Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Stability of Diffusion Adaptive Filters ⋆

Chen Chen * Zhixin Liu * Lei Guo *

* *Key Laboratory of Systems and Control, Institute of Systems Science, AMSS, Chinese Academy of Sciences, Beijing, 100190, China (e-mails: chenchen@amss.ac.cn, lzx@amss.ac.cn, lguo@amss.ac.cn)*

Abstract: In this paper, we consider the diffusion adaptive filters where a set of sensors is required to collectively estimate time-varying signals (or parameters) from noisy measurements in a way of information diffusion. We will establish the stability of the diffusion least mean square (DLMS) algorithm, without requiring stationarity, independency, and boundedness assumptions of the system signals, which means that our results can be applied to more general and practical class of stochastic systems than those studied in the literature. We will present theoretical results concerning stability and bounds on the mean square error(MSE)of the filtering. We will also show that the network of sensors can cooperate to guarantee the stability of the filtering, even though any single sensor does not have such a capability. This clearly reveals the advantages of the DLMS algorithm vs. standard least mean square (LMS) algorithm. Numerical simulations will also be presented to support the theoretical justifications.

1. INTRODUCTION

In the last decade, sensor networks have attracted much attention from researchers and are widely used in various engineering areas including communications, signal processing, controls, robotics and computer technology. Generally speaking, sensor networks are composed of many spatially distributed sensors which are used to collect and deal with the information. In such a context, more information from more regions can be utilized, while new problems arise that one central processors might not handle such a large amount of data. To solve the problem, the in-network processing has been widely studied since it has some advantages in robustness to failure, decreasing computation and reducing network congestion.

It is well-known that parameter estimation or adaptive filtering has been a central issue in the area of control and identification. With the development of the wireless communication networks, how to design the in-network distributed estimation or distributed adaptive filtering algorithm becomes more and more important. In recent years, some distributed adaptive filtering algorithms have been proposed inspired by different motivations. For instance, the authors in [1] and [2] proposed the incremental RLS and the incremental LMS algorithms, where the information at each sensor is circulated through a topological cycle. In [3] and [4], diffusion LMS and diffusion RLS were proposed, where the estimates at each sensor are diffused to its neighbors. In addition to this, other distributed algorithms have also been developed and the corresponding convergence and performance analysis are given, see [5]-[11], [13] and [14] among many others.

To the best of our knowledge, almost all existing results on the distributed adaptive filtering need the independency or strictly stationarity assumptions of the systems signals. In [3] and [10], the authors assume that regression vectors are spatially and temporally independent. In [9], the authors require that the observation matrices at time t are independent of the δ-filed $\mathcal{F}_{t-1}$; In [8], the regression vectors are assumed to be strictly stationary and ergodic. However, in many practical situations, the independency or strictly stationarity assumptions cannot be satisfied. So a naturally important issue is: can we establish stability and conduct performance analysis for distributed adaptive filters for systems with correlated and non-stationary signals ? On the other hand, for the stability of the distributed LMS, most of the proposed conditions in the existing literature, essentially require that each sensor has the ability to guarantee the stability of the LMS algorithm when there is no information exchange between sensors[3][8]. Hence, the superiority of the distributed strategies on cooperation need to be further explored.

In this paper, we focus on the analysis of the diffusion adaptive filter (DLMS) developed in [3], where the signals are generated by a linear time-varying stochastic regression model. We will provide sufficient conditions for the stability and performance analysis of the filtering, which can be regarded as a joint excitation condition of the filtering network. Aiming at relaxing the limitations of the existing theoretical results as mentioned in the last paragraph, we will establish more general theoretical results in the current paper with the following two main features: (i) Different from most existing work, we do not require any independency or stationarity assumptions on the system signals, which means that a more general class of stochastic models can be included. To this end, we need to investigate the product of random matrices, which is the root of the problem. (ii) Compared with the standard LMS algorithm, we find that the sensors in DLMS can cooperate to fulfill the estimation or filtering task, even though each single sensor cannot. This finding clearly displays the advantage

⋆ This work is supported by National Natural Science Foundation (NNSF) of China under Grants 61203141 and 61273221 and the National Key Basic Research Program of China (973 program) under Grant No.2014CB845301/2/3.

of the distributed algorithms, which is rarely mentioned in the existing literature.

The remainder of this article is organized as follows. In Section 2, we present the DLMS and the main results of this paper. In Section 3, we provide the related lemmas and outline the proof of Theorem 2.1. In Section 4, numerical simulations are presented to illustrate the theoretical justifications. Concluding remarks are made in Section 5.

The following notations are used throughout the paper. I_n means $n \times n$-dimensional identity matrix. Operators $(\cdot)', diag(\cdot), \lambda_{max}(\cdot), \lambda_{min}(\cdot)$ denote transpose, diagonal matrix, the largest eigenvalue, the smallest eigenvalue, respectively. $col(\cdots)$ stands for a vector by stacking the specified vectors. $\otimes$ denote matrix Kronecker product. A matrix $A \geq 0$ means A is semi-positive definite, $A \geq B$ means $A - B \geq 0$. We say that the matrix A is stochastic if each element of A is non-negative and the sum of each row equals to1. Furthermore, a matrix is called doubly stochastic if it is a stochastic matrix and the sum of each column is also 1. For any vector x, $\|x\|$ represents the Euclidean norm of x. For any matrix X, $\|X\| = \{\lambda_{max}(XX')\}^{\frac{1}{2}}$.

2. DLMS & MAIN RESULTS

Consider a network comprising of N sensors where only single-hop communication is allowed, i.e., sensor i can only communicate with the sensors in its neighborhood $\mathcal{N}_i \subset \{1, \cdots, N\}$. We use graph $\mathcal{G} = \{V, E\}$ to describe the relationship between sensors, where the vertices are the sensors and edge $(i, j) \in E$ if sensor j is one of the neighbors of i. For convention of analysis, we assume that the graph $\mathcal{G}$ is undirected and contain self-loops, that is, $i \in \mathcal{N}_i, \forall i$. An $N \times N$ matrix $A = \{a_{ij}\}$ is introduced to represent the weights of links, where $a_{ij} > 0$ if and only if $j \in \mathcal{N}_i$ and $\sum_{j=1}^{N} a_{ij} = 1, \forall i$.

The task of the network of sensors is to estimate a sequence of M dimensional time-varying parameter vectors $\{\theta_k, k = 1, 2, \cdots\}$, where the parameter variation at time k is denoted by $\omega_k = \theta_k - \theta_{k-1}$. We assume that the signal $\{y_k^i, \varphi_k^i\}$ that the sensor $i(i = 1, 2, \cdots, N)$ receives obeys the following time-varying stochastic linear regression model:

$$y_k^i = (\varphi_k^i)'\theta_k + v_k^i, \tag{1}$$

where y_k^i and v_k^i are scalar observation and noise at node i, respectively, and φ_k^i is the M-dimensional stochastic regression vector.

At each time step $k \geq 0$, each sensor updates its estimates by using the estimates of its neighbors. In this paper, the sensor $i(i = 1, 2 \cdots, N)$ will adopt the following diffusion LMS (DLMS) algorithm:

$$\begin{cases} \vartheta_k^i = \sum_{j=1}^{N} a_{ij}\hat{\theta}_k^j; \\ \hat{\theta}_{k+1}^i = \vartheta_k^i + \mu_i \dfrac{\varphi_k^i}{1 + \|\varphi_k^i\|^2}(y_k^i - (\varphi_k^i)'\vartheta_k^i), \end{cases} \tag{2}$$

where $0 < \mu_i < 1$ is the step size of i and the initial estimates $\hat{\theta}_1^i, i = 1, \cdots, N$ are arbitrary values.

Remark 2.1. It is worth mentioning that the above DLMS is first introduced in [1][3]. In the above DLMS algorithm, all sensors will estimate the time-varying parameters at the same time, which can reduce the complexity and save communications.

The objective of this paper is to establish the exponentially stability and performance analysis of the DLMS (2).

To proceed our analysis, we introduce the following global quantities:

$$\begin{aligned}
&\Theta_k \triangleq col\{\underbrace{\theta_k, \cdots, \theta_k}_{N}\}, \quad W_k = col\{\underbrace{\omega_k, \cdots, \omega_k}_{N}\}, \\
&Y_k \triangleq col\{y_k^1, \cdots, y_k^N\}, \quad \vartheta_k \triangleq col\{\vartheta_k^1 \cdots \vartheta_k^N\}, \\
&\Psi_k \triangleq diag\{\varphi_k^1, \cdots, \varphi_k^N\}, \quad \hat{\Theta}_k \triangleq col\{\hat{\theta}_k^1, \cdots, \hat{\theta}_k^N\}, \\
&\widetilde{\Theta}_k \triangleq col\{\widetilde{\theta}_k^1, \cdots, \widetilde{\theta}_k^N\} \text{ with } \widetilde{\theta}_k^i = \widehat{\theta}_k^i - \theta_k^i \\
&V_k = col\{v_k^1, \cdots, v_k^N\}, \\
&D_k \triangleq diag\{\frac{\mu_1}{1 + \|\varphi_k^1\|^2}I_M, \cdots, \frac{\mu_N}{1 + \|\varphi_k^N\|^2}I_M\}
\end{aligned}$$

Then (1) can be written as

$$Y_k = \Psi_k'\Theta_k + V_k. \tag{3}$$

Correspondingly, equation (2) can be written as

$$\begin{cases} \vartheta_k = (A \otimes I_M)\hat{\Theta}_k, \\ \hat{\Theta}_{k+1} = \vartheta_k + D_k\Psi_k(Y_k - \Psi_k'\vartheta_k), \end{cases} \tag{4}$$

or, in a more compact form:

$$\hat{\Theta}_{k+1} = (A \otimes I_M)\hat{\Theta}_k + D_k\Psi_k(Y_k - \Psi_k'(A \otimes I_M)\hat{\Theta}_k). \tag{5}$$

Subtracting Θ_k from both sides of (5) and using the fact that $(A \otimes I_M)\Theta_k = \Theta_k$, we obtain

$$\begin{aligned} \widetilde{\Theta}_{k+1} = (I_{MN} - D_k\Psi_k\Psi_k') \cdot (A \otimes I_M)\widetilde{\Theta}_k \\ + D_k\Psi_kV_k + W_{k+1}. \end{aligned} \tag{6}$$

To state the main result of this paper, we will introduce some definitions first.

Definition 2.1. ([17]). A random matrix (or vector) sequence $\{A_k, k \geq 0\}$ defined on the basic probability space $(\Omega, \mathcal{F}, P)$ is called L_p stable$(p > 0)$ if $\sup_{k \geq 0} E\|A_k\|^p < \infty$.

In the sequel, we will refer to $\|A_k\|_{L_p}$ defined by $\|A_k\|_{L_p} \triangleq \{E\|A_k\|^p\}^{\frac{1}{p}}$ as the p norm of A_k. In order to obtain the stability of $\widetilde{\Theta}_k$, some conditions on the coefficient matrices of the homogeneous part of (6) are required. To this end, we give the following definition:

Definition 2.2. ([17]). A sequence of $d \times d$ random matrices $A = \{A_k, k \geq 0\}$ is called L_p-exponentially stable$(p \geq 1)$ with parameter $\lambda \in [0, 1)$, if it belong to the following set

$$\begin{aligned} S_p(\lambda) = \Big\{A : \|\prod_{j=i+1}^{k} A_j\|_{L_p} \leq M\lambda^{k-i}, \forall k \geq i, \\ \forall i \geq 0, \text{for some M} > 0\Big\}. \end{aligned} \tag{7}$$

Condition 2.1. The graph $\mathcal{G}$ is undirected and connected, and the weighted matrix A is symmetric.

Remark 2.2. From the construction of A, A is doubly stochastic with $a_{ii} > 0$.

Condition 2.2. $\{\varphi_k^i, \mathcal{F}_k\}, i = 1, \cdots, N$, are N adapted sequences, and $\lambda_k \in S^o$, where

$$\lambda_k \triangleq \lambda_{min}\left\{E\left[\frac{1}{2N}\sum_{i=1}^{N}\frac{\varphi_{k+1}^i(\varphi_{k+1}^i)'}{1+\|\varphi_{k+1}^i\|^2} \mid \mathcal{F}_k\right]\right\}, \quad (8)$$

and

$$S^o = \Big\{a = (a_i) : a_i \in [0,1], \text{ and there exist constants } \lambda \in (0,1) \text{ and } M > 0 \text{ such that } E\prod_{j=i+1}^{k}(1-a_j) \le M\lambda^{k-i}.\Big\}$$

Remark 2.3. If there exists a constant $\underline{\lambda} \in (0,1)$ such that $\lambda_k \ge \underline{\lambda}$ for all k, then Condition 2.2 is obviously satisfied. More discussions on how to verify Condition 2.2 can be found in [17]. The advantages of Condition 2.2 for diffusion adaptive filters will also be discussed in Remark 2.8 below.

Theorem 2.1. If Conditions 2.1 and 2.2 are satisfied, then the coefficient matrices of the homogeneous part of (6) $\Big\{(I_{MN} - D_k\Psi_k\Psi_k')\cdot(A\otimes I_M), k \ge 1\Big\}$ is L_2-exponentially stable.

Remark 2.4. Different from almost all the existing works on distributed adaptive filtering, for example [1][3][8], we do not require the signals to satisfy the independency or stationarity assumptions. So, our result can be applied to a more general class of stochastic models, and, in fact, the results in [3] can be deduced from our results.

Remark 2.5. The above theorem essentially provides the joint excitation condition for the exponentially stability of the homogeneous part of (6), by which we can know how the bound of the mean square estimation error $E\|\widetilde{\Theta}_k\|^2$ depends on the amplitude of V_k and ω_k.

Remark 2.6. From Theorem 2.1, we can make a short comment on the relationship between our results and the results of other consensus-based distributed estimation algorithms. The similarity lies in that the excitation conditions both have the "sum-form" in space[9][10][11]. And the difference is in the excitation mechanism: our results rely on the "shrink" property of the network adjacency matrix A. Specifically, under the sum-form condition of the estimators and the ergodicity of A, the norm of the coefficient matrix will be strictly less than 1, then the homogeneous part of the error equation is exponentially stable, thereby the algorithm works. While other consensus-based results resort to the "consensus" property of A. In practical terms, as time evolves, the information collected by each individual estimator will asymptotically approximate the sum-form information of the network so that each individual estimator is like a "central unit", then the algorithm works under the sum-form condition.

Theorem 2.2. Suppose that Conditions 2.1 and 2.2 are satisfied, $\|\widetilde{\Theta}_0\|_{L_2} < \infty$, and that for some $\beta > 1$,

$$\sigma_2 \triangleq \sup_k \|\xi_k \log^\beta(e+\xi_k)\|_{L_2} < \infty, \quad (9)$$

where $\xi_k = \|V_k\| + \|W_{k+1}\|$. Then $\{\widetilde{\Theta}_k, k \ge 1\}$ is L_2 stable, and

$$\limsup_{k\to\infty} \|\widetilde{\Theta}_k\|_{L_2} \le c[\sigma_2 \log(e + \sigma_2^{-1})] \quad (10)$$

where c is a constant.

Remark 2.7. The proof of Theorem 2.2 is similar to Theorem 4.2 in [17], and we omit it due to space limitation. Moreover, both Theorems 2.1 and 2.2 can be extended to more general norms as those in [17]. Furthermore, more accurate estimation for the MSE can also be obtained by following the arguments of those in [19].

Remark 2.8. For a single sensor whose signals are generated by the following regression model:

$$y_k = \varphi_k'\theta_k + v_k, \quad (11)$$

where $y_k, \varphi_k, \theta_k, v_k$ are observations, regression vectors, unknown parameters and noises, respectively, if we apply the standard LMS to estimate θ_k, then the corresponding estimation error is L_2 stable, provided the following condition [17]holds:

Condition 2.3. (Excitation Condition).

$$\lambda_{min}\left\{E\left[\frac{1}{1+h}\sum_{j=kh+1}^{(k+1)h}\frac{\varphi_j(\varphi_j)'}{1+\|\varphi_j\|^2} \mid \mathcal{F}_{kh}\right]\right\} \in S^o \quad (12)$$

$$(13)$$

It is easy to find that for a network of sensors, even though none of the sensors satisfies the above Excitation Condition 2.3, the sensors as a whole can still have the possibility to satisfy Condition 2.2, which means that in a certain degree, the DLMS algorithm can make the sensors to fulfill the estimation task in a cooperative fashion, which any single sensor cannot do alone. We will provide a simulation example to explain this point in Section 4.

3. PROOF OF THEOREM 2.1

To prove Theorem 2.1, we need to provide some key lemmas first, with the proof being presented partially.

Lemma 3.1. ([21]). Assume that all the eigenvalues of $A \in \mathbb{R}^{M\times M}$ are $\lambda_1, \cdots, \lambda_M$ with respect to the eigenvectors $a_1, \cdots, a_M$, and all the eigenvalues of $B \in \mathbb{R}^{N\times N}$ are $\mu_1, \cdots, \mu_N$ with respect to the eigenvectors $b_1, \cdots, b_N$, then all the eigenvalues of $A\otimes B$ are $\lambda_i\mu_j, i = 1, \cdots M, j = 1, \cdots N$ with respect to $a_i \otimes b_j, i = 1, \cdots M, j = 1, \cdots N$.

Lemma 3.2. Assume that a family of matrices $\{A_k = (A_k^{ij}) \in \mathbb{R}^{N\times N}\}$ are semi-positive definite, and for any k, $A_k \le I_{N\times N}$, then there exists a constant b, such that for any k, i, j, $A_k^{ij} \le b$.

Lemma 3.3. ([17]). Assume that $\{\alpha_k\} \in S^o$ and $\alpha_k \le \alpha^*$, where α^* is a constant. Then for any $\varepsilon \in (0,1)$, we have

$$\{\varepsilon\alpha_k\} \in S^o. \quad (14)$$

Lemma 3.4. ([17]). Assume that $\alpha = \{\alpha_k, \mathcal{F}_k\}$, $\beta = \{\beta_k, \mathcal{F}_k\}$ are two adapted processes, satisfying

$$\beta_k \in [0,1], E[\beta_{k+1}|\mathcal{F}_k] \ge \alpha_k, k \ge 0. \quad (15)$$

Then $\{\beta_k\} \in S^o$ provided $\{\alpha_k\} \in S^o$.

Lemma 3.5. Assume that $A = (a_{ij})$ is an $N \times N$-dimensional doubly stochastic matrix which satisfies Condition 2.1, and $\{\Phi_k = (\Phi_k^{ij}), k = 1\cdots N\}$ are a sequence of $M \times M$-dimensional symmetric matrices satisfying $0 \le \Phi_k \le I_M, k = 1\cdots N$, then there exists a constant $C < 1$ depending only on A such that

$$\lambda_{max}\big[(A\otimes I)\cdot diag(I_M-\Phi_1, \cdots, I_M-\Phi_N)\cdot(A\otimes I)\big] \le 1 - C\delta, \quad (16)$$

where $\delta = \frac{1}{N}\lambda_{min}(\Phi_1 + \cdots + \Phi_N)$.

Proof From Condition 2.1 and Remark 2.2, A has N real eigenvalues in an ascending order

$$-1 < \lambda_N(A) \le \lambda_{N-1}(A) \le \cdots \le \lambda_2(A) < \lambda_1(A) = 1. \quad (17)$$

Denote $\lambda_{gap}(A) \triangleq \max\{|\lambda_2(A)|, |\lambda_N(A)|\}$.

Let $\{\alpha_1, \cdots \alpha_N\}$ be a system of orthogonal basis of R^N composed of the unit eigenvectors corresponding to $\{\lambda_1, \cdots, \lambda_N\}$, then $\alpha_1 = (\frac{1}{\sqrt{N}}, \cdots \frac{1}{\sqrt{N}})'$. Hence, by Lemma 3.1, the eigenvalues of $A \otimes I_M$ are $\{\lambda_i, i = 1, \cdots N\}$, and the eigenvectors corresponding to λ_i are $\{\alpha_i \otimes e_j, j = 1, \cdots M\}$, where e_i is the ith row of I_M. For convenience, we use $\beta_k (1 \le k \le MN)$ to denote $\alpha_{\lceil k/M \rceil} \otimes e_{k - M\lfloor k/M \rfloor}$, then $\{\beta_k, k = 1, \cdots, MN\}$ is a system of orthogonal eigenvectors of $A \otimes I$.

For any two MN-dimensional vectors x and y, we use $x' * y$ and $x' \diamond y$ to denote

$$x'\big[(A \otimes I) \cdot diag(I_M - \Phi_1, \cdots, I_M - \Phi_N) \cdot (A \otimes I)\big] y \quad (18)$$

and

$$x' diag(\Phi_1, \cdots, \Phi_N) \cdot y \quad (19)$$

respectively.

Pick any MN-dimensional unit vector x, then x can be written as $x = \sum_{i=1}^{MN} x_i \beta_i$ and $\sum_{i=1}^{MN} x_i^2 = 1$. Then

$$\begin{aligned} x' * x &= \sum_{i=1}^{MN} x_i^2 (\lambda_{\lceil i/M \rceil})^2 \\ &- (\sum_{i=1}^{MN} x_i \lambda_{\lceil i/M \rceil} \beta_i)' \diamond (\sum_{i=1}^{MN} x_i \lambda_{\lceil i/M \rceil} \beta_i) \\ &\triangleq S_1 - S_2, \end{aligned} \quad (20)$$

where $S_1 > 0, S_2 > 0$.

Now we consider S_1 and S_2 respectively. Since for $1 \le i \le M$, $\lambda_{\lceil i/M \rceil} = 1$, and for $M + 1 \le i \le MN$, $\lambda_{\lceil i/M \rceil} \le \lambda_{gap}$, then

$$S_1 \le \sum_{i=1}^{M} x_i^2 + \lambda_{gap}^2 (1 - \sum_{i=1}^{M} x_i^2) \le 1. \quad (21)$$

Let $S_2^1 \triangleq \sum_{i=1}^{M} x_i \lambda_{\lceil i/M \rceil} \beta_i$ and $S_2^2 \triangleq \sum_{i=M+1}^{MN} x_i \lambda_{\lceil i/M \rceil} \beta_i$, then

$$\begin{aligned} S_2 &= (S_2^1 + S_2^2)' \diamond (S_2^1 + S_2^2) \\ &= (S_2^1)' \diamond S_2^1 + (S_2^1)' \diamond S_2^2 + (S_2^2)' \diamond S_2^1 + (S_2^2)' \diamond S_2^2. \end{aligned} \quad (22)$$

For the first term of the right hand side of (22),

$$\begin{aligned} |(S_2^1)' \diamond S_2^1| &= \frac{1}{N} \sum_{i=1}^{N} (x_1, \cdots, x_M) \Phi_i (x_1, \cdots, x_M)' \\ &= \frac{1}{N} (x_1, \cdots, x_M) (\sum_{i=1}^{N} \Phi_i) (x_1, \cdots, x_M)' \\ &\ge \delta (\sum_{i=1}^{M} x_i^2). \end{aligned} \quad (23)$$

Notice that $(S_2^2)' \diamond S_2^2$ is the linear combination of the terms $x_i x_j (M + 1 \le i, j \le MN)$ and the coefficients depend on $\{\Phi_k^{ij}, k = 1, \cdots, N\}$ and A. By the condition that $0 < \Phi_k \le I_M$, for any i, j, k, we have $\{\Phi_k^{ij}\}$ have a common upper bound from Lemma 3.2. Then there exists constants $A_{ij} (M + 1 \le i, j \le MN)$ and $C_1 > 1$ only depending on A and M, N, such that

$$|(S_2^2)' \diamond S_2^2| \le \sum_{i=M+1}^{MN} \sum_{j=M+1}^{MN} A_{ij} |x_i x_j| \le C_1 \sum_{i=M+1}^{MN} x_i^2. \quad (24)$$

Similarly, there exist constants C_2, C_3 depending on A, such that

$$|(S_2^1)' \diamond S_2^2| \le C_2 \sum_{j=M+1}^{MN} x_j^2, \quad (25)$$

and

$$|(S_2^2)' \diamond S_2^1| \le C_3 \sum_{j=M+1}^{MN} x_j^2 \quad (26)$$

Combining (23), (24), (25) and (26) together, we obtain

$$S_2 \ge \delta (\sum_{i=1}^{M} x_i^2) - (C_1 + C_2 + C_3)(1 - \sum_{i=1}^{M} x_i^2). \quad (27)$$

Consider the function $f(y) = \delta y - (C_1 + C_2 + C_3)(1 - y)$, which monotonically increases on $[0, 1]$. It can be computed that $f(\frac{\frac{\delta}{2} + C_1 + C_2 + C_3}{\delta + (C_1 + C_2 + C_3)}) = \frac{\delta}{2}$ and $f(1) = \delta$, then let $K = \frac{\frac{\delta}{2} + C_1 + C_2 + C_3}{\delta + (C_1 + C_2 + C_3)}$, we have $f(y) \ge \frac{\delta}{2}$ on $[K, 1]$.

Now we discuss $S_1 - S_2$ in two cases:
(i) If $K < \sum_{i=1}^{M} x_i^2 \le 1$, from the analysis above, we have

$$S_1 - S_2 < 1 - \frac{\delta}{2}. \quad (28)$$

(ii) If $\sum_{i=1}^{M} x_i^2 \le K$, from $S_2 > 0$ we have

$$S_1 - S_2 \le 1 - \frac{(1 - \lambda_{gap}^2)}{2(C_1 + C_2 + C_3)} \delta. \quad (29)$$

Combining (28), (29) and (20) together, and from the arbitrariness of x, we complete the proof with $C = \max\{\frac{1}{2}, \frac{(1 - \lambda_{gap}^2)}{2(C_1 + C_2 + C_3)}\} < 1$. □

Lemma 3.6. Under Conditions 2.1 and 2.2, there exists a constant $C^* < 1$ such that:

$$\lambda_{max} \Big\{ E[\Phi'(k + 1, k) \Phi(k + 1, k) \mid \mathcal{F}_{k-1}] \Big\} \le 1 - C^* \lambda_{k-1}, \quad (30)$$

where $\Phi(\bullet, \bullet)$ is defined as:

$$\Phi(n + 1, m) = (I_{MN} - D_n \Psi_n \Psi_n') \cdot (A \otimes I_M) \Phi(n, m), \quad (31)$$

$$\Phi(m, m) = I_{MN}, \forall n \ge m. \quad (32)$$

Lemma 3.7. Under the same conditions and notations of Lemma 3.6, for any $k_0 \ge 0$, consider the equation

$$x_k = \Phi(k + 1, k) x_{k-1}, \qquad k \ge k_0 + 1, \quad (33)$$

where x_{k_0} is deterministic and $\|x_{k_0}\| = 1$. Then there exists $\alpha_k \in [0, 1]$ such that $\alpha_k \in \mathcal{F}_k$ and

$$\|x_k\| \le (1 - \alpha_k) \|x_{k-1}\|, k \ge k_0 + 1, \quad (34)$$

and

$$E[\alpha_{k+1} | \mathcal{F}_k] \ge \frac{C^*}{2} \lambda_k, k \ge k_0 + 1, \quad (35)$$

where C^* is defined in Lemma 3.6.

Proof of Theorem 2.1:

Since $\frac{C^*}{2}\lambda_k \in [0, \frac{1}{2}]$ and $\{\lambda_k\} \in S^o$, then by Lemma 3.3, we have

$$\{\frac{C^*}{2}\lambda_k\} \in S^o. \tag{36}$$

Using Lemma 3.4, we obtain

$$\{\alpha_k\} \in S^o, \tag{37}$$

where $\{\alpha_k\}$ is defined in Lemma 3.7.

Consider the equation (33), and by (34), we have

$$\|x_k\| \le (1-\alpha_k)\cdots(1-\alpha_{k_0+1})\|x_{k_0}\|, \tag{38}$$

which is followed by

$$\|x_k\|^2 \le (1-\alpha_k)\cdots(1-\alpha_{k_0+1}). \tag{39}$$

Combing this and (37) together, we complete the proof of Theorem 2.1. □

4. SIMULATION RESULTS

In order to illustrate Theorem 2.1 and 2.2, we present some simulation examples in Figure 1-Figure 4. Assume that a network consists of $N = 3$ sensors with

$$A = \begin{pmatrix} 2/3 & 1/3 & 0 \\ 1/3 & 1/2 & 1/6 \\ 0 & 1/6 & 5/6 \end{pmatrix} \tag{40}$$

and we will use the network to estimate the unknown M-dimensional parameters $\{\theta_k\}$ with $M = 3$ cooperatively. The observation noises $\{v_k^i, i = 1, \cdots, N, k \ge 1\}$ are temporally and spatially independently distributed with $v_k^i \sim N(0, 1)$. For $i = 1, 2, 3$, assume that the regression vector $\{\varphi_k^i\}$ are the outputs of the linear stochastic model

$$x_k^i = A^i x_{k-1}^i + B^i \xi_k^i, \quad \forall k \ge 1, \tag{41}$$

$$\varphi_k^i = C^i x_k^i, \tag{42}$$

where $\{\xi_k^i, i = 1, \cdots, N, k \ge 1\}$ are temporally and spatially independently distributed with $\xi_k^i \sim N(0, 1)$. Set

$$x_0^1 = x_0^2 = x_0^3 = (0, 0, 0)',$$

$$A^1 = A^2 = A^3 = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/5 \end{pmatrix},$$

$$B^1 = (1, 0, 0)', B^2 = (0, 1, 0)', B^3 = (0, 0, 1)',$$

$$C^1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, C^1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, C^1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It can be verified that $\{\varphi_k^i, k \ge 0\}$ is neither independent nor stationary. Through the proof similar to that for Example 2 in [20], we can prove that Conditions 2.1 and 2.2 are satisfied for the network of sensors under the weighted matrix A , while Excitation Condition 2.3 is not for each individual sensor. Set $\widetilde{\theta}_0^1 = 2.9580, \widetilde{\theta}_0^2 = 2.2361, \widetilde{\theta}_0^3 = 1.7321$. We plot the mean square error(MSE)(averaged over 50 Runs) in four contexts:

(i) When $\omega_k = 0$, and there is no information exchange between sensors. Each sensor use the standard LMS algorithm to estimate the parameter separately. From Figure 1, we see that the MSE at all sensors is very large.

(ii) When $\omega_k = 0$, and the network of sensors cooperate to estimate the parameter with the weighted matrix A. The sensors apply the DLMS algorithm, then the conditions in Theorem 2.2 holds. Form Figure 2, we see that the MSE at all sensors decreases to small values.

(iii) When ω_k is uniformly and independently distributed in the interval $[-0.15, 0.15]$, and there is no information exchange between sensors. Each sensor use the standard LM-S algorithm to estimate the parameter separately. From Figure 3, we see that the MSE at all sensors is very large.

(iv) When ω_k is uniformly and independently distributed in $[-0.15, 0.15]$, and the network of sensors cooperate to estimate the parameter with the weighted matrix A using the DLMS algorithm, then the conditions in Theorem 2.2 holds. From Figure 4, we see that the MSE at all sensors is within a small neighborhood of 0.

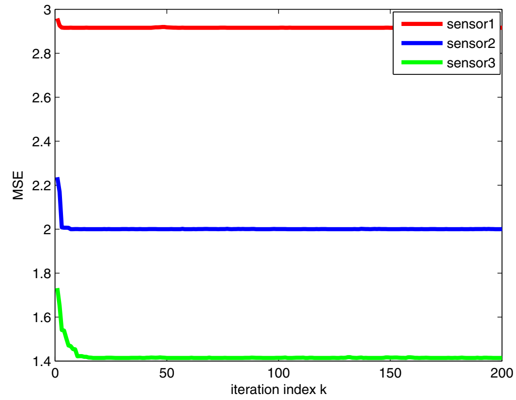

Fig. 1. The parameter is time-invariant and each sensor estimates the parameters by the standard LMS separately without information exchange.

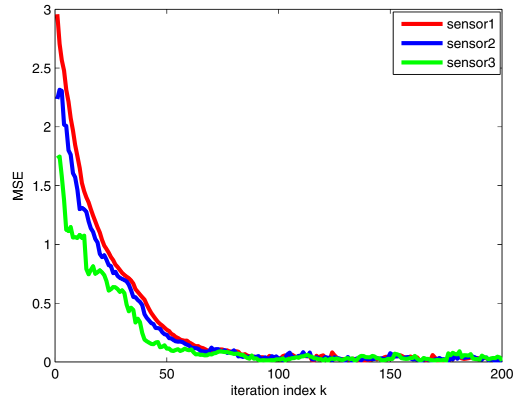

Fig. 2. The parameter is time-invariant and the sensors estimate the parameters cooperatively using the DLMS.

5. CONCLUDING REMARKDS

In this paper, we have established the L_2 stability of the MSE for DLMS algorithm under correlated and non-stationary conditions on the system signals. We have shown that the network of sensors can cooperate to fulfill the estimation or filtering task even though any single sensor cannot. This paper can be regarded as the first step towards the theoretical analysis of distributed adaptive filtering for more general and more practical stochastic signal models. Of course, many problems still remain to be further investigated, for example, how to further relax

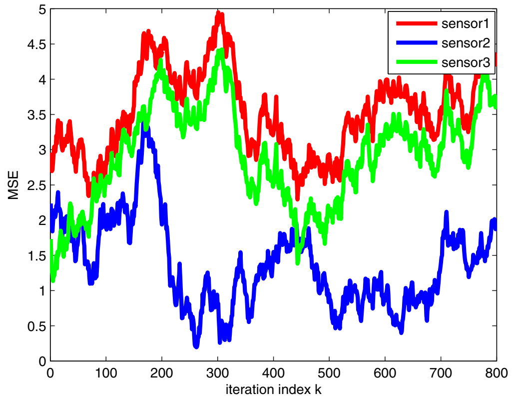

Fig. 3. The parameters are time-variant and each sensor estimates the parameters by the standard LMS separately without information exchange.

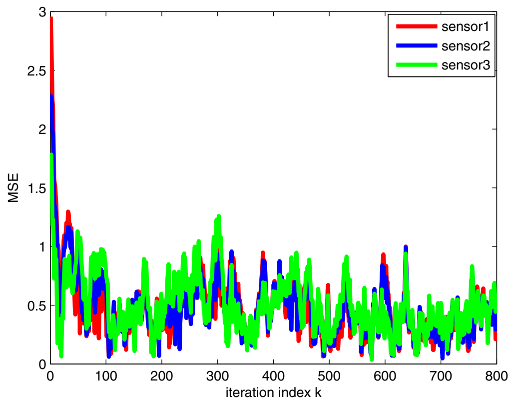

Fig. 4. The parameters are time-variant and the sensors estimate the parameters cooperatively using the DLMS.

the joint excitation condition 2.2? how to analyze other distributed filtering algorithms including the Kalman filtering-based cooperative algorithm? how to combine distributed filtering with distributed control properly?

REFERENCES

[1] A. H. Sayed and C. G. Lopes, Adaptive processing over distributed networks, IEICE *Trans. Fund. Electron., Commun. Comput. Sci.*, vol. E90-A, no. 8, pages 1504-1510, Aug. 2007.

[2] C. G. Lopes and A. H. Sayed, Incremental adaptive strategies over distributed networks, *IEEE Trans. Signal Process.*, vol. 55, no. 8, pages 4064-4077, Aug. 2007.

[3] F. S. Cattivelli, and A.H. Sayed, "Diffusion LMS Strategies for Distributed Estimation", *IEEE Trans. on Signal Processing*, 55, pages 2069-2084, 2010.

[4] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, Diffusion recursive least-squares for distributed estimation over adaptive networks, *IEEE Trans. Signal Process.*, vol. 56, no. 5, pages 1865-1877, May 2008.

[5] F. S. Cattivelli and A. H. Sayed, Diffusion mechanisms for fixedpoint distributed Kalman smoothing, in *Proc. EUSIPCO*, Lausanne, Switzerland, Aug. 2008.

[6] S. S. Ram, A. Nedic, and V. V. Veeravalli, Stochastic incremental gradient descent for estimation in sensor networks, in *Proc. Asilomar Conf. Signals, Syst., Comput., Pacific Grove, CA*, pages 582-586, 2007.

[7] L. Xiao, S. Boyd, and S. Lall, A space-time diffusion scheme for peer-to-peer least-squares estimation, in *Proc. IPSN*, Nashville, TN, Apr. 2006, pages 168-176.

[8] I. D. Schizas, G. Mateos, and G. B. Giannakis, "Distributed LMS for consensus-based in-network adaptive processing," *IEEE Trans. Signal Process.*, vol. 8, no. 6, pages 2365-2381, June, 2009.

[9] Q. Zhang and J.F. Zhang, "Distributed parameter estimation over unreliable networks with Markovian switching topologies", *IEEE Trans. on Automatic Control*, Vol. 57, No. 10, pages 2545-2560, 2012.

[10] S. S. Stankovic, M. S. Stankovic, D. M. Stipanovic, Decentralized parameter estimation by consensus based stochastic approximation. *IEEE Trans. on Automatic Control*, Vol. 56, No. 3, pages 531-543, 2011.

[11] S. Kar, J. M. Moura , H. V. Poor "Distributed linear parameter estimation: Asymptotically efficient adaptive strategies". *SIAM J. on Control and Optimization*, Vol.51, No.3, pages 2200-2229, 2013.

[12] F. S. Cattivelli, and A.H. Sayed , "Diffusion LMS Strategies for Distributed Kalman Filtering and Smoothing", *IEEE Trans. on Automatic Control*, 58, pages 1035-1048, 2010.

[13] R. Olfati-Saber, "Distributed Kalman Filtering for Sensor Networks", in *Proceedings of the 46th Conference on Decision Control*, New Orleans, LA, pages 5492-5498, 2007.

[14] I. D. Schizas, G.B.Giannakis, A. Ribeiro, and S. I. Roumeliotis, "Consensus in Ad Hoc WSNs with Noisy LinkC Part II: Distributed Estimation and Smoothing of Random Signals", *IEEE Trans. on Signal Processing*, Vol. 56, pages 1650-1666, 2008.

[15] A. Rastegarnia, M.A. Tinati, and A. Khalili, "Performance Analysis of Quantized Incremental LMS Algorithm for Distributed Adaptive Estimation", in *Signal Processing*, Vol. 90, pages 2621-2627, 2012.

[16] S. Xie, H. Li, Distributed LMS estimation over networks with quantised communications, International Journal of Control, Vol.86, No.3, pages 478-492,2013.

[17] L. Guo. "Stability of recursive stochastic tracking algorithms", *SIAM J.on Control and Optimization*, Vol. 32, No. 5, pages 1195-1225,1994. .

[18] L. Guo and L. Ljung, "Exponential stability of general tracking algorithms", *IEEE Trans. on Automatic Control*, Vol.40, No.8, pages 1376-1387, 1995.

[19] L. Guo, L.Ljung, "Performance analysis of general tracking algorithms", *IEEE Trans.on Automatic Control*, Vol.40, No.8, pages 1388-1402, 1995.

[20] J. F. Zhang, L. Guo, and H. F. Chen. "L_p-stability of estimation errors of kalman filter for tracking time-varying parameters", *International Journal of Adaptive Control and Signal Processing* Vol.5, No.3, pages 155-174, 1991.

[21] M. Xue and S. Roy,"Kronecker products of defective matrices: Some spectral properties and their implications on observability," in 2012 *American Control Conference*, June 2012.