Communications in Information and Systems
Volume 15, Number 4, 423–451, 2015

Distributed estimation for parameter in heterogeneous linear time-varying models with observations at network sensors

Jinlong Lei and Han-Fu Chen

In this paper, a distributed stochastic approximation based estimation algorithm is proposed to estimate the parameter in heterogeneous linear time-varying models associated with sensors from a network. At any time, each agent updates its estimate using the local observations and the information derived from its neighboring agents. The estimates are shown to converge to the one that minimizes the long run average of the square residuals. Switch of the communication graphs is assumed to be deterministic, and the regressors of the linear models are assumed to satisfy some ergodic property, rather than the conditional independence or strict stationarity. Numerical simulations are given to illustrate the obtained theoretic result.

1. Introduction

With wide applications of sensor networks [1, 2], estimating unknown parameters based on the data gathered by a group of spatially distributed sensors has attracted much attention from researchers. In the centralized approach, all data is transferred to a fusion center and the collected data is processed there. However, this approach may not be preferable, because 1) it costs too much communication resource, 2) it is not robust since a failure of the central node would lose all information achieved, and 3) some agents might be reluctant to share its local data due to privacy concern [3]. Alternatively, for the distributed approach each sensor acts as an individual adaptive filter, which estimates the parameter by using its local observations and the information derived from its neighboring sensors. So, compared with

Dedicated to Professor P. E. Caines on the occasion of his 70th birthday. This work was supported by the NSFC under Grants 61273193, 61120106011, 61134013, the 973 program of China under grant No. 2014CB845301, and the National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences.

the centralized approach, the distributed estimation schema has advantages of enhancing the robustness of the sensor networks, preserving privacy, and reducing the communication and computation costs.

Consensus problems have been widely investigated recently in different aspects [4–6]. There exist many distributed problems that are solved by consensus-based distributed algorithms, for example, sensor localization [7], distributed optimization [8, 9], distributed stochastic approximation [10, 11], and distributed control [12][13]. As for the parameter estimation problem in sensor networks, consensus-based distributed estimation algorithms and their convergence analysis have also been studied in many papers, such as the diffusion least mean square (LMS) algorithm [14, 15], the diffusion recursive least squared (RLS)[16], the distributed LMS algorithm [17], the distributed RLS [18], the distributed Kalman filtering [19][20], SA based distributed estimation algorithm [21–23], and so on.

Most of the above-mentioned works require the (conditionally) independent or strictly stationary ergodic conditions on the observation models. Algorithms proposed in [14],[15] utilize constant step-sizes and give the mean-square errors when the regressors are assumed to be independent Gaussian sequences. In [17], the estimation error norms are shown to be bounded for most of the time when the regressors are assumed to be strictly stationary ergodic. In [21], the regressors are assumed to be independent, the covariance matrix of the regressor is assumed to satisfy the strict diagonal dominance condition, and the fourth order moments of the regressor are assumed to be finite. In [22], the regressors of all sensors are assumed to be iid (independent and identically distributed) sequences, while in [23], the regressors at time k are assumed to be independent of the σ-filed $\mathcal{F}_{t-1}$ generated by the past information. In [22, 23], expectations of regressors are assumed to be known and are used in the distributed estimation algorithm.

In this paper, we estimate the unknown M-dimensional vector θ^* based on the data gathered by N spatially distributed sensors in the network. Every agent $i = 1, \ldots, N$ at time k has access to its local d_i-dimensional vector measurement given by the following linear time-varying model

$$Y_{i,k} = H_{i,k}\theta^* + v_{i,k}, \tag{1}$$

where $H_{i,k} \in \mathbb{R}^{d_i \times M}$ is the regressor accessible to agent i, and $v_{i,k} \in \mathbb{R}^{d_i \times 1}$ is the local observation noise of agent i. Set

$$Y_k \triangleq col\{Y_{1,k}, \ldots, Y_{N,k}\} \in \mathbb{R}^{d \times 1}, \tag{2}$$

$$H_k \triangleq col\{H_{1,k}, \ldots, H_{N,k}\} \in \mathbb{R}^{d \times M} \tag{3}$$

with $d = \sum_{i=1}^{N} d_i$, where $col\{x_1, \ldots, x_N\} \triangleq (x_1^T, \ldots, x_N^T)^T$. The distributed parameter estimation problem is to seek the M-dimensional vector that minimizes the long run average of the square residuals

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \| Y_k - H_k\theta \|^2 . \tag{4}$$

As to be shown this is equivalent to a distributed root-seeking problem with some proper assumptions imposed on the models. Then the distributed estimation algorithm based on the distributed stochastic approximation algorithm with expanding truncations (DSAAWET) given in [11] can be applied. The update rule of each sensor is a combination of the consensus term being the weighted average of the estimates derived at its neighboring agents, and the innovation term processing its current observation. The estimates are shown to converge to the minimum of (4). Compared with the existing results, here we impose neither (conditional) independency nor strict stationarity on regressors, but the switch of communication graphs is assumed to be deterministic rather than random. Numerical simulations are given to demonstrate the theoretic result for the case, where the regressors of all linear time-varying models are generated by AR processes.

The remainder of the paper is organized as follows. The distributed estimation algorithm is proposed and its convergence theorem is formulated in Section 2. The global behavior of the estimate sequence is given in Section 3, while the local properties of the estimates and the noises along bounded subsequences of estimates are presented in Section 4. The proof of the theorem is placed in Section 5. The numerical simulations are demonstrated in Section 6, and some concluding remarks are given in Section 7.

The notations used in the paper are as follows: A given matrix $A = [a_{ij}] \in \mathbb{R}^{n\times n}$ is called nonnegative if $a_{ij} \geq 0 \quad \forall i, j = 1, \ldots, n.$ A nonnegative square matrix A is called doubly stochastic if $A\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^T A = \mathbf{1}^T$, where $\mathbf{1}$ denotes the vectors of compatible dimensions with all entries equal to 1, and X^T denotes the transpose of X. I_m denotes the identity matrix of dimension m. $\mathbf{0}$ denotes the matrices or vectors of compatible dimensions with all entries equal to 0. By $\otimes$ we denote the Kronecker product.

A nonnegative matrix $W(k) = [\omega_{ij}(k)]_{i,j=1}^{N}$ with positive diagonal entries is used to describe the communication relationship among N agents in the network at time k. The corresponding digraph is denoted by $\mathcal{G}_k = (\mathcal{V}, \mathcal{E}_k)$, where $\mathcal{V} = \{1, \ldots, N\}$ is the node set and $\mathcal{E}_k = \{(j, i) : \omega_{ij}(k) > 0\}$ is the edge set. Denote by $\mathcal{N}_i(k) = \{j \in \mathcal{V} : \omega_{ij}(k) > 0\}$ the neighboring agents of agent i at time k.

The digraph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ is called strongly connected if for any given pair $i, j \in \mathcal{V}$, there exists a sequence of distinct nodes $i_1, \ldots, i_p$ such that $(i, i_1) \in \mathcal{E}$, $(i_1, i_2) \in \mathcal{E}, \ldots, (i_p, j) \in \mathcal{E}$.

2. Distributed estimation algorithm

We first show that the distributed parameter estimation (1)–(4) is equivalent to the distributed root-seeking problem for linear local functions. Then a distributed estimation algorithm based on DSAAWET [11] is proposed to recursively estimate the unknown parameter.

2.1. Assumptions

We impose the following assumptions on the linear time-varying models.

C1 For any $i \in \mathcal{V}$, the regressor $H_{i,k}$ satisfies the ergodicity property:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} H_{i,k}^T H_{i,k} \triangleq R_{i,h}, \tag{5}$$

and $\{v_{i,k}, \mathcal{F}_{i,k}\}$ is a martingale difference sequence (mds) with

$$\sup_k E[\|v_{i,k+1}\|^2 | \mathcal{F}_{i,k}] < \infty,$$

where $\mathcal{F}_{i,k} \triangleq \sigma\{H_{i,t}, v_{i,t}, 1 \le t \le k\}$.

C2 $R_h \triangleq \sum_{i=1}^{N} R_{i,h}$ is positive definite.

Remark 1. Clearly, the independent or strictly stationary ergodic regressors imply (5) with probability one. There are some other regressors also satisfying C1. The collective identifiability assumption C2 extends identifiability condition for a centralized estimator that is needed to get a consistent estimate of the unknown parameter θ^*. This together with the ergodic assumption plays a crucial rule in the convergence analysis.

We impose the following assumption on the communication network.

C3 (a) $W(k)\ \ \forall k \ge 0$ are doubly stochastic matrices;

(b) There exists a constant $0 < \eta < 1$ such that

$$\omega_{ij}(k) \ge \eta \quad \forall j \in \mathcal{N}_i(k) \ \ \forall i \in \mathcal{V} \ \ \forall k \ge 0;$$

(c) The digraph $\mathcal{G}_\infty = \{\mathcal{V}, \mathcal{E}_\infty\}$ is strongly connected, where

$$\mathcal{E}_\infty = \{(j,i) : (j,i) \in \mathcal{E}(k) \text{ for infinitely many indices } k\};$$

(d) There exists a positive integer B such that

$$(j,i) \in \mathcal{E}(k) \cup \mathcal{E}(k+1) \cup \cdots \cup \mathcal{E}(k+B-1)$$

for all $(j,i) \in \mathcal{E}_\infty$ and any $k \geq 0$.

Remark 2. If C3 holds, then by [8, Proposition 1] there exist constants $c > 0, 0 < \rho < 1$ such that

$$\left\| \Phi(k,s) - \frac{1}{N}\mathbf{1}\mathbf{1}^T \right\| \leq c\rho^{k-s+1} \quad \forall k \geq s, \tag{6}$$

where $\Phi(k,s)$ is given by

$$\Phi(k,k+1) = \mathrm{I}_N, \ \Phi(k,s) = W(k)\cdots W(s) \ \forall k \geq s.$$

2.2. Equivalent problem

Since $H_{i,k}$ is adapted to $\mathcal{F}_{i,k}$, by the property of the weighted mds [26]

$$\sum_{k=1}^{n} H_{i,k}^T v_{i,k} = O(s_n(2) log^{\frac{1}{2}+\eta}(s_n^2(2)+e)) \quad a.s. \ \forall \eta > 0,$$

where $s_n(2) = (\sum_{k=1}^n \|H_{i,k}\|^2)^{\frac{1}{2}}$. Hence by (5) we obtain

$$\sum_{k=1}^{n} H_{i,k}^T v_{i,k} = O(n^{\frac{1}{2}+\eta}) \quad a.s. \ \forall \eta > 0.$$

Thus, from (5) we obtain

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} H_{i,k}^T Y_{i,k} &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} H_{i,k}^T H_{i,k}\theta^* + \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} H_{i,k}^T v_{i,k} \\ &= R_{i,h}\theta^* \triangleq R_{i,hy} \quad a.s. \end{aligned} \tag{7}$$

From (2) (3) it follows that

$$\frac{1}{n}\sum_{k=1}^{n} H_k^T H_k = \frac{1}{n}\sum_{i=1}^{N}\sum_{k=1}^{n} H_{i,k}^T H_{i,k}, \quad \frac{1}{n}\sum_{k=1}^{n} H_k^T Y_k = \frac{1}{n}\sum_{i=1}^{N}\sum_{k=1}^{n} H_{i,k}^T Y_{i,k},$$

and hence by (5) (7) we derive

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}H_k^TH_k=\sum_{i=1}^{N}R_{i,h}\triangleq R_h,$$
$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}H_k^TY_k=\sum_{i=1}^{N}R_{i,hy}=R_h\theta^*\triangleq R_{hy}.$$

Then we obtain

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\parallel Y_k-H_k\theta\parallel^2=\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}Y_k^TY_k-2\theta^TR_{hy}+\theta^TR_h\theta.$$

So, the minimum of (4) is the root of its gradient function:

$$-R_{hy}+R_h\theta=0.$$

Since R_h is positive definite, the minimum of (4) is uniquely achieved at $\theta=R_h^{-1}R_{hy}=\theta^*$.

In summary, under C1, C2 the problem of finding the minimum of (4) is converted to collectively seeking root of the function

$$f(\theta)=\sum_{i=1}^{N}f_i(\theta)$$

with $f_i(\theta)=-R_{i,h}\theta+R_{i,hy}$. The corresponding root set is $J=\{\theta^*\}$.

2.3. Algorithm

Denote by $\theta_{i,k}$ the estimate for θ^* given by agent i at time k. Since $R_{i,h}$ and $R_{i,hy}$ cannot be directly derived, by replacing $R_{i,h}$ and $R_{i,hy}$ with $H_{i,k}^TH_{i,k}$ and $H_{i,k}^TY_{i,k}$ the observation of the function $f_i(\theta)=-R_{i,h}\theta+R_{i,hy}$ at point $\theta_{i,k}$ is constructed as

$$O_{i,k+1}=H_{i,k}^T(Y_{i,k}-H_{i,k}\theta_{i,k}).\tag{8}$$

Thus, the observation noise $\varepsilon_{i,k+1}$ is

$$\varepsilon_{i,k+1}=O_{i,k+1}-f_i(\theta_{i,k})=H_{i,k}^TY_{i,k}-R_{i,hy}+(R_{i,h}-H_{i,k}^TH_{i,k})\theta_{i,k}.\tag{9}$$

For any $i \in \mathcal{V}$, the estimate is generated by the following algorithm

$$\sigma_{i,0} = 0, \quad \hat{\sigma}_{i,k} \triangleq \max_{j \in \mathcal{N}_i(k)} \sigma_{j,k}, \tag{10}$$

$$\theta'_{i,k+1} = \left(\sum_{j \in \mathcal{N}_i(k)} \omega_{ij}(k) \theta_{j,k} I_{[\sigma_{j,k} = \hat{\sigma}_{i,k}]} + \frac{1}{k} O_{i,k+1} \right) I_{[\sigma_{i,k} = \hat{\sigma}_{i,k}]}, \tag{11}$$

$$\theta_{i,k+1} = \theta'_{i,k+1} I_{[\|\theta'_{i,k+1}\| \le \hat{\sigma}_{i,k}]}, \tag{12}$$

$$\sigma_{i,k+1} = \hat{\sigma}_{i,k} + I_{[\|\theta'_{i,k+1}\| > \hat{\sigma}_{i,k}]}, \tag{13}$$

where $O_{i,k+1}$ is defined in (8), and I_A is the indicator of a random event A, i.e., $I_A(\omega) = 1$ if $\omega \in A$, and $I_A(\omega) = 0$, otherwise.

2.4. Convergence theorem

Theorem 3. *Let $\{\theta_{i,k}\}_{k \ge 1}$ be generated by* (10)–(13) *with $O_{i,k+1}$ defined by* (8) *for any initial value $\theta_{i,0}$. Assume C1, C2, and C3 hold. Then*

$$\lim_{k \to \infty} \theta_{i,k} = \theta^*, \quad a.s., \quad \forall i \in \mathcal{V}.$$

The proof consists in verifying conditions guaranteeing convergence of DSAAWET stated in Appendix. It is noticed that the algorithm (10)–(13) coincides with the algorithm (68)–(71) by identifying $x_{i,k}$, x^*, γ_k and M_k in (68)–(71) with $\theta_{i,k}$, $\mathbf{0}$, $\frac{1}{k}$, and k in (10)–(13), respectively. The detailed proof of Theorem 3 is given in Section 5. The theorem points out that the estimates given by all agents converge to the true parameter θ^* with probability one.

Remark 4. Assume there exists a positive constant $\alpha > 0$ such that

$$\frac{1}{n} \sum_{k=1}^{n} (H_{i,k}^T H_{i,k} - R_{i,h}) = o\left(\frac{1}{n^\alpha}\right). \tag{14}$$

Then Theorem 3 remains true if the step-size $\frac{1}{k}$ in (11) is replaced by $\frac{1}{k^\beta}$ with $1 \ge \beta > 0.5$ and $\alpha + \beta \ge 1$.

Remark 5. Compared with the iid Gaussian assumption [14],[15] and the strictly stationary condition [17], we impose weaker conditions on the regressors. It will be shown in Section 6 that the regressors being AR processes satisfy (5). Besides, the switch of the communication graphs is assumed to be deterministic in the paper, while the random switches are considered in [22, 23], and the fixed undirected graph is considered in [14, 15, 17].

3. Global behavior of estimate sequence

Denote by $\tau_{i,m} \triangleq \inf\{k : \sigma_{i,k} = m\}$ the smallest time when the truncation number of agent i has reached m, by $\tau_m \triangleq \min_{i\in\mathcal{V}} \tau_{i,m}$ the smallest time when at least one of agents has its truncation number reached m, and by $\sigma_k \triangleq \max_{i\in\mathcal{V}} \sigma_{i,k}$ the largest truncation number among all agents at time k. Set $\tilde{\tau}_{j,m} \triangleq \max\{\tau_{j,m}, \tau_{m+1}\}$. We first recall some results from [11] that will be used in the sequel.

Lemma 6. i) *[11, Remark 3.1] For* $\{\theta_{i,k}\}$ *generated by* (10)–(13) *with any initial values the following assertion takes place:*

$$\theta_{i,k+1} = \mathbf{0} \text{ when } \sigma_{i,k+1} > \sigma_{i,k}. \tag{15}$$

ii) *[11, Lemma 4.3] Assume C3 holds. Then*

$$\tilde{\tau}_{j,m} \le \tau_m + B(N-1) \quad \forall j \in \mathcal{V} \quad \text{for } m \ge 0.$$

iii) *[11, Lemma 5.5] Assume C3 holds. If* $\lim_{k\to\infty} \sigma_k = \sigma < \infty$, *then there exists an integer* $k_0 > 0$ *such that*

$$\sigma_{i,k} = \sigma \quad \forall k \ge k_0 \quad \forall i \in \mathcal{V}.$$

Next we give some lemmas that will be used for the proof of Theorem 3.

Lemma 7. *For a sequence of matrices* $\{A_k\}$, *if*

$$\lim_{k\to\infty} \frac{\sum_{m=1}^{k} A_m}{k} = \bar{A}, \tag{16}$$

then for any constant $T > 0$

$$\lim_{k\to\infty} \sum_{m=k}^{m(k,T_k)} \frac{A_m - \bar{A}}{m} = 0 \quad \forall T_k \in [0, T], \tag{17}$$

$$\lim_{k\to\infty} \frac{1}{T} \sum_{m=k}^{m(k,T)} \frac{A_m}{m} = \bar{A}, \tag{18}$$

where $m(k,T) \triangleq \max\{m : \sum_{p=k}^{m} \frac{1}{p} \le T\}$.

Proof. By denoting $Q_k = \sum_{m=1}^{k}(A_m - \bar{A})$, from (16) it follows that $\frac{Q_k}{k} \xrightarrow[k\to\infty]{} 0$. Therefore,

$$\sum_{m=k}^{m(k,T_k)} \frac{A_m - \bar{A}}{m} = \sum_{m=k}^{m(k,T_k)} \frac{1}{m}(Q_m - Q_{m-1})$$
$$= \frac{Q_{m(k,T_k)}}{m(k,T_k)} - \frac{Q_{k-1}}{k} + \sum_{m=k}^{m(k,T_k)-1} \frac{1}{m(m+1)} Q_m \xrightarrow[k\to\infty]{} 0 \quad \forall T_k \in [0,T],$$

and hence (17) holds.

By replacing T_k in (17) with T, it is seen that (18) holds. □

Set $\Theta_k = col\{\theta_{1,k}, \ldots, \theta_{N,k}\}$, $\theta_k = \frac{1}{N}\sum_{i=1}^{N}\theta_{i,k}$, $\Theta_{\perp,k} = \Theta_k - (\mathbf{1} \otimes \mathrm{I}_M)\theta_k$, and $\varepsilon_k = col\{\varepsilon_{1,k}, \ldots, \varepsilon_{N,k}\}$.

Lemma 8. *If C1 and C3 hold, then the sequence of estimates* $\{\Theta_k\}$ *generated by* (10)–(13) *contains at least one bounded subsequence* $\{\Theta_{n_k}\}$ *with* $\sigma_{i,n_k} = \sigma_{n_k}\ \forall i \in \mathcal{V}$.

Proof. In what follows all discussion is carried out for a fixed sample path.

Case 1: If $\lim\limits_{k\to\infty} \sigma_k = \sigma < \infty$, then there exists a positive integer k_0 such that there is no truncation after k_0. Consequently, the estimate sequence $\{\Theta_k\}$ is bounded and by Lemma 6 iii) $\sigma_{i,k} = \sigma \quad \forall k \geq k_0 \quad \forall i \in \mathcal{V}$.

Case 2: For the case $\lim\limits_{k\to\infty} \sigma_k = \infty$ it suffices to show that for sufficiently large $m > m_0 \triangleq 2^D - 1$

$$\sigma_{i,\tau_m+D} = m \quad \forall i \in \mathcal{V}, \quad \| \Theta_{\tau_m+D} \| \leq c_b, \tag{19}$$

where $D = (N-1)B$, and $c_b = \sqrt{N}(2^D - 1)$.

Set $k = \tau_m$. For sufficiently large $m \geq m_0$ and any $q = 1, \ldots, D$ we first show the following facts:

i) For any agent i with $\sigma_{i,k} = m$ it holds that

$$\sigma_{i,k+q} = m, \text{ and } \| \theta_{i,k+q} \| \leq 2^q - 1 \leq M_m; \tag{20}$$

ii) For any agent j with $\sigma_{j,k} < m$ it holds that

$$\begin{aligned} &\sigma_{j,k+q} \leq m, \\ &\text{further, } \| \theta_{j,k+q} \| \leq 2^q - 1, \text{ if, in addition, } \sigma_{j,k+q} = m. \end{aligned} \tag{21}$$

We prove i) and ii) by induction.

Step 1: We first show that i) and ii) hold for $q = 1$.

Since $k = \tau_m$, by the definition of τ_m we derive $\sigma_{j,k} \leq m$ and $\sigma_{j,k-1} < m \ \forall j \in \mathcal{V}$. Then from (10) it follows that $\hat{\sigma}_{i,k} = m$ for any agent i with $\sigma_{i,k} = m$. Since $\sigma_{i,k-1} < \sigma_{i,k}$, $\theta_{i,k} = \mathbf{0}$ by (15). Then from (11) we derive

$$\theta'_{i,k+1} = \frac{1}{k} H_{i,k}^T Y_{i,k}. \tag{22}$$

By (7) we have

$$\lim_{k\to\infty} \frac{1}{k} H_{i,k}^T Y_{i,k} = 0. \tag{23}$$

Thus, there exists a sufficiently large $k \geq m_0$ such that

$$\frac{1}{k} \parallel H_{i,k}^T Y_{i,k} \parallel \leq 1 \quad \forall i \in \mathcal{V}. \tag{24}$$

Since $k = \tau_m \geq m$, from (22) (24), $M_m = m$, and $m_0 = 2^D - 1$ it follows that for sufficiently large $m \geq m_0$

$$\parallel \theta'_{i,k+1} \parallel \leq \frac{1}{k} \parallel H_{i,k}^T Y_{i,k} \parallel \leq 1 \leq M_m \tag{25}$$

for any agent i with $\sigma_{i,k} = m$. Therefore, $\theta_{i,k+1} = \theta'_{i,k+1}$ by (12), and $\sigma_{i,k+1} = \hat{\sigma}_{i,k} = m$ by (13). Hence by (25) we derive $\parallel \theta_{i,k+1} \parallel \leq 1$. Consequently, i) holds for $q = 1$.

By definition (10) for $\hat{\sigma}_{j,k}$ we know that $\hat{\sigma}_{j,k} \leq m$ for any agent j with $\sigma_{j,k} < m$ since $k = \tau_m$. If $\hat{\sigma}_{j,k} = m$, then $\theta'_{j,k+1} = \mathbf{0}$ by (11) since $\sigma_{j,k} < \hat{\sigma}_{j,k}$, and hence $\sigma_{j,k+1} = \hat{\sigma}_{j,k} = m$ by (13); If $\hat{\sigma}_{j,k} < m$, then $\sigma_{j,k+1} \leq \hat{\sigma}_{j,k} + 1 \leq m$ by (13). Thus in the case $\sigma_{j,k} < m$ by considering the alternative cases $\hat{\sigma}_{j,k} < m$ and $\hat{\sigma}_{j,k} = m$ we conclude $\sigma_{j,k+1} \leq m$. From (15) we see $\theta_{j,k+1} = \mathbf{0}$ when $\sigma_{j,k+1} = m$. Therefore, ii) holds for $q = 1$.

Step 2: Assume i) and ii) hold for $q = 1, \ldots, p$ with $p < D$. We show that they hold for $q = p + 1$.

From the inductive assumption by (20) (21) we have $\sigma_{k+p} \leq m$ and $\sigma_{i,k+p} = m$. Hence, $\hat{\sigma}_{i,k+p} = m$. Then by (11)

$$\begin{aligned} \theta'_{i,k+p+1} = & \sum_{j \in N_i(k)} \omega_{ij}(k) \theta_{j,k+p} I_{[\sigma_{j,k+p}=m]} \\ & + \frac{1}{k+p} H_{i,k+p}^T (Y_{i,k+p} - H_{i,k+p}\theta_{i,k+p}). \end{aligned} \tag{26}$$

By (20) (21) we know

$$\| \theta_{i,k+p} \| \le 2^p - 1 \text{ if } \sigma_{i,k+p} = m.$$

Then from (26) it follows that

$$\| \theta'_{i,k+p+1} \| \le 2^p - 1 + \frac{1}{k+p} \| H^T_{i,k+p} Y_{i,k+p} \| \\ + \frac{1}{k+p}(2^p - 1) \| H^T_{i,k+p} H_{i,k+p} \| . \tag{27}$$

Since $H^T_{i,k} H_{i,k}$ satisfies (5), similar to (24) it can be shown that for sufficiently large $k \ge m_0 = 2^D - 1$

$$\frac{1}{k} \| H^T_{i,k} H_{i,k} \| \le 1 \quad \forall i \in \mathcal{V},$$

which incorporating with (24)(27) implies that for sufficiently large $m \ge m_0$

$$\| \theta'_{i,k+p+1} \| \le 2^p - 1 + 1 + (2^p - 1) = 2^{p+1} - 1 \le 2^D - 1 < M_m. \tag{28}$$

So, $\theta_{i,k+p+1} = \theta'_{i,k+p+1}$ by (12), and $\sigma_{i,k+p+1} = \hat{\sigma}_{i,k+p} = m$ by (13). Then by (28) we derive $\| \theta_{i,k+p+1} \| \le 2^{p+1} - 1$.

Thus, we conclude that

$$\sigma_{i,k+p+1} = m, \text{ and } \| \theta_{i,k+p+1} \| \le 2^{p+1} - 1 \le M_m. \tag{29}$$

This proves i) for $q = p + 1$.

We now show ii) for $q = p + 1$. By the inductive assumption we have $\sigma_{j,k+p} \le m$. For the case $\sigma_{j,k+p} = m$, as shown in (26)–(29), we have (29) with i replaced by j. So, ii) is valid for $q = p + 1$ for the case $\sigma_{j,k+p} = m$. For the case $\sigma_{j,k+p} < m$ we separately consider the cases $\hat{\sigma}_{j,k+p} = m$ and $\hat{\sigma}_{j,k+p} < m$. For the case $\hat{\sigma}_{j,k+p} = m$, by (11) we have $\theta'_{j,k+p+1} = \mathbf{0}$ since $\sigma_{j,k+p} < \hat{\sigma}_{j,k+p}$, and hence $\sigma_{j,k+p+1} = \hat{\sigma}_{j,k+p} = m$ by (13). For the case $\hat{\sigma}_{j,k+p} < m$, we have $\sigma_{j,k+p+1} \le \hat{\sigma}_{j,k+p} + 1 \le m$ by (13). Therefore, for the case $\sigma_{j,k+p} < m$ we have $\sigma_{j,k+p+1} \le m$. If, in addition, $\sigma_{j,k+p+1} = m$, then $\sigma_{j,k+p+1} > \sigma_{j,k+p}$ and by (15) we conclude $\theta_{j,k+p+1} = \mathbf{0}$. Thus, we have proved ii) for $q = p + 1$.

Consequently, (20) (21) hold for $q = 1, \ldots, D$ by induction. Hence, we conclude that $\tau_{i,m+1} > k + D \quad \forall i \in \mathcal{V}$ for sufficiently large $m \ge m_0$. Since $k = \tau_m$, we have $\tau_{m+1} - \tau_m > D$. Then by Lemma 6 ii) we obtain $\tau_{i,m} \le \tau_m + D \ \forall i \in \mathcal{V}$, and hence $\sigma_{i,\tau_m + D} \ge m$ by noticing $\sigma_{i,\tau_{i,m}} = m$ by definition. On the other hand, from $\tau_{m+1} > \tau_m + D$ it follows that $\sigma_{i,\tau_m + D} \le$

$m \quad \forall i \in \mathcal{V}$. This yields $\sigma_{i,\tau_m+D} = m \quad \forall i \in \mathcal{V}$. From (20) (21) it is seen that $\| \theta_{i,\tau_m+D} \| \leq 2^D - 1 \ \forall i \in \mathcal{V}$ for large enough $m \geq m_0$. Thereby

$$\| \Theta_{\tau_m+D} \| \leq \sqrt{N} \max_i \| \theta_{i,\tau_m+D} \| \leq \sqrt{N}(2^D - 1).$$

Thus, we have shown (19).

Combining Case 1 and Case 2 proves the lemma. □

4. Local properties along bounded subsequences

The following lemma measures the closeness of the sequence $\{\Theta_k\}$ along the bounded subsequence $\{\Theta_{n_k}\}$ with $\sigma_{i,n_k} = \sigma_{n_k} \quad \forall i \in \mathcal{V}$.

Lemma 9. *Let $\{\Theta_{n_k}\}$ be a bounded subsequence with $\sigma_{i,n_k} = \sigma_{n_k} \quad \forall i \in \mathcal{V}$. Assume C1 and C3 hold. Then there exist constants $c_1 > 0$, $c_2 > 0$, $M_0' > 0$, $T > 0$ such that for sufficiently large k*

$$\| \Theta_{m+1} - \Theta_{n_k} \| \leq c_1 T + M_0', \tag{30}$$

$$\| \theta_{m+1} - \theta_{n_k} \| \leq c_2 T \quad \forall m : n_k \leq m \leq m(n_k, T). \tag{31}$$

Proof. Since $\{\Theta_{n_k}\}$ is a bounded subsequence with $\sigma_{i,n_k} = \sigma_{n_k} \quad \forall i \in \mathcal{V}$, by (10) we see $\sigma_{i,n_k} = \hat{\sigma}_{j,n_k} \ \forall i, j \in \mathcal{V}$. Then by (11) we derive

$$\theta'_{i,n_k+1} = \sum_{j \in \mathcal{N}_i(n_k)} \omega_{ij}(n_k)\theta_{j,n_k} + \gamma_{n_k} O_{i,n_k+1}.$$

If there is no truncation at time $n_k + 1$ for any agent $i \in \mathcal{V}$, then

$$\theta_{i,n_k+1} = \theta'_{i,n_k+1} = \sum_{j \in \mathcal{N}_i(n_k)} \omega_{ij}(n_k)\theta_{j,n_k} + \gamma_k O_{i,n_k+1},$$

and the recursion (10)–(13) can be rewritten in the compact form:

$$\Theta_{n_k+s+1} = (W(n_k + s) \otimes \mathrm{I}_M)\Theta_{n_k+s} + \gamma_{n_k+s}(F(\Theta_{n_k+s}) + \varepsilon_{n_k+1+s}) \tag{32}$$

for $s = 0$.

Since the sequence $\{\Theta_{n_k}\}$ is bounded, there exists a constant $C > 0$ and an integer $k_0 > 0$ such that

$$\| \Theta_{n_k} \| \leq C \quad \forall k \geq k_0. \tag{33}$$

Set $F(\Theta) \triangleq \mathrm{col}\{f_1(\theta_1), \ldots, f_N(\theta_N)\}$ with $\Theta = col\{\theta_1, \ldots, \theta_N\}$. Define

$$M_0' = 1 + C(c\rho + 2), \tag{34}$$

$$H_1 = \max_{\Theta}\{\| F(\Theta) \| : \| \Theta \| \leq M_0' + 1 + C\}, \tag{35}$$

$$c_1 = H_1 + c_0(3 + \frac{c(\rho+1)}{1-\rho}), \quad c_2 = \frac{H_1 + c_0}{\sqrt{N}}, \tag{36}$$

where c and ρ are given in (6), and c_0 is given by

$$c_0 = \sqrt{N} c_3 \max_{i \in \mathcal{V}}(\|R_{i,h}\| + tr(R_{i,h})) + 1, \tag{37}$$

where $c_3 = C + M_0' + 1$ and $tr(R_{i,h})$ denotes the trace of $R_{i,h}$. Select $T > 0$ such that

$$c_1 T < 1. \tag{38}$$

For any $k \geq k_0$ define

$$s_k \triangleq \sup\{s \geq n_k : \| \Theta_j - \Theta_{n_k} \| \leq c_1 T + M_0' \quad \forall j : n_k \leq j \leq s\}. \tag{39}$$

From (33) and (38) it follows that for any $k \geq k_0$

$$\| \Theta_s \| \leq c_1 T + \| \Theta_{n_k} \| + M_0' \leq M_0' + 1 + C = c_3 \quad \forall s : n_k \leq s \leq s_k. \tag{40}$$

We now show $s_k > m(n_k, T)$.

Assume the converse that for sufficiently large $k \geq k_1$

$$s_k \leq m(n_k, T). \tag{41}$$

We first show that there exists integer $k_1 > k_0$ such that for all $k \geq k_1$

$$s_k < \tau_{\sigma_{n_k}+1}. \tag{42}$$

We prove (42) for the two alternative cases: i) $\lim_{k\to\infty} \sigma_k = \infty$ and ii) $\lim_{k\to\infty} \sigma_k = \sigma < \infty$. i) Since $\{M_k\}$ is a sequence of positive numbers increasingly diverging to infinity, there exists a positive integer $k_1 > k_0$ such that

$M_{\sigma_{n_k}} > M_0' + 1 + C$ for all $k \geq k_1$. Hence, from (40) we know $s_k < \tau_{\sigma_{n_k}+1}$.
ii) For this case there exists a positive integer $k_1 > k_0$ such that $\sigma_{n_k} = \sigma$ for all $k \geq k_1$, and hence $\tau_{\sigma_{n_k}+1} = \tau_{\sigma+1} = \infty$. This implies (42).

From (42) it follows that (32) holds for $s : 0 \leq s \leq s_k - n_k - 1$.

Next we investigate the property of the noise sequence $\{\varepsilon_{k+1}\}_{k\geq 0}$. Let us decompose the noise $\varepsilon_{i,k+1}$ given in (9) into two parts as

$$\varepsilon_{i,k+1} = \varepsilon^1_{i,k+1} + \varepsilon^2_{i,k+1}, \tag{43}$$

where $\varepsilon^1_{i,k+1} = H^T_{i,k} Y_{i,k} - R_{i,hy}$, and $\varepsilon^2_{i,k+1} = (R_{i,h} - H^T_{i,k} H_{i,k})\theta_{i,k}$. By setting $A_k = H^T_{i,k} Y_{i,k}$, from (7) we see that (16) holds with $\bar{A} = R_{i,hy}$. Hence by (17) we derive

$$\lim_{k\to\infty} \frac{1}{T} \left\| \sum_{m=n_k}^{m(n_k,T_k)} \frac{1}{m} (H^T_{i,m} Y_{i,m} - R_{i,hy}) \right\| = 0 \quad \forall\, T_k \in [0, T]. \tag{44}$$

Then by (41) we conclude that

$$\lim_{k\to\infty} \frac{1}{T} \left\| \sum_{m=n_k}^{s} \frac{1}{m} \varepsilon^1_{i,m+1} \right\| = 0 \quad \forall s : n_k \leq s \leq s_k. \tag{45}$$

Denote by $\lambda_{max}(A)$ the largest eigenvalue of a square matrix A. We have

$$tr(H^T_{i,m} H_{i,m}) \geq \lambda_{max}(H^T_{i,m} H_{i,m}) = \|H_{i,m}\|^2. \tag{46}$$

Therefore, from (40) and (41) it follows that

$$\begin{aligned} \left\| \sum_{m=n_k}^{s} \frac{1}{m} \varepsilon^2_{i,m+1} \right\| &\leq \sum_{m=n_k}^{s} \frac{1}{m} [\|R_{i,h}\| + \|H_{i,m}\|^2] \|\theta_{i,m}\| \\ &\leq c_3 \sum_{m=n_k}^{m(n_k,T)} \frac{1}{m} [\|R_{i,h}\| + tr(H^T_{i,m} H_{i,m})] \quad \forall s : n_k \leq s \leq s_k \end{aligned} \tag{47}$$

for sufficiently large $k \geq k_1$.

By setting $A_k = H^T_{i,k} H_{i,k}$, from (5) it follows that (16) holds with $\bar{A} = R_{i,h}$. Therefore, by (18) we derive

$$\lim_{k\to\infty} \frac{1}{T} \sum_{m=n_k}^{m(n_k,T)} \frac{1}{m} tr(H^T_{i,m} H_{i,m}) = tr(R_{i,h}).$$

Hence

$$\lim_{k\to\infty}\frac{1}{T}\sum_{m=n_k}^{m(n_k,T)}\frac{1}{m}[\|R_{i,h}\|+tr(H_{i,m}^T H_{i,m})]=tr(R_{i,h})+\|R_{i,h}\|. \tag{48}$$

Combining (47) with (48), we obtain

$$\limsup_{k\to\infty}\frac{1}{T}\left\|\sum_{m=n_k}^{s}\gamma_m\varepsilon_{i,m+1}^2\right\|\le c_3(\|R_{i,h}\|+tr(R_{i,h}))\quad\forall s: n_k\le s\le s_k,$$

which incorporating with (45) yields

$$\limsup_{k\to\infty}\frac{1}{T}\left\|\sum_{m=n_k}^{s}\gamma_m\varepsilon_{i,m+1}\right\|\le c_3(\|R_{i,h}\|+tr(R_{i,h}))\quad\forall s: n_k\le s\le s_k.$$

Thus, for sufficiently large $k\ge k_1$

$$\frac{1}{T}\left\|\sum_{m=n_k}^{s}\gamma_m\varepsilon_{i,m+1}\right\|\le c_3(\|R_{i,h}\|+tr(R_{i,h}))+\frac{1}{\sqrt{N}}\quad\forall s: n_k\le s\le s_k.$$

Consequently, from $\left\|\sum_{m=n_k}^{s}\gamma_m\varepsilon_{m+1}\right\|\le\sqrt{N}\max_{i\in\mathcal{V}}\left\|\sum_{m=n_k}^{s}\gamma_m\varepsilon_{i,m+1}\right\|$ we conclude that for sufficiently large $k\ge k_1$

$$\left\|\sum_{m=n_k}^{s}\gamma_m\varepsilon_{m+1}\right\|\le c_0T\quad\forall s: n_k\le s\le s_k. \tag{49}$$

Define

$$Z_{s_k+1}=(W(s_k)\otimes \mathrm{I}_M)\Theta_{s_k}+\gamma_{s_k}(F(\Theta_{s_k})+\varepsilon_{s_k+1}). \tag{50}$$

Set $z_{s_k+1}=\frac{\mathbf{1}^T\otimes \mathrm{I}_M}{N}Z_{s_k+1}$. By $\frac{\mathbf{1}^T\otimes \mathrm{I}_M}{N}(W(s)\otimes \mathrm{I}_M)=\frac{\mathbf{1}^T\otimes \mathrm{I}_M}{N}\ \ \forall s\ge 0$ we derive

$$\begin{aligned} z_{s_k+1}&=\theta_{s_k}+\frac{\mathbf{1}^T\otimes \mathrm{I}_M}{N}\gamma_s(F(\Theta_{s_k})+\varepsilon_{s_k+1})\\ &=\theta_{n_k}+\frac{\mathbf{1}^T\otimes \mathrm{I}_M}{N}\sum_{m=n_k}^{s_k}\gamma_m(F(\Theta_m)+\varepsilon_{m+1}), \end{aligned} \tag{51}$$

and hence by (35) (40) (41) and (49) it follows that

$$\begin{aligned}
(52)\qquad \|z_{s_k+1}-\theta_{n_k}\| &\le \left\|\frac{\mathbf{1}^T\otimes \mathrm{I}_M}{N}\right\| \left\|\sum_{m=n_k}^{s_k}\gamma_m(F(\Theta_m)+\varepsilon_{m+1})\right\| \\
&\le \frac{1}{\sqrt{N}}\sum_{m=n_k}^{s_k}\gamma_m \parallel F(\Theta_m)\parallel + \frac{1}{\sqrt{N}}\left\|\sum_{m=n_k}^{s_k}\gamma_m\varepsilon_{m+1}\right\| \\
&\le \frac{H_1+c_0}{\sqrt{N}}T = c_2T \quad \text{for sufficiently large } k\ge k_1.
\end{aligned}$$

Define $Z_{\perp,s_k+1}=Z_{s_k+1}-(\mathbf{1}\otimes \mathrm{I}_m)z_{s_k+1}$. Since $W(k)$ $\forall k\ge 0$ are doubly stochastic, it is seen that

$$\begin{aligned}
Z_{\perp,s_k+1} &= \left[\left(\Phi(s_k,n_k)-\frac{1}{N}\mathbf{1}\mathbf{1}^T\right)\otimes \mathrm{I}_M\right]\Theta_{n_k} \\
&\quad+\sum_{m=n_k}^{s_k}\gamma_m\left[\left(\Phi(s_k-1,m)-\frac{1}{N}\mathbf{1}\mathbf{1}^T\right)\otimes \mathrm{I}_M\right]F(\Theta_m) \\
&\quad+\sum_{m=n_k}^{s_k}\gamma_m\left[\left(\Phi(s_k-1,m)-\frac{1}{N}\mathbf{1}\mathbf{1}^T\right)\otimes \mathrm{I}_M\right]\varepsilon_{m+1}.
\end{aligned}$$

Noting that $\|F(\theta_m)\|\le H_1\quad \forall m: n_k\le m\le s_k$ by (35) (40), from (6) (33) we see that for sufficiently large $k\ge k_1$

$$\begin{aligned}
(53)\quad \parallel Z_{\perp,s_k+1}\parallel &\le Cc\rho^{s_k+1-n_k}+\sum_{m=n_k}^{s_k}\gamma_m H_1 c\rho^{s_k-m} \\
&\quad+\left\|\sum_{m=n_k}^{s_k}\gamma_m\left[\left(\Phi(s_k-1,m)-\frac{1}{N}\mathbf{1}\mathbf{1}^T\right)\otimes \mathrm{I}_M\right]\varepsilon_{m+1}\right\|.
\end{aligned}$$

By (49) we derive

$$(54)\qquad \parallel \Gamma_s-\Gamma_{n_k-1}\parallel\le c_0T\quad \forall s: n_k\le s\le s_k,$$

where $\Gamma_n\triangleq\sum_{m=1}^n\gamma_m\varepsilon_{m+1}$. Note that

$$\sum_{m=n_k}^{s} \gamma_m(\Phi(s-1,m)\otimes I_l)\varepsilon_{m+1} = \sum_{m=n_k}^{s} (\Phi(s-1,m)\otimes I_l)(\Gamma_m - \Gamma_{m-1})$$
$$= \sum_{m=n_k}^{s} (\Phi(s-1,m)\otimes I_l)(\Gamma_m - \Gamma_{n_k-1})$$
$$- \sum_{m=n_k}^{s} (\Phi(s-1,m)\otimes I_l)(\Gamma_{m-1} - \Gamma_{n_k-1}).$$

Summing by parts, by (6) (54) we derive

$$\left\| \sum_{m=n_k}^{s} \gamma_m(\Phi(s-1,m)\otimes I_l)\varepsilon_{m+1} \right\|$$
$$\leq \| \Gamma_s - \Gamma_{n_k-1} \| + \sum_{m=n_k}^{s-1} \| \Phi(s-1,m) - \Phi(s-1,m+1) \| \| \Gamma_m - \Gamma_{n_k-1} \|$$
$$\leq c_0 T + \sum_{m=n_k}^{s-1} (c\rho^{s-m-1} + c\rho^{s-m}) c_0 T$$
$$\leq c_0 T + \frac{c(\rho+1)}{1-\rho} c_0 T \quad \forall s : n_k \leq s \leq s_k.$$

This incorporating with (49) yields

$$\left\| \sum_{m=n_k}^{s_k} \gamma_m \left[\left(\Phi(s_k-1,m) - \frac{1}{N}\mathbf{1}\mathbf{1}^T \right) \otimes \mathrm{I}_M \right] \varepsilon_{m+1} \right\| \tag{55}$$
$$\leq (2 + \frac{c(\rho+1)}{1-\rho}) c_0 T \quad \text{for sufficiently large } k \geq k_1.$$

From (53) (55) and $\gamma_k \xrightarrow[k\to\infty]{} 0$ it follows that

$$\| Z_{\perp,s_k+1} \| \leq Cc\rho + 1 + \left(2 + \frac{c(\rho+1)}{1-\rho} \right) c_0 T \tag{56}$$

for sufficiently large $k \geq k_1$.

Since $Z_{s_k+1} = Z_{\perp,s_k+1} + (\mathbf{1}\otimes \mathrm{I}_M) z_{s_k+1}$ we derive

$$\| Z_{s_k+1} - \Theta_{n_k} \| = \| (\mathbf{1}\otimes \mathrm{I}_M) z_{s_k+1} + Z_{\perp,s_k+1} - \Theta_{\perp,n_k} - (\mathbf{1}\otimes \mathrm{I}_M)\theta_{n_k} \|$$
$$\leq \| Z_{\perp,s_k+1} \| + \| \Theta_{\perp,n_k} \| + \sqrt{N} \| z_{s_k+1} - \theta_{n_k} \| .$$

Noting that $\|\Theta_{\perp,n_k}\| \le 2C \quad \forall k \ge k_0$ by (33), from (52) and (56) it is seen that for sufficiently large $k \ge k_1$

$$\| Z_{s_k+1} - \Theta_{n_k} \| \le Cc\rho + 1 + \left(2 + \frac{c(\rho+1)}{1-\rho}\right) c_0 T + 2C + \sqrt{N}\frac{H_1 + c_0}{\sqrt{N}} T = M_0' + c_1 T, \tag{57}$$

where M_0' and c_1 are defined by (34) and (36), respectively. Therefore,

$$\| Z_{s_k+1} \| \le \| \Theta_{n_k} \| + M_0' + c_1 T \le M_0' + 1 + C.$$

This means that (32) holds for $s = s_k - n_k$ and $\Theta_{s_k+1} = Z_{s_k+1}$ for sufficiently large $k \ge k_1$. Hence, from (57) we obtain

$$\| \Theta_{s_k+1} - \Theta_{n_k} \| \le M_0' + c_1 T,$$

which contradicts with the definition (39) for s_k. Thereby, $s_k > m(n_k, T)$ holds for sufficiently large $k \ge k_1$. Consequently, from (39) we know that (30) holds for sufficiently large k.

Since $s_k > m(n_k, T)$, similar to (52) it can be proven that (31) holds for sufficiently large k.

The proof is completed. □

The following lemma gives the property of the noise sequence along the bounded subsequence $\{\Theta_{n_k}\}$ with $\sigma_{i,n_k} = \sigma_{n_k} \quad \forall i \in \mathcal{V}$.

Lemma 10. *If C1, C3 hold, then for any $\{n_k\}$ with $\sigma_{i,n_k} = \sigma_{n_k} \quad \forall i \in \mathcal{V}$ and with $\{\Theta_{n_k}\}$ bounded it holds that*

$$\lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \left\| \sum_{m=n_k}^{m(n_k,T_k)} \frac{1}{m} \varepsilon_{k+1} \right\| = 0 \quad \forall T_k \in [0, T]. \tag{58}$$

Proof. Noticing that $\varepsilon_{i,k+1}$ is divided into two parts by (43), from (44) we know that to prove (58) it suffices to verify

$$\lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \left\| \sum_{m=n_k}^{m(n_k,T_k)} \frac{1}{m} (R_{i,h} - H_{i,m}^T H_{i,m}) \theta_{i,m} \right\| = 0 \quad \forall T_k \in [0, T] \tag{59}$$

Inequality (30) in Lemma 9 assures that there exists a $T \in (0, 1)$ such that $m(n_k, T) < \tau_{\sigma_{n_k}+1}$ and $\{\Theta_s : n_k \le s \le m(n_k, T) + 1\}$ are bounded for sufficiently large k. Therefore, there exists a positive constant c_4 such that

$$\parallel \theta_{i,m} \parallel \leq c_4 \quad \forall m : n_k \leq m \leq m(n_k, T) \tag{60}$$

for sufficiently large k.

Combining (53) and (55) we see that there exist positive constants c_3', c_4', c_5' such that for sufficiently large k

$$\parallel \Theta_{\perp,s+1} \parallel \leq c_3 \rho^{s+1-n_k} + c_4 \sup_{m \geq n_k} \gamma_m + c_5 T \quad \forall s : n_k \leq s \leq m(n_k, T). \tag{61}$$

Since $0 < \rho < 1$, there exists a positive integer m' such that $\rho^{m'} < T$. Then $\sum_{m=n_k}^{n_k+m'} \gamma_m \xrightarrow[k\to\infty]{} 0$ by $\lim\limits_{k\to\infty} \gamma_k = 0$. Thus, $n_k + m' < m(n_k, T)$ for sufficiently large k. Therefore, from (61) we know that for sufficiently large k

$$\parallel \Theta_{\perp,s+1} \parallel \leq o(1) + (c_3 + c_5)T \quad \forall s : n_k + m' \leq s \leq m(n_k, T),$$

where $o(1) \to 0$ as $k \to \infty$. From (31) it follows that for sufficiently large k

$$\parallel \theta_{i,s+1} - \theta_{n_k} \parallel \leq o(1) + (c_2 + c_5)T \quad \forall s : n_k + m^{'} \leq s \leq m(n_k, T).$$

Therefore, from here by (46) (48) (60) we conclude that

$$\begin{aligned} &\left\| \sum_{m=n_k}^{m(n_k,T_k)} \frac{1}{m} (R_{i,h} - H_{i,m}^T H_{i,m})(\theta_{i,m} - \theta_{n_k}) \right\| \\ &\leq 2c_4 \sum_{m=n_k}^{n_k+m^{'}} \frac{1}{m} (\parallel R_{i,h} \parallel + tr(H_{i,m}^T H_{i,m})) \\ &\quad + [o(1) + (c_2 + c_5)T] \sum_{m=n_k}^{m(n_k,T)} \frac{1}{m} (\parallel R_{i,h} \parallel + tr(H_{i,m}^T H_{i,m})) \end{aligned}$$

for sufficiently large k and any $T_k \in [0, T]$. Then by (48) we derive

$$\begin{aligned} &\limsup_{k\to\infty} \left\| \sum_{m=n_k}^{m(n_k,T_k)} \frac{1}{m} (R_{i,h} - H_{i,m}^T H_{i,m})(\theta_{i,m} - \theta_{n_k}) \right\| \\ &\leq (c_2 + c_5)([\parallel R_{i,h} \parallel + tr(R_{i,h})]T^2 \quad \forall T_k \in [0, T]. \end{aligned} \tag{62}$$

By (60) we know

$$\left\| \sum_{m=n_k}^{m(n_k,T_k)} \frac{1}{m}(R_{i,h} - H_{i,m}^T H_{i,m})\theta_{n_k} \right\| \leq c_4 \left\| \sum_{m=n_k}^{m(n_k,T_k)} \frac{1}{m}(R_{i,h} - H_{i,m}^T H_{i,m}) \right\|,$$

which incorporating with (5) (17) yields

$$\limsup_{k\to\infty} \left\| \sum_{m=n_k}^{m(n_k,T_k)} \frac{1}{m}(R_{i,h} - H_{i,m}^T H_{i,m})\theta_{n_k} \right\| = 0 \quad \forall T_k \in [0,T]. \tag{63}$$

Therefore, combining (62) (63) we derive (59). The proof is completed. □

5. Proof of Theorem 3

We first show that the truncation number of each agent converges to the same positive integer. Based on this fact, we then prove the strong consistency of the estimates.

5.1. Finiteness of number of truncations

The following lemma says that the truncation number is finite for all agents.

Lemma 11. *If C1, C2, and C3 hold, then*

$$\lim_{k\to\infty} \sigma_k = \sigma < \infty. \tag{64}$$

Proof. From Lemma 8 we know that $\{\Theta_k\}$ contains a bounded subsequence $\{\Theta_{n_k}\}$ with $\sigma_{i,n_k} = \sigma_{n_k}$ $\forall i \in \mathcal{V}$. For this bounded subsequence $\{\Theta_{n_k}\}_{k\geq 1}$, there exists a positive constant c_0 such that $\|\Theta_{n_k}\| \leq c_0$ $\forall k \geq 1$. Thus, $\{\theta_{n_k}\}$ is in the bounded set $\{\theta \in \mathbb{R}^M : \|\theta\| \leq c_0\}$. Note that $v(\theta) = \|R_h\theta - R_{hy}\|^2$ and R_h is positive definite. Then there exists a constant $c_1 > c_0$ such that $\max_{\|\theta\|\leq c_0} v(\theta) < \inf_{\|\theta\|=c_1} v(\theta)$. Since $J = \{\theta^*\}$, there exists a nonempty interval $[\delta_1, \delta_2] \in (\max_{\|\theta\|\leq c_0} v(\theta), \inf_{\|\theta\|=c_1} v(\theta))$ such that $d([\delta_1, \delta_2], v(J)) > 0$.

Assume the converse that $\lim\limits_{k\to\infty} \sigma_k = \infty$. Similar to [11, Lemma 5.4] it can be proven that θ_{n_k} starting from a point in the set $\{\theta \in \mathbb{R}^M : \|\theta\| \leq c_0\}$ crosses the boundary $\{\theta \in \mathbb{R}^M : \| \theta \| = c_1\}$ infinitely many times. Therefore, for the nonempty interval $[\delta_1, \delta_2]$, there are infinitely many crossings $\{v(\theta_{n_k}), \ldots, v(\theta_{m_k})\}$, where by "crossing $[\delta_1, \delta_2]$ by $\{v(\theta_{n_k}), \ldots, v(\theta_{m_k})\}$" we mean that $v(\theta_{n_k}) \leq \delta_1, v(\theta_{m_k}) \geq \delta_2$, and $\delta_1 < v(\theta_s) < \delta_2 \quad \forall s : n_k < s < m_k$. We now show that this is impossible.

In Section 2.2 it is noted that the distributed parameter estimation is equivalent to the distributed root-seeking problem. We want to show that A1-A3 required by Theorem 12 in Appendix are satisfied. Since $\gamma_k = \frac{1}{k}$ we know that A1 holds. By noticing that $f_i(\theta) = -R_{i,h}\theta + R_{i,hy}\ \forall i \in \mathcal{V}$, we see that $f_i(\theta)$ is continuous, and hence A3 holds. Since $f(\theta) = -R_h\theta + R_{hy}$, R_h is positive definite and $J = \{\theta^*\}$, by setting $v(\theta) = \|R_h\theta - R_{hy}\|^2$ A2 is verified. Thus, A1-A3 hold. Further, the noise sequence satisfies (58) along indices $\{n_k\}$ with $\sigma_{i,n_k} = \sigma_{n_k}\ \forall i \in \mathcal{V}$ and with $\{\Theta_{n_k}\}$ bounded. Then similar to [11, Lemma 5.3] we can show that any nonempty interval $[\delta_1, \delta_2]$ with $d([\delta_1, \delta_2], v(J)) > 0$ cannot be crossed by infinitely many sequences $\{v(\theta_{n_k}), \ldots, v(\theta_{m_k})\}$, which yields a contradiction.

Thus, the converse assumption $\lim_{k\to\infty} \sigma_k = \infty$ is not true, and hence (64) holds. □

5.2. Strong consistency

Proof of Theorem 3. Since the algorithm (10)–(13) is in the same form as DSAAWET, we use Theorem 12 in Appendix to prove Theorem 3. For this it suffices to prove assumptions A1-A4, and C3 required by Theorem 12. It has already been shown in the proof of Lemma 11 that A1-A3 hold. Since C3 is assumed in the theorem, we only need to prove that noise condition A4 is satisfied.

Since $\lim_{k\to\infty} \sigma_k = \sigma < \infty$ by Lemma 11, there exists $c_1 > 0$ such that $\|\theta_{i,k}\| \leq c_1\ \forall k \geq 1$. Note that $\lim_{k\to\infty} \frac{1}{k} tr(H_{i,k}^T H_{i,k}) = 0$ by (5). Then by (46)

$$\begin{aligned}
&\frac{1}{k}\|(R_{i,h} - H_{i,k}^T H_{i,k})\theta_{i,k}\| \leq \frac{c_1}{k}(\|R_{i,h}\| + \|H_{i,k}\|^2) \\
&\leq \frac{c_1}{k}(\|R_{i,h}\| + tr(H_{i,k}^T H_{i,k})) \xrightarrow[k\to\infty]{} 0.
\end{aligned}$$

Therefore,

$$\frac{1}{k}(R_{i,h} - H_{i,k}^T H_{i,k})\theta_{i,k} \xrightarrow[k\to\infty]{} 0 \quad \forall i \in \mathcal{V},$$

which incorporating with (23) implies A4 a).

By Lemma 11 and Lemma 6 iii) we know that there exists some positive integer k_0 such that

$$\sigma_{i,k} = \sigma \quad \forall k \geq k_0.$$

Hence the sequence $\{\theta_{i,k}\}$ is bounded for any $i \in \mathcal{V}$. Then by Lemma 10

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\left\|\sum_{m=k}^{m(k,T_k)}\frac{1}{m}\varepsilon_{k+1}\right\| = 0 \quad \forall T_k \in [0,T].$$

So, by the definition of ε_{k+1} we conclude that for any $i \in \mathcal{V}$

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\left\|\sum_{m=k}^{m(k,T_k)}\frac{1}{m}\varepsilon_{i,k+1}\right\| = 0 \quad \forall T_k \in [0,T],$$

which implies A4 b).

By Theorem 12 we conclude that the estimates given by each agent converge to θ^*. $\square$

6. Numerical simulations

In this section, numerical simulations demonstrate the convergence of the proposed algorithm. It is also shown how do both the size of the network and the dimension of the unknown parameter affect the convergence rate.

The i-th component of the parameter $\theta^* \in \mathbb{R}^M$ is set to be $(1+0.1i)\sqrt{i}$. For any $i \in \mathcal{V}$, the regressor $H_{i,k} = (h_{i,k}, \ldots, h_{i,k-M+1})$ is a row vector with its entry $h_{i,k}$ generated by

$$h_{i,k} = (1-\kappa_i)\alpha_i h_{i,k-1} + \sqrt{\kappa_i}\xi_{i,k}. \tag{65}$$

For each agent $i \in \mathcal{V}$, κ_i and α_i are generated according to the uniform distributions over the intervals $[0.2, 0.4]$ and $[0.8, 1.2]$, respectively. $\{\xi_{i,k}\}$ is a sequence of iid random variables $\in \mathcal{N}(0, 0.36)$, and the initial values $h_{i,0}$ for all agents are mutually independent and uniformly distributed over the interval $[-2, 2]$. The sequence $\{h_{i,k}\}$ produced by (65) is neither independent nor strictly stationary, while by [24, Theorem 1.5.3] we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} h_{i,k+m}h_{i,k} = R_i(m) \quad a.s. \tag{66}$$

for all $m \geq 0$ with $R_i(m) = 0.36\kappa_i \sum_{k=0}^{\infty}[(1-\kappa_i)\alpha_i]^{2k+m}$. Since $(1-\kappa_i)\alpha_i \in (0,1)$, by (66) we see that (5) holds. Moreover, it can also be verified that (14) holds for any $0 < \alpha < 0.5$.

The observation noise $v_{i,k}$ in (1) for agent i is a sequence of iid random variables $\in \mathcal{N}(0, 0.09)$. All components of the initial values $x_{i,0}$ for all agents are set to be mutually independent and uniformly distributed over the interval $[-2, 2]$.

Let $\{\theta_{i,k}\}_{k\geq 1}$ be generated by (10)–(13) with the step-size $\frac{1}{k}$ in (11) replaced by $\frac{1}{k^{0.6}}$. Let $\theta_k^j,\ j = 1, \ldots, M$ denote the j-th component of $\theta_k = \frac{1}{N}\sum_{i=1}^N \theta_{i,k}$. Let the average of the square errors be $e(k) = \sum_{i=1}^N \|\theta_{i,k} - \theta^*\|^2/N$. Three simulations are carried out.

Simulation One: Let $N = 100$, and $M = 8$. We divide agents into two groups: the first group is composed of agents indexed from $i = 1$ to $i = 50$, while the second group from $i = 51$ to $i = 100$. For any $k \geq 1$,

- at time $3k - 2$, the communication graph for agents in the first group is strongly connected, while agents in the second group do not communicate with any agent;
- at time $3k - 1$, the communication graph for agents in the second group is strongly connected, while agents in the first group do not communicate with any agent;
- at time $3k$, agent $i : 1 \leq i \leq 50$ in the first group communicates with agent $i + 50$ in the second group through a bidirectional edge, and all nonzero weights are set to be $\frac{1}{2}$.

Explicitly, the matrix $W(k)$ is as follows:

$$W(3k-2) = \begin{pmatrix} W_1 & \mathbf{0} \\ \mathbf{0} & \mathrm{I}_{N/2} \end{pmatrix}, \quad W(3k-1) = \begin{pmatrix} \mathrm{I}_{N/2} & \mathbf{0} \\ \mathbf{0} & W_2 \end{pmatrix},$$
$$W(3k) = \begin{pmatrix} \frac{1}{2}\mathrm{I}_{N/2} & \frac{1}{2}\mathrm{I}_{N/2} \\ \frac{1}{2}\mathrm{I}_{N/2} & \frac{1}{2}\mathrm{I}_{N/2} \end{pmatrix},$$

where $W_1 \in \mathbb{R}^{N/2\times N/2}$ and $W_2 \in \mathbb{R}^{N/2\times N/2}$ are doubly stochastic matrices with positive diagonal entries. Further, the digraphs of W_1 and W_2 are strongly connected.

The simulation results are shown in Figure 1, where it is seen that the estimates converge to the true parameter.

Simulation Two: Set $M = 8$. Let $G(N, p_N)$ denote the Poisson random graph model on N nodes defined in [25], by which it is meant that each edge is included in the graph with probability p_N independently of the rest. Take samples from such a graph $G(N, p_N)$ with $p_N = 6/N$ but only consider the connected ones. Denote by d_i the number of neighboring agents of agent i. Set $W(k) = W = [\omega_{ij}]_{i,j=1}^N\ \ \forall k \geq 1$ with $\omega_{ij} = \frac{1}{d_i}$ when j is the neighboring agent of i.

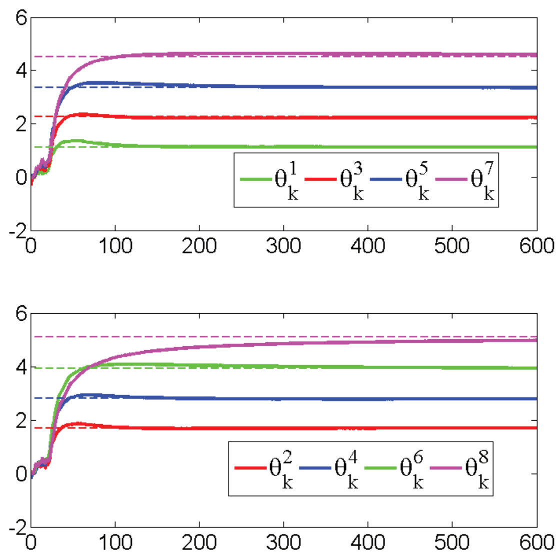

Figure 1: The average estimate θ_k.

The simulation is carried out separately for $N = 20, 100,$ and 1000. The results are shown in Figure 2. It is seen that the network size affects the convergent rate: the larger N yields the slower convergent rate.

Simulation Three: Set $N = 100$. Take a sample graph that is connected from the Poisson random graph $G(N, p)$ with $p = 6/N$. Denote by d_i the number of neighboring agents of agent i. Set $W(k) = W = [\omega_{ij}]_{i,j=1}^{N}$ $\forall k \geq 1$ with $\omega_{ij} = \frac{1}{d_i}$ when j is the neighboring agent of i.

The simulation is carried out separately for $M = 2, 4,$ and 10. The results are shown in Figure 3, from which it is seen that the dimension of the unknown parameter affects the convergent rate: the larger M leads to the lower convergent rate.

7. Concluding remarks

In the paper, a distributed estimation algorithm is calculated at each sensor in the network to estimate the unknown parameter. The estimates are shown to converge to the true parameter when the regressors satisfy some ergodic property. The numerical simulations are given to demonstrate the theoretic result.

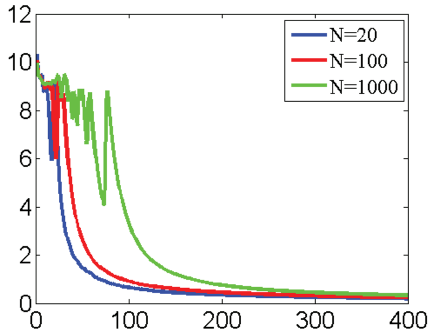

Figure 2: The average square error $e(k)$ for different N.

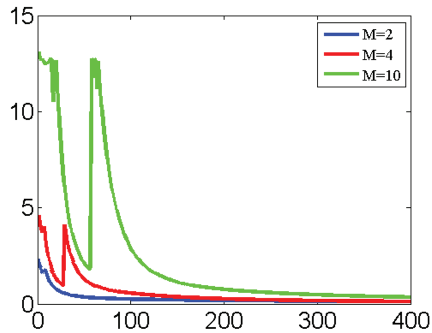

Figure 3: The average square error $e(k)$ for different M.

Some directions for further research include investigating the convergence rate and the asymptotic properties of the distributed algorithm, estimating the unknown time-varying parameter in the sensor networks, and considering the distributed estimation problem with the sign errors [27] or with the regressors corrupted by noises.

8. Appendix: Convergence for DSAAWET

DSAAWET proposed in [11] is used to collectively find the root of the following function

$$f(x) = \sum_{i=1}^{N} f_i(x),$$

where $f_i(\cdot) : \mathbb{R}^M \to \mathbb{R}^M$ is the local function assigned to agent i that can only be observed by i itself. The root set of $f(\cdot)$ is denoted by $J \triangleq \{x \in \mathbb{R}^M : f(x) = 0\}$.

Denote by $x_{i,k} \in \mathbb{R}^M$ the estimate for the root of $f(\cdot)$ given by agent i at time k. While obtaining the information shared from its neighboring agents, agent i has its local observation

$$O_{i,k+1} = f_i(x_{i,k}) + \varepsilon_{i,k+1}, \tag{67}$$

where $\varepsilon_{i,k+1}$ is the observation noise.

With $\{M_k\}$ being a sequence of positive numbers increasingly diverging to infinity and $x^* \in \mathbb{R}^M$ being a given point known to all agents, the estimation sequence $\{x_{i,k}\}_{k\geq 1}$ of agent i is generated as follows:

$$\sigma_{i,0}=0,\quad \hat{\sigma}_{i,k}\triangleq \max_{j\in\mathcal{N}_i(k)}\sigma_{j,k}, \tag{68}$$

$$x'_{i,k+1}=\Bigg(\sum_{j\in\mathcal{N}_i(k)}\omega_{ij}(k)\big(x_{j,k}I_{[\sigma_{j,k}=\hat{\sigma}_{i,k}]}+x^*I_{[\sigma_{j,k}<\hat{\sigma}_{i,k}]}\big) \tag{69}$$
$$+\gamma_k O_{i,k+1}\Bigg)I_{[\sigma_{i,k}=\hat{\sigma}_{i,k}]}+x^*I_{[\sigma_{i,k}<\hat{\sigma}_{i,k}]},$$

$$x_{i,k+1}=x^*I_{[\|x'_{i,k+1}\|>M_{\hat{\sigma}_{i,k}}]}+x'_{i,k+1}I_{[\|x'_{i,k+1}\|\leq M_{\hat{\sigma}_{i,k}}]}, \tag{70}$$

$$\sigma_{i,k+1}=\hat{\sigma}_{i,k}+I_{[\|x'_{i,k+1}\|>M_{\hat{\sigma}_{i,k}}]}, \tag{71}$$

where $O_{i,k+1}$ is defined by (67), $\gamma_k>0$ is the step size.

Let us introduce the following conditions.

A1 $\gamma_k>0, \gamma_k \xrightarrow[k\to\infty]{} 0$, and $\sum\limits_{k=1}^{\infty}\gamma_k=\infty$.

A2 There exists a continuously differentiable function $v(\cdot):\mathbb{R}^l\to\mathbb{R}$ such that

a) $\sup\limits_{\delta\leq d(x,J)\leq\Delta} f^T(x)v_x(x)<0$ for any $\Delta>\delta>0$, where $v_x(\cdot)$ denotes the gradient of $v(\cdot)$ and $d(x,J)=\inf_y\{\| x-y \|: y\in J\}$,

b) $v(J)\triangleq\{v(x): x\in J\}$ is nowhere dense,

c) $\|x^*\|<c_0$ and $v(x^*)<\inf_{\|x\|=c_0}v(x)$ for some positive constant c_0, where x^* is used in (69) (70).

A3 The local functions $f_i(\cdot)$ $\forall i\in\mathcal{V}$ are continuous.

A4 For any $i\in\mathcal{V}$, the noise sequence $\{\varepsilon_{i,k+1}\}_{k\geq 0}$ satisfies

a) $\gamma_k\varepsilon_{i,k+1}\xrightarrow[k\to\infty]{} 0$,

b) $\lim\limits_{T\to 0}\limsup\limits_{k\to\infty}\frac{1}{T}\|\sum\limits_{m=n_k}^{m(n_k,t_k)}\gamma_m\varepsilon_{i,m+1}I_{[\|x_{i,m}\|\leq K]}\|=0$ $\forall t_k\in[0,T]$ for any sufficiently large K along indices $\{n_k\}$ whenever $\{x_{i,n_k}\}$ converges, where $m(k,T)\triangleq\max\{m:\sum_{i=k}^m\gamma_i\leq T\}$.

Theorem 12. *[11, Theorem 3.3] Let $\{x_{i,k}\}$ be generated by* (68)–(71) *for any initial value $x_{i,0}$. Assume A1-A3, and C3 hold. Then*

$$X_{\perp,k}\xrightarrow[k\to\infty]{}\mathbf{0}\quad\textit{and}\quad d(x_k,J)\xrightarrow[k\to\infty]{}0$$

for the sample path ω *for which A4 holds for all agents, where* $x_k = \frac{1}{N}\sum_{i=1}^{N} x_{i,k}$, $X_k = col\{x_{1,k}, \ldots, x_{N,k}\}$, *and* $X_{\perp,k} = X_k - (\mathbf{1} \otimes I_M)x_k$.

References

[1] I. F. Akyildiz, T. Melodia and K. R. Chowdhury, *A survey on wireless multimedia sensor networks.* Computer Networks, **51**(4): 921–960, 2007.

[2] J. Yick, B. Mukherjee and D. Ghosal, *Wireless sensor network survey,* Computer Networks. **52**(12): 2292–2330, 2008.

[3] A. H. Sayed, *Adaption, Learning, and Optimization over Networks.* 2014.

[4] W. Ren and R. W. Beard, *Consensus seeking in multi-agent systems under dynamically changing interaction topologies.* IEEE Trans. Autom. Control, **50**(5): 655–661, 2005.

[5] R. O. Saber and R. M. Murray, *Consensus problems in networks of agents with switching topology ans time-delays.* IEEE Trans. Autom. Control, **49**(9): 1520–1533, 2004.

[6] T. C, Aysal, M. J. Coates and M. G. Rabbat, *Distributed average consensus with dithered quantization.* IEEE Trans. Signal Processing, **56**(10): 4905–4918, 2008.

[7] U. A. Khan, S. Kar and J. M. F. Moura, *Distributed sensor localization in random environments using minimal number of anchor nodes.* IEEE Trans. Signal Processing, **57**(5): 2000–2016, 2009.

[8] A. Nedić and A. Ozdaglar, *Distributed subgradient methods for multi-agent optimization.* IEEE Trans. Autom. Control, **54**(1): 48–61, 2009.

[9] I. Lobel and A. Ozdaglar, *Distributed subgradient methods for convex optimization over random networks.* IEEE Trans. Autom. Control, **56**(6): 1291–1306, 2011.

[10] P. Bianchi, G. Fort and W. Hachem, *Performance of a distributed stochastic approximation algorithm.* IEEE Trans. Inf. Theory, **59**(11): 7405–7418, 2013.

[11] J. Lei and H. F. Chen, *Distributed stochastic approximation algorithm with expanding truncations: algorithm and applications.* `arXiv:1410.7180v3`, 2014.

[12] K. You, Z. Li and L. Xie, *Consensus condition for linear multi-agent systems over randomly switching topologies.* Automatica, **49**(10): 3125–3132, 2013.

[13] R. O. Saber, J. A. Fax and R. M. Murray, *Consensus and cooperation in networked multi-agent systems.* Proc. IEEE, **95**(1): 215–233, 2007.

[14] C. G. Lopes and A. H. Sayed, *Diffusion least-mean square over adaptive networks: formulation and performance analysis.* IEEE Trans. Signal Processing, **56**(7): 3122-3136, 2008.

[15] F. S. Cattivelli and A. H. Sayed, *Diffusion LMS strategies for distributed estimation.* IEEE Trans. Signal Processing, **58**(3): 1035-1048, 2010.

[16] F. S. Cattivelli, C. G. Lopes and A. H. Sayed, *Diffusion recursive least-squares for distributed estimation over adaptive networks.* IEEE Trans. Signal Processing, **56**(5): 1865–1877, 2008.

[17] I. D. Schizas, G. Mateos and G. B. Giannakis, *Distributed LMS for consensus-based in-network adaptive processing.* IEEE Trans. Signal Processing, **57**(6): 2365–2382, 2009.

[18] G. Mateos, I. D. Schizas and G. B. Giannakis, *Distributed recursive least-squares for consensus-based in-network adaptive estimation.* IEEE Trans. Signal Processing, **57**(11): 4583–4588, 2009.

[19] U. Khan and J. M. F. Moura, *Distributed Kalman filtering for sensor networks.* Proc. 46th IEEE Conf. Decision Control, pp. 5492–5498, 2007.

[20] U. Khan and J. M. F. Moura, *Distributing the Kalman filter for large-scale systems.* IEEE Trans. Inf. Theory, **58**(6): 3575–3605, 2008.

[21] S. Stankovic, M. Stankovic and D. Stipanovic, *Decentralized parameter estimation by consensus based stochastic approximation.* IEEE Trans. Autom. Control, **56**(3): 531–543, 2011.

[22] S. Kar, and J. M. F. Moura and K. Ramanan, *Distributed parameter estimation in sensor networks: Nonlinear Observation Models and Imperfect Communication.* IEEE Trans. Inf. Theory, **58**(6): 3575–3605, 2012.

[23] Q. Zhang and J. F. Zhang, *Distributed parameter estimation over unreliable networks with Markovian switching topologies.* IEEE Trans. Autom. Control, **57**(10): 2545–2560, 2012.

[24] H. F. Chen and W. X. Zhao, *Recursive Identification and Parameter Estimation.* CRC Press, 2014.

[25] M. O. Jackson, *Social and Economic Networks.* Princeton University Press, 2008.

[26] Y. S. Chow and H. Teicher, *Probability Theory.* Springer, 1997.

[27] H. F. Chen and G. Yin, *Asymptotic properties of sign algorithms for adaptive filtering.* IEEE Trans. Autom. Control, **58**(9): 1545–1556, 2003.

The Key Laboratory of Systems and Control
Institute of Systems Science
Academy of Mathematics and Systems Science
Chinese Academy of Sciences
Beijing 100190, P. R. China
E-mail address: leijinlong11@mails.ucas.ac.cn; hfchen@iss.ac.cn