\Phi(k+1,i+L+1)\Phi(i+L+1,i+1)\cdot[I_{mn}-(I_{mn}-\mu F_i)\boldsymbol{A}]S(k,i)$. For any $i>0$, by the fact that $\{F_j\}$ is a bounded matrix sequence, we have $\Phi(i+L,i)=\prod_{j=i}^{i+L-1}(I_{mn}-\mu F_j)\boldsymbol{A}=(A^L)\otimes I_m+O(\mu L)$, where we assume that $\mu$ is sufficiently small such that $\mu L=O(\mu\log\frac{1}{\mu})\leqslant\frac{1}{2}$. Hence, by (23), we have $\Phi(i+L,i)=A_{\rm ave}\otimes I_m+O(\mu\log\frac{1}{\mu})$, where $A_{\rm ave}=\lim_{k\to\infty}A^k$ with $(A_{\rm ave})_{ij}=1/n$. Then

$$\begin{aligned}&\Phi(i+L,i)\big[I_{mn}-(I_{mn}-\mu F_{i-1})\boldsymbol{A}\big]\\&=\left[(A_{\rm ave}\otimes I_m)+O\left(\mu\log\frac{1}{\mu}\right)\right]\big[I_{mn}-(I_{mn}-\mu F_{i-1})\boldsymbol{A}\big]\\&=(A_{\rm ave}\otimes I_m)-(A_{\rm ave}\otimes I_m)(I_{mn}-\mu F_{i-1})\boldsymbol{A}+O(\mu\log\frac{1}{\mu})\\&=\mu(A_{\rm ave}\otimes I_mF_{i-1}A\otimes I_m)+O\left(\mu\log\frac{1}{\mu}\right)\\&=O\left(\mu\log\frac{1}{\mu}\right).\end{aligned} \tag{25}$$

Thus, we have

$$\|S_1\|_{L_s}=O\left(\mu\log\frac{1}{\mu}\sum_{i=1}^{k-L}\|\Phi(k+1,i+L+1)\|_{L_p}\|S(k,i)\|_{L_r}\right)=O\left(\mu\log\frac{1}{\mu}\sum_{i=1}^{k-L}(1-\mu\alpha)^{k-i-L}\sqrt{k-i}\right)$$
$$=O\left(\mu^{-\frac{1}{2}}\log\frac{1}{\mu}\right), \tag{26}$$

where the equality (26) holds true because of iii) of Lemma A.1 in [16]. Combing (21), (24) with (26), we can complete the proof of the lemma.

Proof of Theorem 2. By using (18) and Lemmas 5 and 7, we can obtain the results of Theorem 2 immediately.

## 5 Concluding remarks

In this paper, we have established the stability and tracking performance bounds of the DLMS adaptive filtering algorithm, under general conditions of system signals which may be correlated and nonstationary. Furthermore, we demonstrated that general connected networks of sensors can cooperate to fulfill the stochastic estimation or filtering task even though any individual sensor cannot because of possible degeneration of covariance matrices of the regression vectors. However, many questions need to be investigated, for example, further relaxation of the cooperative excitation assumption for the DLMS filters presented in this study, analysis and comparison with other types of distributed filtering algorithms including the Kalman filtering-based cooperative algorithm [23–25], and the combination of the distributed filtering with distributed adaptive control.

**Acknowledgements** This work was supported by National Natural Science Foundation of China (Grant No. 61273221) and National Basic Research Program of China (973) (Grant No. 2014CB845302).

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1 Macchi O. Adaptative Processing: the Least Mean Squares Approach With Applications in Transmission. New York: John Wiley & Sons, Ltd., 1995
2 Sayed A H. Fundamentals of Adaptive Filtering. Hoboken: Wiley-IEEE Press, 2003
3 Haykin S S. Adaptive Filter Theory. Englewood Cliffs: Prentice Hall, 2008
4 Akyildiz I, Su W, Sankarasubramaniam Y, et al. Wireless sensor networks: a survey. Comput Netw, 2002, 38: 393–422
5 Zhang Q, Zhang J F. Distributed parameter estimation over unreliable networks with Markovian switching topologies. IEEE Trans Automat Control, 2012, 57: 2545–2560
6 Kar S, Moura J M F, Poor H V. Distributed linear parameter estimation: asymptotically efficient adaptive strategies. SIAM J Control Optim, 2013, 51: 2200–2229
7 Kar S, Moura J M F. Convergence rate analysis of distributed gossip (linear parameter) estimation: fundamental limits and tradeoffs. IEEE J Sel Top Signal Process, 2011, 5: 674–690
8 Chen W S, Wen C Y, Hua S Y, et al. Distributed cooperative adaptive identification and control for a group of continuous-time systems with a cooperative PE condition via consensus. IEEE Trans Automat Control, 2014, 59: 91–106
9 Cattivelli F S, Sayed A H. Diffusion LMS Strategies for distributed estimation. IEEE Trans Signal Process, 2010, 55: 2069–2084
10 Schizas I D, Mateos G, Giannakis G B. Distributed LMS for consensus-based in-network adaptive processing. IEEE Trans Signal Process, 2009, 8: 2365–2381
11 Stankovic S S, Stankovic M S, Stipanovic D M. Decentralized parameter estimation by consensus based stochastic approximation. IEEE Trans Automat Control, 2014, 56: 531–543
12 Sayed A H. Adaptive networks. Proc IEEE, 2014, 102: 460–497
13 Sayed A H. Adaptation, learning, and optimization over networks. Found Trends Mach Learn, 2014, 7: 311–801
14 Guo L, Chen H F. Identification and Stochastic Adaptive Control. Boston: Birkhauser, 1991
15 Sayed A H, Lopes C G. Adaptive processing over distributed networks. IEICE Trans Fund Electron Commun Comput Sci, 2007, E90-A: 1504–1510
16 Guo L, Ljung L. Performance analysis of general tracking algorithms. IEEE Trans Automat Control, 1995, 40: 1388–1402
17 Guo L. Stability of recursive stochastic tracking algorithms. SIAM J Control Optim, 1994, 32: 1195–1225
18 Guo L, Ljung L. Exponential stability of general tracking algorithms. IEEE Trans Automat Control, 1995, 40: 1376–1387
19 Chen C, Liu Z X, Guo L. Stability of diffusion adaptive filters. In: Proceedings of the 19th World Congress of the International Federation of Automatic Control, Cape Town, 2014. 10409–10414
20 Chen C, Liu Z X, Guo L. Performance analysis of distributed adaptive filters. Commun Inform Syst, 2015, 15: 453–476
21 Xue M, Roy S. Kronecker products of defective matrices: some spectral properties and their implications on observability. In: Proceedings of the 2012 American Control Conference, Montréal, 2012. 5202–5207
22 Saloff-Coste L, Zúñiga J. Convergence of some time inhomogeneous Markov chains via spectral techniques. Stoch Proc Appl, 2007, 117: 961–979
23 Carli R, Chiuso A, Schenato L, et al. Distributed Kalman filtering based on consensus strategies. IEEE J Sel Areas Commun, 2008, 26: 622–633
24 Khan U A, Moura J M F. Distributing the Kalman filter for large-scale systems. IEEE Trans Signal Process, 2008, 56: 4919–4935
25 Olfati-Saber R. Distributed Kalman filtering for sensor networks. In: Proceedings of the 46th Conference on Decision Control, New Orleans, 2007. 5492–5498