Proceedings of the 34th Chinese Control Conference
July 28-30, 2015, Hangzhou, China

Stability of Distributed LMS Under Cooperative Stochastic Excitation

Siyu Xie[1], Lei Guo[1]
1. Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, CAS, Beijing 100190, P. R. China
E-mail: xiesiyu7@gmail.com, lguo@amss.ac.cn

Abstract: Distributed estimation algorithms can estimate a dynamic process of interest by a set of estimators cooperatively, when any individual estimator cannot. The paper considers this kind of problems and focuses on a class of distributed least mean-square(LMS) algorithms in discrete time. Stability analysis will be provided under a weakest possible cooperative stochastic excitation condition. We remark that the stringent conditions such as independency and stationarity that have been used in almost all the existing literatures are no longer used in this paper. In addition, a simulation result comparing with the non-cooperative LMS algorithm will be provided to show the advantage of the distributed LMS algorithm studied in the paper.

Key Words: Distributed LMS algorithm, Cooperative stochastic excitation, L_p-exponentially stability

1 Introduction

It is conceivable that centralized solution of network estimation problem is more accurate in some sense compared with distributed solutions. However, the centralized solution needs the distributed sensors to transmit their information to a central station to estimate the parameters, which means that this method lacks robustness at the central station and needs strong communication capability. Due to the limitation of communications, sensors or stations may only exchange information with their neighbors and this is the main motivation of the development of distributed network estimation algorithms. In addition, distributed algorithms may have better performance than centralized algorithms by optimizing adjacency matrix[1]. However, many distributed estimation algorithms are investigated under the background of distributed adaptive filtering. For example, Sayed and Lopes introduce the incremental strategy to analyze LMS and RLS algorithms and give formulation and performance analysis of this method [2][3][4]. In the meanwhile, Cattivelli et al. introduces a diffusion RLS scheme for the distributed estimation [5]. Sayed et al. introduces distributed LMS algorithm' formulation and stability analysis under independency assumption and low power consumption [2][6][7]. Beyond that, Cattivelli and Saber et al. investigate distributed Kalman filtering strategy and give some performance analysis of the algorithm [8][9][10]. Moreover, Weisheng Chen et al. investigates a class of distributed cooperative adaptive identification algorithms in continuous-time under a deterministic cooperative excitation condition [11].

There are many other researches about distributed filtering algorithms. However, there exist some essential restrictions in the theoretical analysis of stability, for example, they always need stationarity, independency or robustness of the system signals, which, however cannot be guaranteed in many situations. In contrast, the results that we are going to present in this paper do not need the above mentioned properties. Inspired by [7][11][12], we consider another distributed LMS algorithm. The main contribution of this paper is to establish the stability of this kind of distributed LMS algorithm under a weakest possible stability condition, i.e., cooperative stochastic excitation condition. We will prove that under this condition, the distributed adaptive algorithm can accomplish the estimation task cooperatively just like the centralized LMS algorithm, even when any individual sensor cannot realize the estimation task.

This work is organized as follows. After introducing some necessary preliminaries in Section 2, we will present a stability analysis of the ditributed LMS algorithm in Section 3. In Section 4, we give a simulation result and some conclusions are made in Section 5.

2 Preliminaries

2.1 Graph Theory and Kronecker Product

In this paper, we consider a set of n nodes which is distributed throughout the geographic region. In order to describe their interconnections, we model it as a weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ with the set of nodes $\mathcal{V} = \{v_1, v_2, ..., v_n\}$, the set of edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ and the weighted adjacency matrix $\mathcal{A} = (a_{ij})_{n\times n}$ with $\sum_{j=1}^{n} a_{ij} = 1, \forall i = 1, \ldots, n$. We assume that the graph $\mathcal{G}$ is undirected. Node v_i denotes the ith sensor and e_{ij} denotes the connection between sensor i and j. Note that $e_{ij} \in \mathcal{E} \Leftrightarrow a_{ij} > 0$ and we have $a_{ij} = a_{ji}$. The set of neighbors of sensor i is denoted as $\mathcal{N}_i$. The Laplacian matrix $\mathcal{L}$ of the graph $\mathcal{G}$ is defined as $\mathcal{L} = \mathcal{D} - \mathcal{A}$, where $\mathcal{D} = diag\{d_1, d_2, ..., d_n\}$ and $d_i = \sum_{j=1}^{n} a_{ij} = 1$, i.e. $\mathcal{L} = I - \mathcal{A}$.

For the matrix $\mathcal{L}$, there is an basic lemma as follows.

Lemma 1 [13]: The Laplacian matrix $\mathcal{L}$ of the graph $\mathcal{G}$ has at least one zero eigenvalue, with other eigenvalues positive and not more than 2. Furthermore, if the graph $\mathcal{G}$ is connected, then $\mathcal{L}$ has only one zero eigenvalue.

Definition 1 [14]: For matrices $A \in \mathcal{R}^{m\times n}$ and $B \in \mathcal{R}^{p\times q}$, the Kronecker Product of A and B is defined as

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{n1}B & \cdots & a_{mn}B \end{pmatrix} \in \mathcal{R}^{mp\times nq}.$$

We may also need the following result in the next section.

Lemma 2 [14]: Let $\lambda_i(i=1,\ldots,n),\mu_j(j=1,\ldots,m)$ be the eigenvalues of matrix $A\in\mathcal{R}^{n\times n}$ and $B\in\mathcal{R}^{m\times m}$ respectively. Then $\lambda_1\mu_1,\ldots,\lambda_1\mu_m,\lambda_2\mu_1,\ldots,\lambda_2\mu_m,\ldots,\lambda_n\mu_1,\ldots,\lambda_n\mu_m$ are eigenvalues of matrix $A\otimes B$. What's more, if $x_1,\ldots,x_p$ are linearly independent right eigenvectors of A corresponding to $\lambda_1,\ldots,\lambda_p(p\leq n)$ and $y_1,\ldots,y_q$ are linearly independent right eigenvectors of B corresponding to $\mu_1,\ldots,\mu_q(q\leq m)$, then $x_i\otimes y_j$ are linearly independent right eigenvectors of $A\otimes B$ corresponding to $\lambda_i\mu_j$.

2.2 Model

The purpose of this work is to develop a distributed strategy to update the estimation of an $m-$dimensional time-varying parameter vector θ_k in real-time through local interactions between connected sensors. It is usually convenient to denote the iteration of θ_k as follows

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

where γ is a scalar value and ω_k is an yet undefined variable.

Let us consider the following time-varying stochastic linear regression model at sensor $i(i=1,\ldots,n)$

$$y_k^i=(\varphi_k^i)^\tau\theta_k+v_k^i,\quad k\geq 0, \tag{2}$$

where y_k^i,v_k^i are scalar observation and scalar noise, φ_k^i is the $m-$dimensional stochastic regressor of sensor i.

It is well known that many problems from different application areas can be cast as (2) [7][15].

2.3 Algorithm

The normalized least mean square(LMS) algorithm for an individual system is as follows [16]

$$\widehat{\theta}_{k+1}=\widehat{\theta}_k+\mu\frac{\varphi_k}{1+\parallel\varphi_k\parallel^2}(y_k-\varphi_k^\tau\widehat{\theta}_k),\quad k\geq 0 \tag{3}$$

where $\mu\in(0,1]$ is called step size or adaption rate and the increment of the algorithm is opposite to the stochastic gradient of the mean square error(MSE)

$$e_k(\theta)=E(y_k-\varphi_k^\tau\theta)^2.$$

In network systems, if sensor i has access only to the information from its neighbors $\{l\in\mathcal{N}_i\}$, as in literature [7], we can obtain the following normalized recursion for the estimation of θ_k^i

$$\widehat{\theta}_{k+1}^i=\widehat{\theta}_k^i+\frac{\mu_i}{1+\parallel\varphi_k^i\parallel^2}(R_{y\varphi}^i-R_\varphi^i\widehat{\theta}_k^i)-\nu\sum_{l\in\mathcal{N}_i}a_{li}(\widehat{\theta}_k^i-\omega_l). \tag{4}$$

where $R_{y\varphi}^i=E(\varphi_k^iy_k^i),R_\varphi^i=E[\varphi_k^i(\varphi_k^i)^\tau]$, a_{li} is the parameter of weighted adjacency matrix $\mathcal{A}$ and ω_l is an instantaneous approximation of θ_l.

Different from the previous literatures [7][12], we can accomplish the iteration by generating estimates as follows

$$\begin{aligned}R_{y\varphi}^i&\approx\varphi_k^iy_k^i,\\R_\varphi^i&\approx\varphi_k^i(\varphi_k^i)^\tau,\\\omega_l&\approx\widehat{\theta}_k^l.\end{aligned}$$

Consequently, we can obtain the following distributed LMS algorithm

$$\begin{aligned}\widehat{\theta}_{k+1}^i=\widehat{\theta}_k^i&+\mu_i\frac{\varphi_k^i}{1+\parallel\varphi_k^i\parallel^2}[y_k^i-(\varphi_k^i)^\tau\widehat{\theta}_k^i]\\&-\nu\sum_{l\in\mathcal{N}_i}a_{li}(\widehat{\theta}_k^i-\widehat{\theta}_k^l).\end{aligned} \tag{5}$$

For convenience of analysis, we introduce the following notations

$$\begin{aligned}&\Theta_k\triangleq col\{\theta_k,\ldots,\theta_k\}_{n\times1},\widehat{\Theta}_k\triangleq col\{\widehat{\theta}_k^1,\ldots,\widehat{\theta}_k^n\},\\&\widetilde{\Theta}_k\triangleq col\{\widetilde{\theta}_k^1,\ldots,\widetilde{\theta}_k^n\},\ where\ \widetilde{\theta}_k^i=\widehat{\theta}_k^i-\theta_k,\\&\Omega_k\triangleq col\{\omega_k,\ldots,\omega_k\}_{n\times1},V_k\triangleq col\{v_k^1,\ldots,v_k^n\},\\&Y_k\triangleq col\{y_k^1,\ldots,y_k^n\},\Phi_k\triangleq diag\{\varphi_k^1,\ldots,\varphi_k^n\},\\&L_k\triangleq diag\{\frac{\varphi_k^1}{1+\parallel\varphi_k^1\parallel^2},\ldots,\frac{\varphi_k^n}{1+\parallel\varphi_k^n\parallel^2}\},\\&F_k\triangleq L_k\Phi_k^\tau,\\&\Lambda\triangleq diag\{\mu_1I_m,\ldots,\mu_nI_m\}.\end{aligned}$$

By (2), we have

$$Y_k=\Phi_k^\tau\Theta_k+V_k, \tag{6}$$

From (5), we have

$$\widehat{\Theta}_{k+1}=\widehat{\Theta}_k+\Lambda L_k(Y_k-\Phi_k^\tau\widehat{\Theta}_k)-\nu(\mathcal{L}\otimes I_m)\widehat{\Theta}_k, \tag{7}$$

where matrix $\mathcal{L}$ is the Laplacian matrix of graph $\mathcal{G}$. Let us denote $\widetilde{\Theta}_k=\widehat{\Theta}_k-\Theta_k$, we can get

$$\begin{aligned}\widetilde{\Theta}_{k+1}=&\widetilde{\Theta}_k-\Lambda L_k\Phi_k^\tau\widetilde{\Theta}_k-\nu(\mathcal{L}\otimes I_m)\widehat{\Theta}_k\\&+\Lambda L_kV_k-\gamma\Omega_{k+1}.\end{aligned}$$

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

$$\begin{aligned}\widetilde{\Theta}_{k+1}=&\{I_{mn}-[\Lambda F_k+\nu(\mathcal{L}\otimes I_m)]\}\widetilde{\Theta}_k\\&+\Lambda L_kV_k-\gamma\Omega_{k+1},\end{aligned} \tag{8}$$

So far, we have obtained the distributed LMS error equation (8) which will be analyzed in the following section. Before that, we first give some notations and definitions.

2.4 Notations and Definitions

In the sequel, $\mathcal{R}^{m\times n}$ denotes the set of $m\times n$ matrices. Let $A\in\mathcal{R}^{n\times n}$ and $B\in\mathcal{R}^{n\times n}$ be two symmetric matrices, then $A\geq B$ means $A-B$ is a positive semidefinite matrix. $\lambda_{max}(\cdot)$ and $\lambda_{min}(\cdot)$ denote the largest and the smallest eigenvalues of matrix $(\cdot)$ respectively. For any matrix $X\in\mathcal{R}^{m\times n}$, its norm is defined as $\parallel X\parallel=\{\lambda_{max}(XX^\tau)\}^{\frac{1}{2}}$ and for any random matrix A_k, its L_p-norm is defined as $\parallel A_k\parallel_{L_p}=\{E\parallel A_k\parallel^p\}^{\frac{1}{p}}$. Also, $\mathcal{F}_k=\sigma\{\varphi_i^j,\omega_i,v_{i-1}^j,j=1,\ldots,n,i\leq k\}$.

To proceed with further discussions, we need the following definitions.

Definition 2: For a random matrix sequence $\{A_k,k\geq 0\}$ defined on the basic probability space $(\Omega,\mathcal{F},P)$, if $sup_{k\geq0}E\parallel A_k\parallel_{L_p}<\infty$, then $\{A_k\}$ is called L_p-stable.

Definition 3: For a sequence of $d \times d$ random matrices $A = \{A_k, k \geq 0\}$, if it belongs to the following set

$$S_p(\lambda) = \Big\{A : \Big\| \prod_{j=i+1}^{k} (I - A_j) \Big\|_{L_p} \leq M\lambda^{k-i}, \quad \forall k \geq i, \forall i \geq 0, \ for\ some\ M > 0\Big\}, \tag{9}$$

then A is called stably exciting of order $p (p \geq 1)$ with parameter $\lambda \in [0, 1)$. Correspondingly, $I - A$ is called L_p-exponentially stable($p \geq 0$) with parameter $\lambda \in [0, 1)$.

For convenience of discussions, we introduce the following subclass of $S_1(\lambda)$ for a scalar sequence $a = \{a_k, k \geq 0\}$

$$S^0(\lambda) = \Big\{a : a_k \in [0, 1], E \prod_{j=i+1}^{k} (1 - a_j) \leq M\lambda^{k-i}, \quad \forall k \geq i, \forall i \geq 0, \ for\ some\ M > 0\Big\}, \tag{10}$$

where $\lambda \in [0, 1)$. In addition

$$S^0 \triangleq \bigcup_{\lambda \in (0,1)} S^0(\lambda).$$

3 Stability Analysis

In this section, we study the stability of the error equation (8).

3.1 Conditions

To proceed with the analysis, we need the following two conditions.

Condition 1: The graph $\mathcal{G}$ is connected.

Condition 2 (Cooperative Stochastic Exctation) : Let $\{\varphi_k^i, \mathcal{F}_k, k \geq 0\}, i = 1, \ldots, n$, be n adapted sequences and there exists an integer $h > 0$ such that $\{\lambda_k, k \geq 0\} \in S^0(\lambda)$ for some $\lambda \in (0, 1)$, where λ_k is defined by

$$\lambda_k \triangleq \lambda_{min}\Big\{E\Big[\frac{1}{n(h+1)} \sum_{i=1}^{n} \sum_{j=k+1}^{k+h} \frac{\varphi_j^i (\varphi_j^i)^\tau}{1 + \|\varphi_j^i\|^2} \Big| \mathcal{F}_k\Big]\Big\}. \tag{11}$$

We should notice that the above cooperative stochastic excitation condition is a very general condition for the algorithm. When none of the sensors satisfies the excitation condition needed for the traditional single LMS algorithm as in [17], the whole system can still satisfy Condition 2. In other words, these sensors can work in a cooperative way to fulfill their task even when any individual sensor cannot. Moreover, this condition is arguably the weakest possible excitation condition (cf. [17][18][19]).

3.2 Stability Theorem

We are now in a position to give our main result.

Theorem 1: Consider the signed model (2) and suppose that Conditions 1 and 2 are satisfied. Then for any $\Lambda \geq 0, \nu \geq 0$ and $0 \leq \Lambda + 2\nu I_{mn} \leq I_{mn}$, the coefficient matrix sequence of the homogeneous part of (8), i.e., $\{I_{mn} - [\Lambda F_k + \nu(\mathcal{L} \otimes I_m)]\}$ is L_p-exponentially stable $(p \geq 1)$.

Before proving the theorem, we first list and prove some lemmas.

Lemma 3 [17]: If the sequences α_k and β_k satisfy $0 \leq \alpha_k \leq \beta_k \leq 1$ and $\{\alpha_k\} \in S^0(\lambda)$, then $\{\beta_k\} \in S^0(\lambda)$.

Lemma 4 [17]: Let $\{\alpha_k\} \in S^0(\lambda)$ and $\alpha_k \leq \alpha^* < 1$ where α^* is a constant, then $\forall \epsilon \in (0, 1)$, $\{\epsilon\alpha_k\} \in S^0(\lambda^{(1-\alpha^*)\epsilon})$.

Lemma 5 [17]: Let $\{A_k^i, \mathcal{F}_k\}, i = 1, \ldots, n$, be n adapted sequences of random matrices and $0 \leq A_k^i \leq I, \forall i = 1, \ldots, n$ and denote $A_k = diag\{A_k^1, \ldots, A_k^n\}$. If there existes an integer $h > 0$ such that $\{\rho_k\} \in S^0(\rho), \rho \in (0, 1)$ where

$$\rho_k \triangleq \lambda_{min}\Big\{E\Big[\frac{1}{1+h} \sum_{j=k+1}^{k+h} A_j \Big| \mathcal{F}_k\Big]\Big\}, \tag{12}$$

then $\{A_k\} \in S_p(\rho^\alpha)$, where

$$\alpha = \begin{cases} \frac{1}{8h(1+h)^2}, & 1 \leq p \leq 2 \\ \frac{1}{4h(1+h)^2 p}, & p > 2. \end{cases}$$

Lemma 6: For any $\Lambda \geq 0, \nu \geq 0$ and $0 \leq \Lambda + 2\nu I_{mn} \leq I_{mn}$, we have

$$0 \leq \Lambda F_k + \nu(\mathcal{L} \otimes I_m) \leq I_{mn}.$$

Proof. By Condition 1, matrix $\mathcal{L}$ has n real eigenvalues in an ascending order

$$0 = t_1 < t_2 \leq t_3 \leq \cdots \leq t_n \leq 2.$$

Correspondingly, $\mathcal{L} \otimes I_m$ has mn real eigenvalues in an ascending order

$$0 = l_1 = l_2 = \cdots = l_m < l_{m+1} \leq l_{m+2} \leq \cdots \leq l_{mn} \leq 2.$$

Then the matrix $\mathcal{L} \otimes I_m$ is positive semidefinite and $\mathcal{L} \otimes I_m \leq 2I_{mn}$.

Since $\Lambda \geq 0, \nu \geq 0$ and $0 \leq \Lambda + 2\nu I_{mn} \leq I_{mn}$, the conclusion holds. □

Now, we give the most critical lemma of the paper.

Lemma 7: Suppose the conditions in Theorem 1 are satisfied, then $\rho_k \in S^0(\rho)$, where

$$\rho_k \triangleq \lambda_{min}\Big\{E\Big[\frac{1}{1+h} \sum_{j=k+1}^{k+h} [\Lambda F_j + \nu(\mathcal{L} \otimes I_m)] \Big| \mathcal{F}_k\Big]\Big\}. \tag{13}$$

and $\rho = \lambda^\epsilon$ for some constant $\epsilon > 0$.

Proof. According to Condition 1, $\mathcal{L}$ has only one zero eigenvalue whose unit eigenvector is $\frac{1}{\sqrt{n}}(1, \ldots, 1)^\tau$, i.e., $\frac{1}{\sqrt{n}}\mathbf{1}$ where $\mathbf{1} = (1, \ldots, 1)^\tau$. Correspondingly, $\mathcal{L} \otimes I_m$ has m zero eigenvalues whose orthogonal unit eigenvectors are

$$\xi_1 = \frac{1}{\sqrt{n}}\mathbf{1} \otimes e_1, \ldots, \xi_m = \frac{1}{\sqrt{n}}\mathbf{1} \otimes e_m,$$

where e_i is a unit column vector with the ith element is 1. The other eigenvalues of $\mathcal{L} \otimes I_m$ are $l_{m+1} \leq \cdots \leq l_{mn}$ whose orthogonal unit eigenvectors are denoted as $\xi_{m+1}, \ldots, \xi_{mn}$ correspondingly. Here, for an arbitrary unit vector $\eta \in \mathcal{R}^{mn}$, it can be expressed as

$$\eta = \sum_{j=1}^{m} x_j \xi_j + \sum_{j=m+1}^{mn} x_j \xi_j \triangleq \eta_1 + \eta_2,$$

where $\sum_{j=1}^{m} x_j^2 + \sum_{j=m+1}^{mn} x_j^2 = 1$. Now, let [11]

$$\begin{aligned} H_k^i &= \sum_{j=k+1}^{k+h} \frac{\varphi_j^i (\varphi_j^i)^\tau}{1 + \|\varphi_j^i\|^2} \\ H_k &= diag\{H_k^1, \ldots, H_k^n\}. \end{aligned}$$

By the definition of F_j, we have

$$\begin{aligned} \Delta_k &\triangleq E\bigg[\frac{1}{1+h} \sum_{j=k+1}^{k+h} [\Lambda F_j + \nu(\mathcal{L} \otimes I_m)] \Big| \mathcal{F}_k\bigg] \\ &= E\bigg\{\frac{1}{1+h}[\Lambda H_k + h\nu(\mathcal{L} \otimes I_m)] \Big| \mathcal{F}_k\bigg\}, \end{aligned} \tag{14}$$

Note also that

$$\begin{aligned} \Gamma_k &\triangleq E\bigg[\frac{1}{n(h+1)} \sum_{i=1}^{n} \sum_{j=k+1}^{k+h} \frac{\varphi_j^i (\varphi_j^i)^\tau}{1 + \|\varphi_j^i\|^2} \Big| \mathcal{F}_k\bigg] \\ &= E\bigg[\frac{1}{n(h+1)} \sum_{i=1}^{n} H_k^i \Big| \mathcal{F}_k\bigg]. \end{aligned} \tag{15}$$

By (14), let us consider the following form

$$\begin{aligned} &\eta^\tau \Delta_k \eta = (\eta_1 + \eta_2)^\tau \Delta_k (\eta_1 + \eta_2) \\ =&\eta_1^\tau E\bigg[\frac{1}{1+h}\Lambda H_k | \mathcal{F}_k\bigg]\eta_1 + \eta_2^\tau E\bigg[\frac{1}{1+h}\Lambda H_k | \mathcal{F}_k\bigg]\eta_2 \\ &+ 2\eta_1^\tau E\bigg[\frac{1}{1+h}\Lambda H_k | \mathcal{F}_k\bigg]\eta_2 \\ &+ \eta_1^\tau \bigg[\frac{h}{1+h}\nu(\mathcal{L} \otimes I_m)\bigg]\eta_1 + \eta_2^\tau \bigg[\frac{h}{1+h}\nu(\mathcal{L} \otimes I_m)\bigg]\eta_2 \\ &+ 2\eta_1^\tau \bigg[\frac{h}{1+h}\nu(\mathcal{L} \otimes I_m)\bigg]\eta_2 \\ \triangleq& S_1 + S_2 + S_3 + S_4 + S_5 + S_6. \end{aligned} \tag{16}$$

Since

$$0 \leq E\bigg[\frac{1}{1+h}\Lambda H_k | \mathcal{F}_k\bigg] \leq I_{mn},$$

we can decompose it as follows

$$\begin{aligned} &E\bigg[\frac{1}{1+h}\Lambda H_k | \mathcal{F}_k\bigg] \\ =&\bigg\{E\bigg[\frac{1}{1+h}\Lambda H_k | \mathcal{F}_k\bigg]\bigg\}^{\frac{1}{2}} \bigg\{E\bigg[\frac{1}{1+h}\Lambda H_k | \mathcal{F}_k\bigg]\bigg\}^{\frac{1}{2}} \\ \triangleq& M_k \cdot M_k, \end{aligned}$$

and $M_k \geq 0$.

For matrices ζ_1 and ζ_2, We have following inequality

$$2\zeta_1^\tau \zeta_2 \leq \delta \zeta_1^\tau \zeta_1 + \frac{1}{\delta} \zeta_2^\tau \zeta_2. \tag{17}$$

where $\delta > 0$ can be any constant. Let

$$\begin{aligned} \zeta_1 &\triangleq -M_k \eta_1 \\ \zeta_2 &\triangleq M_k \eta_2 \end{aligned}$$

and substituting this into (17), it is easy to have

$$\begin{aligned} &2\zeta_1^\tau \zeta_2 \\ =&-2\eta_1^\tau E\bigg[\frac{1}{1+h}\Lambda H_k | \mathcal{F}_k\bigg]\eta_2 \\ \leq& \delta \zeta_1^\tau \zeta_1 + \frac{1}{\delta}\zeta_2^\tau \zeta_2 \\ =&\delta \eta_1^\tau E\bigg[\frac{1}{1+h}\Lambda H_k | \mathcal{F}_k\bigg]\eta_1 + \frac{1}{\delta}\eta_2^\tau E\bigg[\frac{1}{1+h}\Lambda H_k | \mathcal{F}_k\bigg]\eta_2. \end{aligned} \tag{18}$$

Then we can obtain

$$\begin{aligned} S_3 =& 2\eta_1^\tau E\bigg[\frac{1}{1+h}\Lambda H_k | \mathcal{F}_k\bigg]\eta_2 \\ \geq& -\delta \eta_1^\tau E\bigg[\frac{1}{1+h}\Lambda H_k | \mathcal{F}_k\bigg]\eta_1 \\ &- \frac{1}{\delta}\eta_2^\tau E\bigg[\frac{1}{1+h}\Lambda H_k | \mathcal{F}_k\bigg]\eta_2 \\ =& -\delta S_1 - \frac{1}{\delta} S_2. \end{aligned} \tag{19}$$

From (16) and (19), it is obvious that

$$\eta^\tau \Delta_k \eta \geq (1-\delta) S_1 + (1 - \frac{1}{\delta}) S_2 + S_4 + S_5 + S_6. \tag{20}$$

Now we will estimate S_1, S_2, S_4, S_5 and S_6, respectively.

Note that

$$\begin{aligned} S_1 =& \frac{1}{1+h} E[\eta_1^\tau \Lambda H_k \eta_1 | \mathcal{F}_k] \\ =& \frac{1}{1+h} E\bigg[\bigg(\sum_{j=1}^{m} x_j \xi_j\bigg)^\tau \Lambda H_k \bigg(\sum_{j=1}^{m} x_j \xi_j\bigg) \Big| \mathcal{F}_k\bigg] \\ =& \frac{1}{1+h} E[X^\tau \xi^\tau \Lambda H_k \xi X | \mathcal{F}_k], \end{aligned} \tag{21}$$

where $X = [x_1, \ldots, x_m]^\tau, \xi = [\xi_1, \ldots, \xi_m]$.

Since

$$\xi = \frac{1}{\sqrt{n}} \begin{pmatrix} e_1 & e_2 & \cdots & e_m \\ e_1 & e_2 & \cdots & e_m \\ \vdots & \vdots & \ddots & \vdots \\ e_1 & e_2 & \cdots & e_m \end{pmatrix}_{n \times m}.$$

it is easy to show that

$$\begin{aligned} H_k \xi =& \frac{1}{\sqrt{n}} \begin{pmatrix} H_k^1 e_1 & H_k^1 e_2 & \cdots & H_k^1 e_m \\ H_k^2 e_1 & H_k^2 e_2 & \cdots & H_k^2 e_m \\ \vdots & \vdots & \ddots & \vdots \\ H_k^n e_1 & H_k^n e_2 & \cdots & H_k^n e_m \end{pmatrix} \\ =& \frac{1}{\sqrt{n}} \begin{pmatrix} H_k^1 & H_k^2 & \cdots & H_k^n \end{pmatrix}^\tau. \end{aligned}$$

Similarly, we can obtain

$$\xi^\tau H_k \xi = \frac{1}{n} \begin{pmatrix} e_1^\tau H_k^1 + e_1^\tau H_k^2 + \cdots + e_1^\tau H_k^n \\ e_2^\tau H_k^1 + e_2^\tau H_k^2 \cdots + e_2^\tau H_k^n \\ \vdots \\ e_m^\tau H_k^1 + e_m^\tau H_k^2 + \cdots + e_m^\tau H_k^n \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n} H_k^i.$$

From this, we have

$$E[\xi^\tau H_k \xi | \mathcal{F}_k] = E\left[\frac{1}{n}\sum_{i=1}^{n} H_k^i \Big| \mathcal{F}_k\right] = (h+1)\Gamma_k. \quad (22)$$

Substituting (22) into (21), it can be deduced that

$$\begin{aligned} S_1 &\geq \mu_{min} E[X^\tau \Gamma_k X | \mathcal{F}_k] \\ &\geq \mu_{min}\lambda_k \sum_{j=1}^{m} x_j^2, \end{aligned} \quad (23)$$

where μ_{min} is the minimum diagonal element of matrix Λ.

Notice that

$$|S_2| \leq \|\eta_2\|^2 = \sum_{j=m+1}^{mn} x_j^2. \quad (24)$$

Since $\eta_1 = \sum_{j=1}^{m} x_j \xi_j$ and $\xi_j (1 \leq j \leq m)$ is the eigenvector corresponding to the zero eigenvalue, we have

$$S_4 = S_6 = 0. \quad (25)$$

For S_5, we know that

$$\begin{aligned} S_5 \geq & \frac{h}{1+h}\nu \sum_{j=m+1}^{mn} l_j x_j^2 \\ \geq & \frac{h}{1+h}\nu l_{m+1} \sum_{j=m+1}^{mn} x_j^2. \end{aligned} \quad (26)$$

Denote

$$y = \sum_{j=1}^{m} x_j^2.$$

By (20) and since $\rho_k = \lambda_{min}(\Delta_k)$, we have

$$\begin{aligned} \rho_k \geq & (1-\delta)\mu_{min}\lambda_k y - (1 - \frac{1}{\delta})(1-y) \\ & + \frac{h}{1+h}\nu l_{m+1}(1-y) \\ = & \left[(1-\delta)\mu_{min}\lambda_k - (\frac{h}{1+h}\nu l_{m+1} - 1 + \frac{1}{\delta})\right] y \\ & + \frac{h}{1+h}\nu l_{m+1} - 1 + \frac{1}{\delta}, \quad y \in [0,1]. \end{aligned} \quad (27)$$

Now, denoting

$$\frac{h}{1+h}\nu l_{m+1} = b$$

and setting

$$(1-\delta)\mu_{min}\lambda_k = b - 1 + \frac{1}{\delta},$$

we have

$$\mu_{min}\lambda_k \delta^2 - (1 - b + \mu_{min}\lambda_k)\delta + 1 = 0.$$

It is evident that

$$\delta_1 = \frac{(1-b+\mu_{min}\lambda_k) + \sqrt{(1-b+\mu_{min}\lambda_k)^2 - 4\mu_{min}\lambda_k}}{2\mu_{min}\lambda_k}$$

$$\delta_2 = \frac{(1-b+\mu_{min}\lambda_k) - \sqrt{(1-b+\mu_{min}\lambda_k)^2 - 4\mu_{min}\lambda_k}}{2\mu_{min}\lambda_k}$$

where $\delta_{1,2} > 0$. Thus we choose $\delta = \delta_1$ and have

$$\begin{aligned} & \rho_k \\ \geq & (1-\delta)\mu_{min}\lambda_k \\ = & \frac{(\mu_{min}\lambda_k + b - 1) - \sqrt{(\mu_{min}\lambda_k + b - 1)^2 - 4\mu_{min} b\lambda_k}}{2} \\ = & \frac{4\mu_{min} b \lambda_k}{2[(\mu_{min}\lambda_k + b - 1) + \sqrt{(\mu_{min}\lambda_k + b - 1)^2 - 4\mu_{min} b\lambda_k}]} \\ \geq & \frac{4\mu_{min} b\lambda_k}{2[(1+1-1) + \sqrt{(1+1-1)^2}]} \\ = & \mu_{min} b \lambda_k. \end{aligned} \quad (28)$$

In conclusion, we have

$$\begin{aligned} \rho_k &\geq b\mu_{min}\lambda_k \\ &= \frac{h}{1+h} l_{m+1}\nu\mu_{min}\lambda_k. \end{aligned}$$

By Lemmas 4 and 5 and since $\lambda_k \in [0, \alpha^*], \alpha^* = \frac{h}{h+1}$, we have

$$\{\rho_k\} \in S^0(\rho),$$

where $\rho = \lambda^\epsilon$ and

$$\epsilon = \frac{h}{h^2 + 2h + 1} \cdot l_{m+1}\nu\mu_{min} > 0.$$

□

Now we present the proof of Theorem 1.

Proof. From Lemmas 5 to 7, we know that

$$\{\Lambda F_j + \nu(\mathcal{L} \otimes I_m)\} \in S_p(\rho^\alpha),$$

where α is defined in Lemma 5. Then by Definition 3, it is obvious that $\{I_{mn} - [\Lambda F_j + \nu(\mathcal{L} \otimes I_m)]\}$ is L_p-exponentially stable $(p \geq 1)$. □

4 Simulation

Following [12], let us take $n = 3$ with the following adjacency matrix

$$\mathcal{A} = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 1/3 & 1/6 \\ 1/2 & 1/6 & 1/3 \end{pmatrix},$$

then the corresponding graph is connected. Let $\gamma = 0.5, \omega_k \sim N(0, 0.1)$ in (1), $v_k^i \sim N(0, 0.1)$ in (2) and $\varphi_k^i (i = 1, 2, 3)$ generated by

$$\begin{cases} x_k^i = & A_i x_{k-1}^i + B_i \xi_k^i, \\ \varphi_k^i = & C_i x_k^i + \varsigma_k^i, \end{cases}$$

where $\xi_k^i \sim N(0,1), \varsigma_k^i \sim N(0,1)$ and

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

$$B_1 = (1,0,0)^\tau, B_2 = (0,1,0)^\tau, B_3 = (0,0,1)^\tau,$$

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

Let $x_0^1 = x_0^2 = x_0^3 = (0,0,0)^\tau$, then the mean square errors(MSEs) are showing in Fig.1.

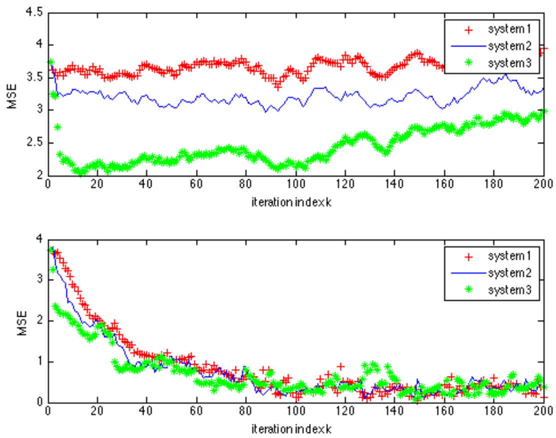

Fig. 1: Mean square errors of three sensors

In the above figure, the first one is the non-cooperative LMS algorithm in which the MSEs of three sensors cannot converge to zero because all the sensors don't satisfy the excitation condition in [17] and the other one is the distributed LMS algorithm studied in the paper in which all the MSEs converge to a small neighborhood of zero, since the whole system satisfies Condition 2.

5 Conclusion

In this paper, we have investigated a class of distributed LMS algorithms and proved that such algorithms can fulfil the estimation task under a very general cooperative stochastic excitation condition, even when any individual sensor or subsystem cannot. Of course, there are still a number of interesting problems for further research, for example,

- To give the performance analysis of the distributed LMS algorithm and compare it with centralized LMS algorithm.
- To extend the stability analysis to other distributed filtering algorithms, e.g., Kalman filtering and RLS algorithms.
- To explore real applications of the distributed LMS algorithm, since it has nice properties.

References

[1] X. Zhao and A. H. Sayed, Performance limits for distributed estimation over LMS adaptive networks, *IEEE. Trans. Sig. Proc.*, 60(10): 5107–5124, 2012.

[2] A. H. Sayed and C. G. Lopes, Adaptive Processing over distributed networks, *IEICE Trans. Fund. Electron.,Commun. Comput. Sci.*, E90-A(8): 1504–1510, 2007.

[3] C. G. Lopes and A. H. Sayed, Distributed adaptive incremental strategies: Formulation and performance analysis, *inProc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Toulouse, France, vol. 3: 584–587, 2006.

[4] C. G. Lopes and A. H. Sayed, Incremental adaptive strategies over distributed networks, *IEEE Trans. on Signal Process.*, 55(8): 4064–4077, 2007.

[5] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, A diffusion RLS scheme for distributed estimation over adaptive networks, in *Proc. IEEE SPAWC*, Helsinki, Finland, 1–5, 2007.

[6] I. D. Schizas, G. Mateos, and G. B. Giannakis, Distributed LMS for consensusbased in-network adaptive processing, *IEEE Trans. Signal Process.*, 8(6): 2365–2381, 2009.

[7] F. S. Cattivelli and A. H. Sayed, Difussion LMS strategies for distributed estimation, *IEEE Trans. on Signal Process*, 58(3): 1035–1048, 2010.

[8] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, Diffusion strategies for distributed Kalman filtering: Formulation and performance analysis, in *Proc. Workshop on Cognitive Inf. Process.*, Santorini, Greece, 2008.

[9] R. Carli, A. Chiuso, L. Schenato, et al., Distributed Kalman filtering based on consensus strategies, *Selected Areas in Communications, IEEE Journal on*, 26(4): 622–633, 2008.

[10] R. O. Saber, Distributed Kalman Filtering for Sensor Networks, in *Proceedings of the 46th Conference on Decision Control*, New Orleans, LA, 5492–5498, 2007.

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

[12] C. Chen, Z. X. Liu and L. Guo, Stability of diffusion adaptive filters, *Proc. IFAC world congress*, Cape Town, 2014.

[13] R. P. Agaev and P. Y. Chebotarev, The matrix of maximum out forests of a digraph and its applications, *Autom. Remote Control*, vol.61, no.9, pp.1424–1450, 2000.

[14] A. J. Laub, Matrix analysis for scientists and Engineers, Philadelphia, PA, USA: SIAM, 2005.

[15] L. Ljung and T. Soderstrom, Theory and practice of recursive identification, MIT Press, Cambridge, MA, 1983.

[16] B. Widerow and S. Stearns, Adaptive signal processing, Prentice-Hall Englewood Cliffs NJ, 1985.

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

[18] L. Guo and L. Ljung, Performance analysis of general tracking algorithms, *IEEE Trans. on Automatic Control*, 40(8): 1388-1402, 1995.

[19] L. Guo, L. Ljung and G. J. Wang, Necessary and sufficient conditions for stability of LMS, *IEEE Trans. on Automatic Control*, 42(6) : 761–770, 1997.