J. Math. Anal. Appl. 530 (2024) 127633

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

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

Regular Articles

A new approach to decentralized constrained nonlinear estimation over noisy communication links ☆

Jiexiang Wang [a], Bao-Zhu Guo [a,b,*]

[a] *School of Mathematics and Physics, North China Electric Power University, Beijing, 102206, China*
[b] *School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa*

ARTICLE INFO

Article history:
Received 19 April 2023
Available online 28 July 2023
Submitted by P.-F. Yao

Keywords:
Decentralized nonlinear estimation
Convergence analysis
Unbalanced graph
Combination
Communication noise

ABSTRACT

This paper investigates constrained nonlinear estimation over noisy communication links. The estimation is solved by a set of nodes in a fully distributed fashion. The unknown parameter of interest is assumed to be belonging to a closed convex set. The measurement of each node is nonlinearly related to the unknown parameter. The communication among adjacent nodes is corrupted by the additive communication noises. We propose a decentralized projection consensus+innovation algorithm with communication noises to solve the nonlinear estimation problem and develop a novel approach to analyze its convergence. For the case of fixed graph, by introducing an auxiliary matrix and combination of the graph Laplacian and the auxiliary matrix, we prove that the algorithm converges in mean square and almost surely if the combined persistence of excitation (**CPE**) condition holds and the measurement function satisfies the Lipschitz continuity and monotonicity conditions. Furthermore, for the case of the time-varying graphs, we establish the jointly combined persistence of excitation (**JCPE**) condition guaranteeing convergence in mean square. Both the **CPE** and **JCPE** conditions are proposed for the first time and do not require that the graph is balanced. The **JCPE** condition holds even when the graph is disconnected at infinitely many time instants. A simulation example is presented to demonstrate our theoretical results.

1. Introduction

Recently, decentralized estimation has received a considerable attention due to its wide applications like power spectrum estimation [21], traffic network [7], state estimation in power systems [1,31] and tracking and navigation [30], to name just a few. This paper considers decentralized estimation by a multi-agent system, where each agent/node estimates an underlying parameter of interest by interacting with its neigh-

☆ This work was supported by the National Natural Science Foundation of China under grant no. 12131008.
* Corresponding author.
E-mail address: bzguo@iss.ac.cn (B.-Z. Guo).

https://doi.org/10.1016/j.jmaa.2023.127633

bours and assimilating the local sensed data simultaneously. So far, there has been a large body of works on decentralized linear parameter estimation. Sayed and Lopes [28] proposed a decentralized incremental recursive least-square algorithm, which requires a Hamiltonian cyclic path across agents. Das and Mesbahi [8] proposed a decentralized linear parameter estimation algorithm based on Laplacian dynamics consensus. Schizas et al. [29] studied the decentralized linear estimation problem using the alternating direction method of multipliers. Liu et al. [19] focused on the effects of network topologies on the decentralized linear estimation algorithm. In [2,4,11,20], a diffusion strategy for decentralized estimation was proposed, where each agent fuses both the local estimates and raw measurement data from its neighbours. Kar et al. [15][16] proposed a decentralized consensus+innovations estimation algorithm where each agent fuses local estimates received from its neighbours and assimilates simultaneously the local sensed data to obtain the unknown parameter. Zhang et al. [38] considered decentralized linear estimation with finite Markovian switching graphs and proved that the algorithm converges in mean square and almost surely if all graphs are balanced and jointly contain a spanning tree. The papers [17][18] generalized the work [15] to the nonlinear case where the measurement equation is nonlinear. Sahu et al. [27] further generalized [18] to the constrained case where all the local estimates are constrained in a convex set which contains the unknown parameter. Wang et al. [32] studied decentralized linear consensus+innovation estimation in an uncertain environment. They developed a new approach to analyze the convergence and established weak convergence conditions. Meng and Li [22] proposed a decentralized nonlinear estimator, which is claimed to be able to run over unbalanced networks. The estimator updates two variables simultaneously and maintains a row stochastic matrix and a column stochastic matrix, respectively. Under the smoothness and quadratic gradient growth conditions, they proved consistency of the estimator if the graph is fixed and strongly connected. Besides, there are also many works studying decentralized dynamic estimation [3,5,6,9,23,33–35,37], event-triggered decentralized linear parameter estimation [10][13][36], decentralized estimation under asynchronous communications [14] and decentralized estimation under network attacks [6][37].

However, most of the aforementioned literatures considered the linear case. Although a few works such as [17][18][27] studied the nonlinear case, they made use of the contraction property of the graph Laplacian matrix to analyze the convergence of the algorithms. As such, it is unavoidably required that the network graph is balanced. Although the work [22] has no such limitation, the algorithm therein is more complex than the algorithm developed in this paper. It is noted that the balance of the graph is a very strong condition because it requires that each node's in-degree is always equal to its out-degree. This paper is focused on decentralized nonlinear estimation under time-varying and unbalanced graphs. The communication among adjacent nodes is assumed to be polluted by the additive communication noises. We propose a decentralized projection estimation algorithm of the consensus+innovation type with communication noises to solve the estimation problem and develop a new approach to analyze convergence of the algorithm. For the case of fixed graph, by introducing an auxiliary matrix and combination of the graph Laplacian and the auxiliary matrix, we prove that the algorithm achieves mean square and almost sure convergence if the combined persistence of excitation (**CPE**) condition holds and the measurement function satisfies the Lipschitz continuity and monotonicity conditions. For the case of time-varying graphs, we establish the jointly combined persistence of excitation (**JCPE**) condition under which and the Lipschitz continuity and monotonicity conditions on the measurement function the mean square convergence is obtained. Both the **CPE** and **JCPE** conditions are proposed for the first time and don't require that the graph is balanced. The **JCPE** condition holds even when the graph is disconnected at infinitely many time instants. Compared with [22,27,32] which are mostly relevant to this work, the contributions of this work are summarized as follows:

- We remove the balance restriction on the digraph required in [27] and consider a more realistic scenario that the digraph is time-varying and the communication is noisy.
- Although the approach developed in this paper is inspired by the approach in [32], this paper considers that the measurement is a nonlinear function of the unknown parameter, whereas [32] considered the

linear case. Nonlinearity of the measurement equation causes that the method developed in [32] can't be directly used to address the decentralized nonlinear estimation problem and thus is essentially different with ours.

- Compared with [22], the algorithm of this paper is simpler and we consider that the communication is corrupted by noises and the graph is time-varying.

The rest of the paper is arranged as follows. The problem is formulated in Section 2. Section 3 presents the result for the case of fixed graph and Section 4 presents the result for the case of time-varying graphs. Section 5 is devoted to the proof of the results. In Section 6, a simulation example is presented, followed up by conclusions in Section 7.

2. Problem formulation

Let $\theta \in \Theta \subset \mathbb{R}^n$ be an n-dimensional vector to be estimated, where $\mathbb{R}$ represents the set of real numbers and Θ is the parameter set satisfying the following standard assumption.

Assumption 1. The parameter set Θ is a closed convex set with non-empty interior int(Θ) and $\theta \in$int(Θ).

Assumption 1 is realistic since the parameter to be estimated usually belongs to a known closed convex set such as temperature.

An ad hoc network comprised of a set of agents/nodes is deployed to estimate the unknown parameter. The communication structure of the network is modelled by a directed graph $\mathcal{G} = \{\mathcal{V}, \mathcal{A}_{\mathcal{G}}\}$. Here, $\mathcal{V} = \{1, 2, \cdots, N\}$ is the agent set, $\mathcal{A}_{\mathcal{G}} = [a_{ij}] \in \mathbb{R}^{N\times N}$ where a_{ij} is the weight associated with the link from node j to node i and $a_{ij} > 0$ if and only if the link from node j to node i exists, $a_{ij} = 0$ otherwise. Let $\mathcal{N}_i$ represent the neighbourhood of agent i, given by $\mathcal{N}_i = \{j|a_{ij} > 0\}$. Assume no self-loops, i.e., $a_{ii} = 0, i \in \mathcal{V}$. Denote the in-degree of agent i by $\sum_{j=1}^{N} a_{ij}$ and the out-degree of agent i by $\sum_{j=1}^{N} a_{ji}$. Balance of $\mathcal{G}$ means that for each agent the in-degree is always equal to its out-degree.

We focus on the discrete-time decentralized nonlinear estimation. The measurement of each node at any time instant is nonlinearly related to the true but unknown parameter, modelled by the following equation.

$$z_i(k) = f_i(\theta) + v_i(k), i \in \mathcal{V}, k = 0, 1, \cdots \tag{1}$$

Here, $z_i(k) \in \mathbb{R}^{m_i}$ is the measurement vector of agent i at time k, $f_i(\cdot) : \Theta \to \mathbb{R}^{m_i}$ is the measurement function of agent i with $m_i < n$, $v_i(k)$ is the measurement noise. We assume throughout the paper that $f_i(\cdot)$ is continuously differentiable on int(Θ) and has bounded gradient for all i. Denote the gradient of $f_i(x)$ at any point $y \in \Theta$ by $H_{i,y} = \frac{\partial f_i(x)}{\partial x}\Big|_{x=y}$, which is $m_i \times n$ dimensional. Let $\|\cdot\|$ define the 2-norm of the argument throughout the paper. We make the following assumption.

Assumption 2. For any $x, y \in \Theta$ and each i, there exists a constant $l_i > 0$ such that $\|f_i(x)-f_i(y)\| \le l_i\|x-y\|$.

Remark 1. Assumption 2 is a smoothness condition. It is commonly assumed in the decentralized estimation literature to ensure convergence ([18,22,27]).

The task of each agent in the network is to estimate the unknown parameter when a central processing unit is absent. For each agent i, we use $x_i(k)$ to represent its local estimate state for the unknown parameter at time instant k. Each agent follows a procedure described in the sequel to complete the estimate task. At each time instant $k \ge 0$, agent i first communicates with its neighbour agents through a directed communication link and receives neighbours' local estimates corrupted by the communication noises, which is modelled as

$$y_{ji}(k) = x_j(k) + w_{ji}(k), j \in \mathcal{N}_i, \tag{2}$$

where $y_{ji}(k)$ is the signal received by agent i from its neighbour j and $w_{ji}(k)$ is the associated communication noise at time instant k. Next, a weighted sum of these received signals and its own signal is executed to drive all agents' local estimates to reach consensus and then an innovation term incorporating the measurement information is added to further drive the estimate to update in the direction of the true parameter. Finally a projection operation is needed to make the estimate values fall into the constrained set. The algorithm is formulated as

$$x_i(k+1) = \mathcal{P}_\Theta\Big[x_i(k) + c(k)\sum_{j\in\mathcal{N}_i} a_{ij}(y_{ji}(k) - x_i(k)) + c(k)H^T_{i,x_i(k)}[z_i(k) - f_i(x_i(k))]\Big], i \in \mathcal{V}, \tag{3}$$

where $c(k)$ is the algorithm gain, $\mathcal{P}_\Theta[\cdot]$ is the projection operator and $y_{ji}(k)$ is given by (2). When communication noise is not considered, i.e. $y_{ji}(k) = x_j(k)$, the algorithm (3) degenerates to that of [27].

Denote the overall vector of measurement noises $v(k) = [v_1^T(k), \cdots, v_N^T(k)]^T$ and the overall vector of communication noises $w(k) = [w_{11}(k), ..., w_{N1}(k); ...; w_{1N}(k), ..., w_{NN}(k)]^T$. Denote the σ-fields generated by the measurement and communication noises $\mathcal{F}(k) = \sigma(v(t), w(t), 0 \le t \le k), k \ge 0$. We make the following assumptions.

Assumption 3. The communication noise $w(k)$ is independent of the measurement noise $v(k), \forall k \ge 0$. Both $\{w(k), \mathcal{F}(k), k \ge 0\}$ and $\{v(k), \mathcal{F}(k), k \ge 0\}$ are martingale difference sequences.

Assumption 4. The conditional second-order moments of noises are bounded a.s., i.e.

$$\sup_{k\ge0} \mathbb{E}[\|v(k)\|^2|\mathcal{F}(k-1)] < \infty \ a.s. \ \sup_{k\ge0} \mathbb{E}[\|w(k)\|^2|\mathcal{F}(k-1)] < \infty \ a.s.$$

Assumption 5. $\{c(k), k \ge 0\}$ is a strictly positive real sequence satisfying

$$c(k) \to 0, \sum_{k=0}^{\infty} c(k) = \infty, \sum_{k=0}^{\infty} c^2(k) < \infty.$$

Remark 2. In Assumption 3, the communication noise only needs to be independent of the measurement noise. Both $v_i(k)$ and $w_{ij}(k)$ do not need to be independent with respect to i, j and time k. In other words, we allow spatial and temporal correlation for the measurement and communication noises, respectively. In Assumption 5, the algorithm gain has a general form and doesn't need to have some special form as required in [18,27] and needn't be monotonically decreasing as required in [32].

We next present the result for the case of fixed graph.

3. Case of fixed graph

We consider in this section the case of fixed graph. Denote $\widehat{\mathcal{L}}_\mathcal{G} = \frac{\mathcal{L}_\mathcal{G}^T + \mathcal{L}_\mathcal{G}}{2}$, where $\mathcal{L}_\mathcal{G} = \mathcal{D}_\mathcal{G} - \mathcal{A}_\mathcal{G}$ with $\mathcal{D}_\mathcal{G} \triangleq \text{diag}(\sum_{j=1}^N a_{1j}, \cdots, \sum_{j=1}^N a_{Nj})$. Let $\widehat{\mathcal{G}}$ be the symmetrized graph of $\mathcal{G}$ (see [24] for the definition). It is well known in the graph theory that $\widehat{\mathcal{L}}_\mathcal{G}$ is the Laplacian matrix of $\widehat{\mathcal{G}}$ if and only if $\mathcal{G}$ is balanced ([24]). In this work, $\widehat{\mathcal{L}}_\mathcal{G}$ is no longer the Laplacian matrix of $\widehat{\mathcal{G}}$ due to unbalance of $\mathcal{G}$.

We have the following theorem. The proof is put in Section 5.

Theorem 3.1. *Under Assumptions 1-5, if there exist real symmetric matrices $W_i \in \mathbb{R}^{n\times n}, i \in \mathcal{V}$ such that the following two conditions hold:*

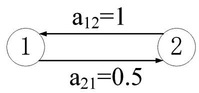

Fig. 1. An unbalanced weighted graph.

(i) $\lambda_{\min}[\widehat{\mathcal{L}}_{\mathcal{G}} \otimes I_n + \mathcal{W}] > 0$,
(ii) $(x-y)^T W_i(x-y) \leq (x-y)^T H_{i,x}^T[f_i(x) - f_i(y)], \forall x, y \in \Theta, \forall i \in \mathcal{V}$,
where I_n *is the n-dimensional identity matrix and* $\mathcal{W} = \text{diag}\{W_1, ..., W_N\}$ *is a block diagonal matrix, then all the local estimates generated by the algorithm (3) converge to* θ *in mean square and almost surely.*

Remark 3. Condition (i) is called the combined persistence of excitation (**CPE**) condition, which is proposed for the first time. It introduces an auxiliary matrix $\mathcal{W}$ and requires that the combined matrix $\widehat{\mathcal{L}}_{\mathcal{G}} \otimes I_n + \mathcal{W}$ is positive definite. At the same time, the auxiliary matrix $\mathcal{W}$ satisfies Condition (ii), which is a monotonicity condition imposed on the measurement function. By choosing $\mathcal{W}$ appropriately, the two conditions can hold under unbalanced graphs, as is illustrated in the following Example 1.

Example 1. For any vector x, let x_i represent its i-th element. Consider an unbalanced digraph with two links shown in Fig. 1. By the definition we have $\widehat{\mathcal{L}}_{\mathcal{G}} = [1, -0.75; -0.75, 0.5]$, which is indefinite. Let $\theta = [\theta_1, \theta_2]^T \in \Theta$ with $\Theta = [-\frac{\pi}{4}, \frac{\pi}{4}]^2 \subset \mathbb{R}^2$. Let the measurement functions $f_1(\theta) = \sin(\theta_1), f_2(\theta) = \sin(\theta_2)$. Thus each agent is not globally observable. The gradients at any $x \in \Theta$ are $H_{1,x} = [\cos(x_1), 0], H_{2,x} = [0, \cos(x_2)]$. Choose $W_1 = \text{diag}\{0.5, 0\}, W_2 = \text{diag}\{0, 0.5\}$. Then,

$$\widehat{\mathcal{L}}_{\mathcal{G}} \otimes I_2 + \mathcal{W} = \begin{pmatrix} 1.5 & 0 & -0.75 & 0 \\ 0 & 1 & 0 & -0.75 \\ -0.75 & 0 & 0.5 & 0 \\ 0 & -0.75 & 0 & 1 \end{pmatrix}. \tag{4}$$

We have $\lambda_{\min}[\widehat{\mathcal{L}}_{\mathcal{G}} \otimes I_n + \mathcal{W}] = 0.0986 > 0$. For node 1, it can be verified that $(x-y)^T H_{1,x}^T[f_1(x) - f_1(y)] - (x-y)^T W_1(x-y) = (x_1 - y_1)(\sin(x_1) - \sin(y_1))\cos(x_1) - 0.5(x_1 - y_1)^2 \geq 0, \forall x_1, y_1 \in [-\frac{\pi}{4}, \frac{\pi}{4}]$. Similarly, we can verify that the condition (ii) holds for node 2. Thus both Conditions (i) and (ii) hold under unbalanced graphs.

Remark 4. In verifying Conditions (i) and (ii) in Theorem 3.1, an important thing is to properly select the matrix $\mathcal{W}$. On the one hand, $\mathcal{W}$ must satisfy that $\sum_{i=1}^N W_i$ is positive definite. To see this, for any $x \in \mathbb{R}^n$, since $\mathcal{L}_{\mathcal{G}}\mathbf{1}_N = \mathbf{0}_N$ by the definition, we have $[\mathbf{1}_N \otimes x]^T[\widehat{\mathcal{L}}_{\mathcal{G}} \otimes I_n + \mathcal{W}][\mathbf{1}_N \otimes x] = x^T[\sum_{i=1}^N W_i]x > 0$ given that Condition (i) holds. On the other hand, the gradient $H_{i,x}$ is usually sparse and we should select W_i to be sparse accordingly to make Condition (ii) hold. For example, in Example 1, $(x-y)^T H_{1,x}^T[f_1(x) - f_1(y)]$ depends on only x_1, y_1 since $H_{1,x}$ does. Then, W_1 should be selected such that $(x-y)^T W_1(x-y)$ also depends on only x_1, y_1, otherwise Condition (ii) doesn't hold.

Remark 5. The matrix $\mathcal{W}$ can be intuitively interpreted as an intermediate bridge that balances the requirements for the strength of network graph information and measurement information to ensure convergence. The matrix $\widehat{\mathcal{L}}_{\mathcal{G}}$ in Condition (i) represents the network graph information, and $(x-y)^T H_{i,x}^T[f_i(x) - f_i(y)]$ on the right side of Condition (ii) represents the measurement information. When the elements of $W_i, i \in \mathcal{V}$ are chosen to be larger, the requirement for the strength of network graph information is lower (i.e., Condition (i) is easier to be satisfied), but at the same time, the requirement for the strength of measurement information is higher, i.e., Condition (ii) is harder to satisfy. Conversely, when the elements of $\mathcal{W}$ are chosen to be smaller, the requirement for the strength of measurement information is lower, but the requirement for the strength of network graph information is higher. For example, if $W_i = I_n$ for all i, then Condition

(i) is easy to hold even for unbalanced and disconnected digraphs but it can be verified that Condition (ii) will require that global observability holds for each node, which is obviously too strong. Hence $W_i = I_n$ shouldn't be chosen.

Remark 6. When the measurement function is linear, i.e., $f_i(x) = H_i x$ with H_i the observation matrix of agent i, Condition (ii) becomes

$$x^T(H_i^T H_i - W_i)x \geq 0, \forall x \in \Theta, i \in \mathcal{V}.$$

If we further choose $W_i = H_i^T H_i, i \in \mathcal{V}$, the above condition automatically holds. Hence, in this case the persistence of excitation condition in [32]:

$$\lambda_{\min}[\widehat{\mathcal{L}}_{\mathcal{G}} \otimes I_n + \operatorname{diag}(H_1^T H_1, \cdots, H_N^T H_N)] > 0$$

serves as the sufficient and necessary condition for the existence of auxiliary variable $\mathcal{W}$ in Theorem 3.1.

Following the proof of Lemma 3.1 of [38], we can prove simply the following Corollary 3.1

Corollary 3.1. *If $\mathcal{W}$ is positive semi-definite, $\mathcal{G}$ is balanced and contains a spanning tree, then Condition (i) in Theorem 3.1 holds if and only if $\lambda_{\min}(\sum_{i=1}^N W_i) > 0$, i.e. $\sum_{i=1}^N W_i$ is positive definite.*

Remark 7. Corollary 3.1 gives an equivalent condition of Condition (i) in Theorem 3.1 given that the digraph is balanced and contains a spanning tree. For general graphs, it is an interesting topic in the future to construct an equivalent condition of $\lambda_{\min}[\widehat{\mathcal{L}}_{\mathcal{G}} \otimes I_n + \mathcal{W}] > 0$ or answer such a mathematical question that whether there exists a function $\mathcal{H}$ such that $\lambda_{\min}[\widehat{\mathcal{L}}_{\mathcal{G}} \otimes I_n + \mathcal{W}] = \mathcal{H}(\lambda_j(\widehat{\mathcal{L}}_{\mathcal{G}}), \lambda_i(W_j), i = 1, ..., n, j = 1, ..., N)$. In particular, if exists, what form of $\mathcal{H}$ is.

4. Case of time-varying graphs

In real networks, the graphs often change over time due to such as packet losses and link failures. In this section, we study that under what conditions the decentralized nonlinear consensus+innovations estimation algorithm with communication noises converges when the graph is time-varying and unbalanced. At this time, the graph weight a_{ij} and the neighbour set $\mathcal{N}_i$ in the algorithm (3) is replaced with $a_{ij}(k)$ and $\mathcal{N}_i(k)$ to accommodate the time-varying feature of the graphs, that is, the algorithm in this case is slightly modified as

$$x_i(k+1) = \mathcal{P}_\Theta\Big[x_i(k) + c(k)\sum_{j \in \mathcal{N}_i(k)} a_{ij}(k)(y_{ji}(k) - x_i(k)) + c(k)H_{i,x_i(k)}^T[z_i(k) - f_i(x_i(k))]\Big], i \in \mathcal{V}, \quad (5)$$

where $y_{ji}(k)$ is the same as in the algorithm (3). We have the following Theorem 4.1. The proof is put in the next section.

Theorem 4.1. *Suppose Assumptions 1-5 and $\sup_{k\geq 0}\|\mathcal{L}_{\mathcal{G}(k)}\| < \infty$. If there exist symmetric matrices $W_i \in \mathbb{R}^{n\times n}, i \in \mathcal{V}$ such that the following two conditions hold:*
(i) $\sum_{k=0}^{\infty}(c(k)\lambda_{\min}[\widehat{\mathcal{L}}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W}]) = \infty$,
(ii) $(x-y)^T W_i(x-y) \leq (x-y)^T H_{i,x}^T[f_i(x) - f_i(y)], \forall x, y \in \Theta, i \in \mathcal{V}$,
where $\mathcal{W} = \operatorname{diag}\{W_1, ..., W_N\}$ is a block diagonal matrix, then all the local estimates generated by the algorithm (5) converge to θ in the mean square sense.

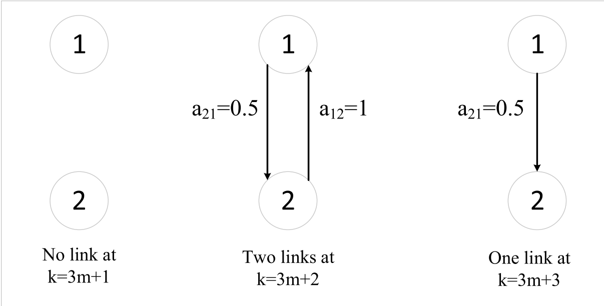

Fig. 2. The graph is empty at time $k = 3m + 1$, is the same as the one in Example 1 at time $k = 3m + 2$ and is with only one link at time $k = 3m + 3, m \geq 0$.

Remark 8. Condition (i) in Theorem 4.1, which is called the jointly combined persistence of excitation (**JCPE**) condition, is very weak. It doesn't require that $\lambda_{\min}[\widehat{\mathcal{L}}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W}] > 0$ at all time instants. Instead, since $\sum_{k=0}^{\infty} c(k) = \infty$, it allows $\lambda_{\min}[\widehat{\mathcal{L}}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W}] = 0$ or < 0 at even infinitely many time instants, which usually means that the graph is disconnected at infinitely many time instants. Take Example 1 in the last section for illustration. We modify the graph therein to be time-varying as shown in Fig. 2 and leave other settings unchanged. From Fig. 2, we know that the graph is unbalanced at each time instant and disconnected at time instants $k = 3m + 1, k = 3m + 3, m \geq 0$. We have $\lambda_{\min}[\widehat{\mathcal{L}}_{\mathcal{G}(3m+1)} \otimes I_n + \mathcal{W}] = 0, \lambda_{\min}[\widehat{\mathcal{L}}_{\mathcal{G}(3m+2)} \otimes I_n + \mathcal{W}] = 0.0986, \lambda_{\min}[\widehat{\mathcal{L}}_{\mathcal{G}(3m+3)} \otimes I_n + \mathcal{W}] = -0.059,\ m \geq 0$. Choose $c(k) = 1/k$ and $\mathcal{W}$ is the same as in Example 1. Then, it can be verified that $\sum_{k=0}^{\infty}(c(k)\lambda_{\min}[\widehat{\mathcal{L}}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W}]) = \sum_{m=0}^{\infty} 0.0986c(3m + 2) - 0.059c(3m + 3) = \sum_{m=0}^{\infty} \frac{0.1188m+0.1778}{(3m+2)(3m+3)} = \infty$. Hence, the **JCPE** condition holds.

Remark 9. We discuss here two variants of Condition (i) in Theorem 4.1. By confining the summation in the **JPE** condition in a fixed time interval $[mh, (m + 1)h - 1], m = 0, 1, \ldots$ with h an positive integer, the condition

$$\sum_{k=mh}^{(m+1)h-1} (c(k)\lambda_{\min}[\widehat{\mathcal{L}}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W}]) \geq \theta c(mh), \tag{6}$$

where $\theta > 0$ is a constant, is sufficient for the **JCPE** since $\sum_{m=0}^{\infty} c(mh) = \infty$. We further exchange the eigenvalue operator and the summation operator in the condition (6) and the condition (6) becomes

$$\lambda_{\min}\Big[\sum_{k=mh}^{(m+1)h-1} (c(k)[\widehat{\mathcal{L}}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W}])\Big] \geq \theta c(mh). \tag{7}$$

The above can hold even if the graph is disconnected at each time instant, as long as the union of the graphs over the interval $[mh, (m + 1)h - 1]$ is strongly connected for all $m \geq 0$. This, however, is not true for the **JCPE** condition because otherwise, $\lambda_{\min}[\widehat{\mathcal{L}}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W}] = 0$ or < 0 for all k, making the **JCPE** fails. From this point of view, it would be better if in Theorem 4.1 the **JCPE** condition is replaced with the condition (7). This has not been solved due to the technical obstacles caused by both the projection operator in the algorithm and the nonlinearity of the measurement function.

5. Proofs of Theorem 3.1 and Theorem 4.1

This section needs three lemmas.

Lemma 5.1 *([25]). Assume* $\{u(k), k \geq 0\}, \{q(k), k \geq 0\}$ *and* $\{\alpha(k), k \geq 0\}$ *are real sequences where* $0 < q(k) \leq 1, \alpha(k) \geq 0, k \geq 0, \sum_{k=0}^{\infty} q(k) = \infty, \frac{\alpha(k)}{q(k)} \to 0, k \to \infty,$ *and*

$$u(k+1) \leq (1-q(k))u(k) + \alpha(k).$$

Then, $\limsup_{k\to\infty} u(k) \leq 0$. *Specially, if* $u(k) \geq 0, k \geq 0$, *then* $u(k) \to 0, k \to \infty$.

Lemma 5.2 *([26]). Assume* $\{x(k), \mathcal{F}(k)\}, \{\alpha(k), \mathcal{F}(k)\}, \{\gamma(k), \mathcal{F}(k)\}$ *are nonnegative adaptive sequences satisfying*

$$\mathbb{E}(x(k+1)|\mathcal{F}(k)) \leq (1+\alpha(k))x(k) - \beta(k) + \gamma(k), k \geq 0, a.s.$$

and $\sum_{k=0}^{\infty}(\alpha(k)+\gamma(k)) < \infty$ *a.s. Then,* $x(k)$ *converges to a finite random variable, a.s. and* $\sum_{k=0}^{\infty} \beta(k) < \infty$. *a.s.*

Lemma 5.3 *([12]). Assume that* $\{s_1(k), k \geq 0\}$ *and* $\{s_2(k), k \geq 0\}$ *are real sequences satisfying* $0 \leq s_2(k) < 1$, $\sum_{k=0}^{\infty} s_2(k) = \infty$ *and* $\lim_{k\to\infty} \frac{s_1(k)}{s_2(k)}$ *exists. Then,*

$$\lim_{k\to\infty} \sum_{i=1}^{k} s_1(i) \prod_{l=i+1}^{k} (1-s_2(l)) = \lim_{k\to\infty} \frac{s_1(k)}{s_2(k)}. \tag{8}$$

Proof of Theorem 3.1. We make the following notation for convenience: $\tilde{x}_i(k+1) \triangleq x_i(k) + c(k) \times \sum_{j\in\mathcal{N}_i}[a_{ij}(x_j(k) - x_i(k))] + c(k)H_{i,x_i(k)}^T[z_i(k) - f_i(x_i(k))], i \in \mathcal{V}$, $\tilde{x}(k) \triangleq [\tilde{x}_1^T(k), \cdots, \tilde{x}_N^T(k)]^T$, $x(k) \triangleq [x_1^T(k), \cdots, x_N^T(k)]^T$, $z(k) \triangleq [z_1^T(k), \cdots, z_N^T(k)]^T$, $f_{\mathbf{1}_N\otimes\theta} \triangleq [f_1^T(\theta), \cdots, f_N^T(\theta)]^T$, $f_{x(k)} \triangleq [f_1^T(x_1(k)), \cdots, f_N^T(x_N(k))]^T$, $D = diag(\alpha_1, ..., \alpha_N)\otimes I_n$ where α_i is the i-th row of $\mathcal{A}_{\mathcal{G}}$, $\mathcal{H}_{x(k)} \triangleq diag(H_{1,x_1(k)}, \cdots, H_{N,x_N(k)})$. Then,

$$\begin{aligned}
&\tilde{x}(k+1) = x(k) + c(k)[\mathcal{A}_{\mathcal{G}} \otimes I_n]x(k) \\
&-c(k)[\mathcal{D}_{\mathcal{G}} \otimes I_n]x(k) + c(k)Dw(k) + c(k)\mathcal{H}_{x(k)}^T[z(k) - f_{x(k)}] \\
&= [I_{Nn} - c(k)\mathcal{L}_{\mathcal{G}} \otimes I_n]x(k) + c(k)Dw(k) \\
&+c(k)\mathcal{H}_{x(k)}^T[f_{\mathbf{1}_N\otimes\theta} - f_{x(k)} + v(k)].
\end{aligned} \tag{9}$$

Denote the overall estimation error by $e(k) \triangleq x(k) - \mathbf{1}_N \otimes \theta$. Let $\tilde{e}(k) \triangleq \tilde{x}(k) - \mathbf{1}_N \otimes \theta$. Noting that $\mathcal{L}_{\mathcal{G}}\mathbf{1}_N = \mathbf{0}_N$ by the definition, we have, by (9), that

$$\begin{aligned}
&\tilde{e}(k+1) \\
&= [I_{Nn} - c(k)\mathcal{L}_{\mathcal{G}} \otimes I_n]e(k) + c(k)Dw(k) \\
&+c(k)\mathcal{H}_{x(k)}^T[f_{\mathbf{1}_N\otimes\theta} - f_{x(k)}] + c(k)\mathcal{H}_{x(k)}^T v(k) \\
&= [I_{Nn} - c(k)\mathcal{L}_{\mathcal{G}} \otimes I_n - c(k)\mathcal{W}]e(k) + c(k)\mathcal{W}e(k) \\
&-c(k)\mathcal{H}_{x(k)}^T[f_{x(k)} - f_{\mathbf{1}_N\otimes\theta}] + c(k)\mathcal{H}_{x(k)}^T v(k) + c(k)Dw(k) \\
&= P(k)e(k) + c(k)A(k) + c(k)B(k) + c(k)Dw(k),
\end{aligned} \tag{10}$$

where $P(k) \triangleq I_{Nn} - c(k)(\mathcal{L}_{\mathcal{G}} \otimes I_n + \mathcal{W})$, $A(k) \triangleq \mathcal{W}e(k) - \mathcal{H}_{x(k)}^T[f_{x(k)} - f_{\mathbf{1}_N\otimes\theta}]$, $B(k) \triangleq \mathcal{H}_{x(k)}^T v(k)$ and the matrix $\mathcal{W}$ is defined in the theorem. By Assumption 1 and the non-expansiveness property of the projection, it follows that $\|x(k+1) - \mathbf{1}_N \otimes \theta\|^2 \leq \|\tilde{x}(k+1) - \mathbf{1}_N \otimes \theta\|^2$. Taking 2-norm for (10) gives

$$\begin{aligned}
&\|e(k+1)\|^2 \leq \|\tilde{e}(k+1)\|^2 \\
&= \|P(k)e(k)\|^2 + c^2(k)\|A(k)\|^2 \\
&+c^2(k)\|B(k)\|^2 + 2c(k)e^T(k)P^T(k)A(k) \\
&+2c(k)e^T(k)P^T(k)B(k) + 2c^2(k)A^T(k)B(k) \\
&+c(k)(P(k)e(k) + c(k)A(k) + c(k)B(k))^T Dw(k) + c^2(k)\|Dw(k)\|^2.
\end{aligned} \tag{11}$$

From Assumption 3, it follows that

$$\begin{aligned}
&\mathbb{E}[(P(k)e(k)+c(k)A(k)+c(k)B(k))^T Dw(k)] \\
&=\mathbb{E}\{\mathbb{E}[(P(k)e(k)+c(k)A(k)+c(k)B(k))^T Dw(k)|\mathcal{F}(k-1)]\} \\
&=\mathbb{E}\{(P(k)e(k)+c(k)A(k)+c(k)B(k))^T D\mathbb{E}[w(k)|\mathcal{F}(k-1)]\}=0.
\end{aligned}$$

Then, taking mathematical expectation on (11) yields

$$\begin{aligned}
&\mathbb{E}[\|e(k+1)\|^2] \le \mathbb{E}[\|\tilde{e}(k+1)\|^2] \\
&=\mathbb{E}[\|P(k)e(k)\|^2]+c^2(k)\mathbb{E}[\|A(k)\|^2] \\
&+c^2(k)\mathbb{E}[\|B(k)\|^2]+2c(k)\mathbb{E}[e^T(k)P^T(k)A(k)] \\
&+2c(k)\mathbb{E}[e^T(k)P^T(k)B(k)]+2c^2(k)\mathbb{E}[A^T(k)B(k)]+c^2(k)\mathbb{E}[\|Dw(k)\|^2]. \qquad (12)
\end{aligned}$$

We next estimate the bounds of the six terms on the right-hand side of (12). From the definition of $P(k)$, we have $P^T(k)P(k)=[I_{Nn}-c(k)(\mathcal{L}_\mathcal{G}^T\otimes I_n+\mathcal{W})][I_{Nn}-c(k)(\mathcal{L}_\mathcal{G}\otimes I_n+\mathcal{W})]=I_{Nn}-c(k)((\mathcal{L}_\mathcal{G}^T+\mathcal{L}_\mathcal{G})\otimes I_n+2\mathcal{W})+c^2(k)(\mathcal{L}_\mathcal{G}^T\otimes I_n+\mathcal{W})(\mathcal{L}_\mathcal{G}\otimes I_n+\mathcal{W})$. By this, it follows that

$$\begin{aligned}
&\|P(k)e(k)\|^2=e^T(k)P^T(k)P(k)e(k) \\
&=\|e(k)\|^2+c^2(k)\|(\mathcal{L}_\mathcal{G}\otimes I_n+\mathcal{W})e(k)\|^2-c(k)e^T(k)[(\mathcal{L}_\mathcal{G}^T+\mathcal{L}_\mathcal{G})\otimes I_n+2\mathcal{W}]e(k) \\
&=\|e(k)\|^2+c^2(k)\|(\mathcal{L}_\mathcal{G}\otimes I_n+\mathcal{W})e(k)\|^2-2c(k)e^T(k)[\widehat{\mathcal{L}}_\mathcal{G}\otimes I_n+\mathcal{W}]e(k) \\
&\le \|e(k)\|^2+c^2(k)\|\mathcal{L}_\mathcal{G}\otimes I_n+\mathcal{W}\|^2\|e(k)\|^2-2c(k)\lambda_{\min}[\widehat{\mathcal{L}}_\mathcal{G}\otimes I_n+\mathcal{W}]\|e(k)\|^2 \\
&\le \|e(k)\|^2-2c(k)c_1\|e(k)\|^2+c^2(k)c_2\|e(k)\|^2 \\
&=[1-2c_1c(k)+c_2c^2(k)]\|e(k)\|^2, \qquad (13)
\end{aligned}$$

where the second equality follows from the definition $\widehat{\mathcal{L}}_\mathcal{G}=\frac{\mathcal{L}_\mathcal{G}^T+\mathcal{L}_\mathcal{G}}{2}$, $c_1\triangleq\lambda_{\min}[\widehat{\mathcal{L}}_\mathcal{G}\otimes I_n+\mathcal{W}]$ and $c_2\triangleq\|\mathcal{L}_\mathcal{G}\otimes I_n+\mathcal{W}\|^2$.

By Assumption 2 and the definition of $A(k)$, we have

$$\begin{aligned}
&c^2(k)\|A(k)\|^2=c^2(k)\|\mathcal{W}e(k)-\mathcal{H}_{x(k)}^T[f_{x(k)}-f_{\mathbf{1}_N\otimes\theta}]\|^2 \\
&\le 2c^2(k)\|\mathcal{W}e(k)\|^2+2c^2(k)\|\mathcal{H}_{x(k)}^T[f_{x(k)}-f_{\mathbf{1}_N\otimes\theta}]\|^2 \\
&\le 2c^2(k)[\|\mathcal{W}\|^2\|e(k)\|^2+[\max_{i\in\mathcal{V}} l_i]^2\|\mathcal{H}_{x(k)}\|^2\|e(k)\|^2] \\
&\le c_3c^2(k)\|e(k)\|^2, \qquad (14)
\end{aligned}$$

where $c_3\triangleq 2\|\mathcal{W}\|^2+2[\max_{i\in\mathcal{V}} l_i]^2\sup_{x\in\Theta}\max_{i\in\mathcal{V}}\|H_{i,x}\|^2$.

By Assumption 4 and the definition of $B(k)$, there exists constant $c_4>0$ such that

$$\begin{aligned}
&c^2(k)(\mathbb{E}[\|B(k)\|^2]+\mathbb{E}[\|Dw(k)\|^2]) \\
&\le c^2(k)(\mathbb{E}[\|\mathcal{H}_{x(k)}\|^2\|v(k)\|^2]+\|D\|^2\mathbb{E}[\|w(k)\|^2]) \\
&\le c_4c^2(k), \qquad (15)
\end{aligned}$$

where $c_4=\sup_{x\in\Theta}\max_{i\in\mathcal{V}}\|H_{i,x}\|^2\sup_{k\ge0}\mathbb{E}[\|v(k)\|^2]+\|D\|^2\sup_{k\ge0}\mathbb{E}[\|w(k)\|^2]$.

By the condition (ii) and the definition of $A(k)$, we have

$$e^T(k)A(k)=e^T(k)\mathcal{W}e(k)-e^T(k)\mathcal{H}_{x(k)}^T[f_{x(k)}-f_{\mathbf{1}_N\otimes\theta}]$$

$$
\begin{aligned}
&= \sum_{i=1}^{N}(x_i(k)-\theta)^T W_i(x_i(k)-\theta) \\
&\quad - \sum_{i=1}^{N}(x_i(k)-\theta)^T H_{i,x_i(k)}^T[f_i(x_i(k))-f_i(\theta)] \le 0.
\end{aligned} \tag{16}
$$

From (14) we have $\|A(k)\| \le \sqrt{c_3}\|e(k)\|$. This, together with (16), yields

$$
\begin{aligned}
&c(k)e^T(k)P^T(k)A(k) \\
&= c(k)e^T(k)[I_{Nn} - c(k)(\mathcal{L}_{\mathcal{G}}^T \otimes I_n + \mathcal{W})]A(k) \\
&= c(k)e^T(k)A(k) - c^2(k)e^T(k)[\mathcal{L}_{\mathcal{G}}^T \otimes I_n + \mathcal{W}]A(k) \\
&\le -c^2(k)e^T(k)[\mathcal{L}_{\mathcal{G}}^T \otimes I_n + \mathcal{W}]A(k) \\
&\le c^2(k)\|e(k)\|\|\mathcal{L}_{\mathcal{G}} \otimes I_n + \mathcal{W}\|\|A(k)\| \\
&\le c^2(k)\sqrt{c_2c_3}\|e(k)\|^2.
\end{aligned} \tag{17}
$$

By Assumption 4, it follows that

$$
\begin{aligned}
&\mathbb{E}[e^T(k)P^T(k)B(k)] = \mathbb{E}[e^T(k)P^T(k)\mathcal{H}_{x(k)}^T v(k)] \\
&= \mathbb{E}\{\mathbb{E}[e^T(k)P^T(k)\mathcal{H}_{x(k)}^T v(k)|\mathcal{F}(k-1)]\} \\
&= \mathbb{E}\{e^T(k)P^T(k)\mathcal{H}_{x(k)}^T \mathbb{E}[v(k)|\mathcal{F}(k-1)]\} = 0.
\end{aligned} \tag{18}
$$

Following a similar procedure to the above, we have

$$
\mathbb{E}[A^T(k)B(k)] = 0. \tag{19}
$$

Applying (13)-(15) and (17)-(19) into (12), we have

$$
\begin{aligned}
&\mathbb{E}[\|e(k+1)\|^2] \\
&\le [1 - 2c_1c(k) + c_2c^2(k)]\mathbb{E}[\|e(k)\|^2] \\
&+c_3c^2(k)\mathbb{E}[\|e(k)\|^2] + c_4c^2(k) + 2\sqrt{c_2c_3}c^2(k)\mathbb{E}[\|e(k)\|^2] \\
&= [1 - 2c_1c(k) + (c_2 + c_3 + 2\sqrt{c_2c_3})c^2(k)]\mathbb{E}[\|e(k)\|^2] + c_4c^2(k).
\end{aligned} \tag{20}
$$

Since $c(k) \to 0$, there exists an integer k_1 such that when $k \ge k_1$, $0 < 2c_1c(k) - (c_2 + c_3 + 2\sqrt{c_2c_3})c^2(k) \le 1$. Hence, by Assumption 5, we know that the relation (20) satisfies the condition of Lemma 5.1 and hence $\mathbb{E}[\|e(k)\|^2] \to 0, k \to \infty$, i.e. all the local estimates converge to the unknown parameter θ in mean square.

By Assumption 4 and the definition of $B(k)$,

$$
\mathbb{E}[2c(k)e^T(k)P^T(k)B(k) + 2c^2(k)A^T(k)B(k)|\mathcal{F}(k-1)] = 0.
$$

Next, taking conditional expectation on (11) and following the steps from (13)-(20), we have

$$
\begin{aligned}
&\mathbb{E}[\|e(k+1)\|^2|\mathcal{F}(k-1)] \\
&\le [1 - 2c_1c(k) + (c_2 + c_3 + 2\sqrt{c_2c_3})c^2(k)]\|e(k)\|^2 + c_4c^2(k) \\
&\le [1 + (c_2 + c_3 + 2\sqrt{c_2c_3})c^2(k)]\|e(k)\|^2 + c_4c^2(k).
\end{aligned} \tag{21}
$$

By Assumption 5, we know that the inequality (21) satisfies the condition of Lemma 5.2. Hence, $\|e(k)\|^2$ converges to a finite random variable as $k \to \infty$ a.s., which together with $\mathbb{E}[\|e(k)\|^2] \to 0, k \to \infty$ gives $\|e(k)\|^2 \to 0, k \to \infty$ a.s., i.e. all the local estimates converge to the unknown parameter θ almost surely. □

Proof of Theorem 4.1. Let $c_2' \triangleq \sup_{k\geq 0} \|\mathcal{L}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W}\|^2$. Replacing c_2 in the proof of Theorem 3.1 with c_2' and then applying (14) and (17) into (11), we obtain

$$\begin{aligned}
&\|e(k+1)\|^2 \\
&\leq \|P(k)e(k)\|^2 + c^2(k)\|A(k)\|^2 + c^2(k)\|B(k)\|^2 \\
&+2c(k)e^T(k)P^T(k)A(k) + 2c(k)e^T(k)P^T(k)B(k) \\
&+2c^2(k)A^T(k)B(k) + c^2(k)\|D(k)w(k)\|^2 \\
&+c(k)(P(k)e(k) + c(k)A(k) + c(k)B(k))^T D(k)w(k) \\
&\leq \|P^T(k)P(k)\|\|e(k)\|^2 + c^2(k)(c_3 + 2\sqrt{c_2'c_3})\|e(k)\|^2 \\
&+c^2(k)\|B(k)\|^2 + 2c(k)e^T(k)P^T(k)B(k) \\
&+2c^2(k)A^T(k)B(k) + c^2(k)\|D(k)w(k)\|^2 \\
&+c(k)(P(k)e(k) + c(k)A(k) + c(k)B(k))^T D(k)w(k).
\end{aligned}$$

Taking the mathematical expectation on the above, by (15), (18), (19) and Assumption 3, we have

$$\begin{aligned}
&\mathbb{E}[\|e(k+1)\|^2] \\
&\leq [\|P^T(k)P(k)\| + c^2(k)(c_3 + 2\sqrt{c_2'c_3})]\mathbb{E}[\|e(k)\|^2] + c^2(k)c_4, \forall k \geq 0.
\end{aligned} \tag{22}$$

Since $c(k) \to 0$, there exists an integer k_2 such that $\lambda_i(2c(k)(\widehat{\mathcal{L}}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W})) < 1, i = 1, ..., Nn, \forall k \geq k_2$ where $\lambda_i(\cdot)$ means the i-th eigenvalue of the argument in the ascending order. Then, by the definition of $P(k)$, we have

$$\begin{aligned}
&\|P^T(k)P(k)\| \\
&= \|I_{Nn} - c(k)((\mathcal{L}^T_{\mathcal{G}(k)} + \mathcal{L}_{\mathcal{G}(k)}) \otimes I_n + 2\mathcal{W}) \\
&+c^2(k)(\mathcal{L}^T_{\mathcal{G}(k)} \otimes I_n + \mathcal{W})(\mathcal{L}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W})\| \\
&\leq \|I_{Nn} - 2c(k)(\widehat{\mathcal{L}}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W})\| + c^2(k)\|\mathcal{L}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W}\|^2 \\
&= \max_{1\leq i\leq Nn} |1 - 2c(k)\lambda_i(\widehat{\mathcal{L}}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W})| + c^2(k)\|\mathcal{L}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W}\| \\
&\leq 1 - 2c(k)\lambda_{\min}(\widehat{\mathcal{L}}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W}) + c^2(k)c_2'.
\end{aligned} \tag{23}$$

Set $m_k = 1 - 2c(k)\lambda_{\min}(\widehat{\mathcal{L}}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W}) + c^2(k)(c_3 + 2\sqrt{c_2'c_3} + c_2')$. Substituting the above formula into (22), we obtain

$$\mathbb{E}[\|e(k+1)\|^2] \leq m_k\mathbb{E}[\|e(k)\|^2] + c^2(k)c_4 + c_4\sum_{i=0}^{k}(m_k\cdots m_{i+1})c^2(i), \forall k \geq k_2. \tag{24}$$

Since

$$m_k \leq e^{-2c(k)\lambda_{\min}(\widehat{\mathcal{L}}_{\mathcal{G}(k)}\otimes I_n+\mathcal{W})+c^2(k)(c_3+2\sqrt{c_2'c_3}+c_2')},$$

it has

$$m_k\cdots m_0 \leq e^{-2\sum_{i=0}^{k} c(i)\lambda_{\min}(\widehat{\mathcal{L}}_{\mathcal{G}(i)}\otimes I_n+\mathcal{W})+(c_3+2\sqrt{c_2'c_3}+c_2')\sum_{i=0}^{k} c^2(i)}.$$

Thus, $\lim_{k\to\infty} m_k\cdots m_0 = 0$ by $\sum_{i=0}^{\infty} c(i)\lambda_{\min}(\widehat{\mathcal{L}}_{\mathcal{G}(i)} \otimes I_n + \mathcal{W}) = \infty$ and $\sum_{i=0}^{\infty} c^2(i) < \infty$. By Lemma 5.3 and the fact $|\lambda_{\min}(\widehat{\mathcal{L}}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W})|$ being bounded, we have

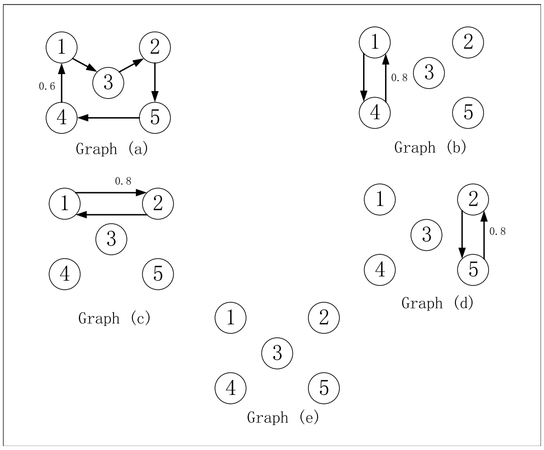

Fig. 3. The graph switches among the five unbalanced digraphs (a), (b), (c), (d), (e) where (e) is an empty graph. All links have weights of 1 except that the link $4 \to 1$ in Graph (a) has weight of 1, and the link $4 \to 1$ in Graph (b), the link $1 \to 2$ in Graph (c) and the link $5 \to 2$ in Graph (d) have weights of 0.8. The graph switches to (a) at $k = 5m + 1$, to (b) at $k = 5m + 2$, to (c) at $k = 5m + 3$, to (d) at $k = 5m + 4$ and to (e) at $k = 5m + 5, m \geq 0$.

$$\lim_{k\to\infty} \sum_{i=0}^{k} (m_k \cdots m_{i+1}) c^2(i) =$$

$$\lim_{k\to\infty} \frac{c^2(k)}{2c(k)\lambda_{\min}(\widehat{\mathcal{L}}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W}) - c^2(k)(c_3 + 2\sqrt{c_2' c_3} + c_2')} = 0.$$

By (22), $\lim_{k\to\infty} \mathbb{E}[\|e(k)\|^2] = 0$. The mean square convergence is thus proved. □

To end this section, we summarize the main contributions of this paper. We propose a decentralized consensus + innovation projection algorithm with communication noises to solve the nonlinear estimation problem and provide a novel approach to analyze convergence of the algorithm. The established convergence conditions are weaker than the existing ones. This is reflected from two aspects. First, the network graph does not need to be balanced, and second for time-varying graph case, the established convergence conditions can still hold even if the network graph is disconnected at even infinitely many time instants, without requiring connectivity to be maintained at all times. The key to achieving these is that we introduce an intermediate matrix $\mathcal{W}$ in (10) in the proof and combine it with the network graph matrix $\mathcal{L}_{\mathcal{G}}$. Then, we use $\mathcal{L}_{\mathcal{G}} + \mathcal{W}$ to contract the difference inequality of the overall estimation error (11) and obtain the key relation (13). In Example 1 of Section 3 and in simulation section, we give examples to illustrate the choice of matrix $\mathcal{W}$.

It is worth noting that in [32], the measurement equations are considered to be linear and hence the observation matrix can be directly combined with the network graph matrix $\mathcal{L}_{\mathcal{G}}$. However, in this paper, the nonlinear case is considered, and there is no observation matrix. Therefore, the method proposed in this paper is sharp contrast to [32].

Remark 10. It seems from (16) in the proof of Theorem 3.1 that only the data pair $(x_i(k), \theta)$ is required to satisfy Condition (ii) in Theorem 3.1. However, we still require all the data pair $(x, y), x, y \in \Theta$ to satisfy Condition (ii) because θ is the unknown parameter that needs to be estimated.

6. Numerical simulation

Consider a network of five agents. The underlying graph switches over time as shown in Fig. 3. Assume $\theta = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5]^T, \Theta = [-\pi/4, \pi/4]^5 \subset \mathbb{R}^5$. The measurement functions are $f_1(x) = \sin x_1, f_2(x) = \sin x_2,$

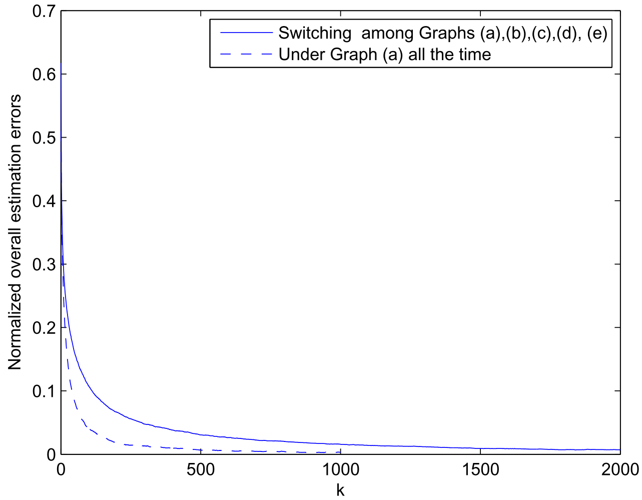

Fig. 4. For the case without communication noises, the figure shows the curves of the normalized overall estimation errors, $\frac{\sum_{i=1}^{5}\|x_i(k)-\theta\|}{25}$ under different kinds of graphs. The solid line represents the error curve under switching graphs; For the purpose of comparison, the dashed line represents the error curve under the fixed Graph (a) all the time.

$f_3(x) = \sin x_3,\ f_4(x) = \sin x_4,\ f_5(x) = \sin x_5$. The measurement noise $\{v(k), k \geq 0\}$ and communication noise $\{w(k), k \geq 0\}$ are assumed to be i.i.d. processes with the standard normal distribution. The gradients

$$H_{1,x} = [\cos(x_1), 0, 0, 0, 0], H_{2,x} = [0, \cos(x_2), 0, 0, 0],$$
$$H_{3,x} = [0, 0, \cos(x_3), 0, 0], H_{4,x} = [0, 0, 0, \cos(x_4), 0], H_{5,x} = [0, 0, 0, 0, \cos(x_5)],$$

where x_i represents the i-th element of x. For each i, W_i is chosen as a diagonal matrix and all elements are zero except that the i-th diagonal element is 0.5. We now verify the condition (i) and (ii) in Theorem 4.1. For node 1, $(x-y)^T W_1(x-y) - (x-y)^T H_{1,x}^T[f_1(x) - f_1(y)] = 0.5(x_1-y_1)^2 - (x_1-y_1)(\sin(x_1) - \sin(y_1))\cos(x_1) \leq 0, \forall x_1, y_1 \in [-\pi/4, \pi/4]$ with y_1 the first element of y. Similarly, we can verify that $(x-y)^T W_i(x-y) - (x-y)^T H_{i,x}^T[f_i(x) - f_i(y)] \leq 0, \forall x, y \in \Theta$ for the other nodes $i = 2, 3, 4, 5$. Hence, the condition (ii) holds. From Fig. 3, the values of $\lambda_{\min}[\widehat{\mathcal{L}}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W}]$ under Graphs (a), (b), (c), (d), (e) are respectively $0.0232, -0.0055, -0.0055, -0.0055, 0$. Choose $c(k) = \frac{1}{k}$. Then, it can be verified that

$$\begin{aligned}&\sum_{k=0}^{\infty}(c(k)\lambda_{\min}[\widehat{\mathcal{L}}_{\mathcal{G}(k)} \otimes I_n + \mathcal{W}]) = \sum_{m=0}^{\infty}\Big[\frac{0.0232}{5m+1} - 0.0055\big(\frac{1}{5m+2} + \frac{1}{5m+3} + \frac{1}{5m+4}\big)\Big]\\ &= \infty.\end{aligned}$$

The condition (i) thus holds. Also, it is easy to see that Assumptions 1-5 hold.

Set the true parameter $\theta = [-\pi/4, -\pi/6, 0, \pi/6, \pi/4]^T$. The initial estimates of all agents are generated according to the standard normal distribution. For the case without communication noises, Fig. 4 shows the curves of the normalized overall estimation errors under switching graphs and fixed graph respectively. It reveals that all the local estimates converge to the unknown parameter asymptotically and the convergence under the switching graphs is slower than that of the fixed graph. For the case with communication noises, we depict Fig. 5. Compared with Fig. 4, Fig. 5 shows that communication noises can greatly reduce the convergence rate of the algorithm. Furthermore, Fig. 6 is depicted with the error curves under different standard deviations of the communication noises, which reveals that the larger the standard deviation, the slower the convergence speed of the proposed algorithm, and the more violent the oscillation of the estimation error curve.

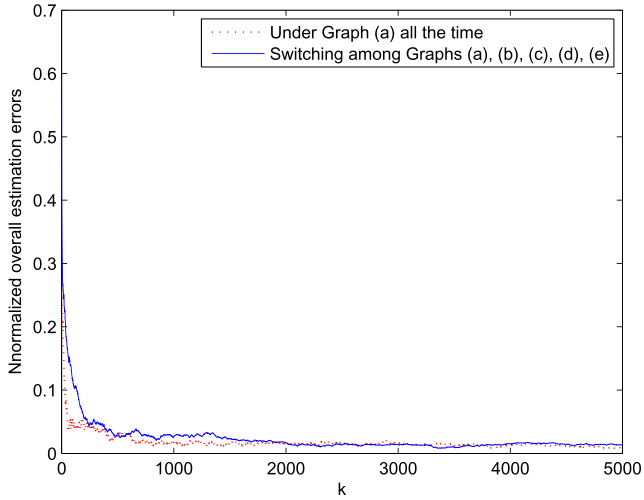

Fig. 5. The case with communication noises under different graphs.

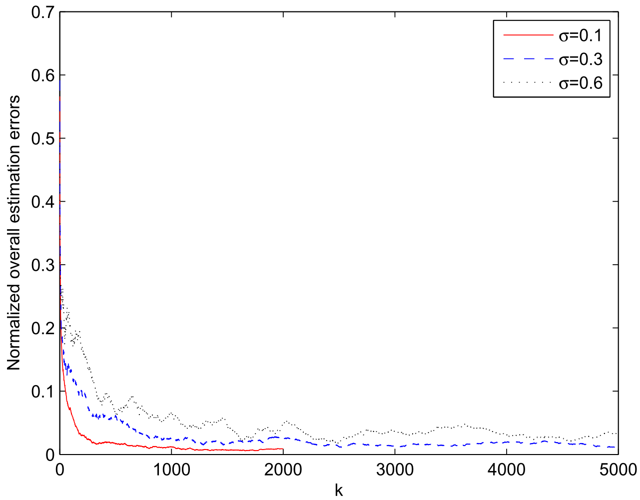

Fig. 6. The case with communication noises under different standard deviations of the communication noises.

7. Conclusions

This paper investigates the decentralized nonlinear estimation using a novel approach. We propose a decentralized consensus+innovation algorithm with communication noises. For the case of fixed graph, by introducing an auxiliary matrix and combination of the graph Laplacian and the auxiliary matrix, we prove that the algorithm achieves mean square and almost sure convergence if the **CPE** condition holds and the measurement function satisfies the Lipschitz continuity and monotonicity conditions. Furthermore, for the case of time-varying graphs, we establish the **JCPE** condition guaranteeing the mean square convergence. Both the **CPE** and **JCPE** conditions do not require that the graph is balanced and the **JCPE** condition holds even when the graph is disconnected at infinitely many time instants. Both the communication noise and measurement noise are allowed to be spatial and temporal correlation. From the examples given in Example 1 and in the simulation section, it seems that finding the matrix $\mathcal{W}$ satisfying the established conditions is not hard for a specific problem. It is an interesting topic in the future to prove that there always exists at least one matrix $\mathcal{W}$ such that Conditions (i) and (ii) of both Theorem 3.1 and Theorem 4.1 hold when the graph is unbalanced and each node is not globally observable.

References

[1] A. Abur, A.G. Exposito, Power System State Estimation: Theory and Implementation, CRC Press, Boca Raton, FL, USA, 2004.

[2] S. Al-Sayed, A.M. Zoubir, A.H. Sayed, Robust distributed estimation by networked agents, IEEE Trans. Signal Process. 65 (15) (2017) 3909–3921.

[3] G. Battistelli, L. Chisci, Stability of consensus extended Kalman filter for distributed state estimation, Automatica 68 (7) (2016) 169–178.

[4] F.S. Cattivelli, A.H. Sayed, Diffusion LMS strategies for distributed estimation, IEEE Trans. Signal Process. 58 (3) (2010) 1035–1048.

[5] F.S. Cattivelli, A.H. Sayed, Diffusion strategies for distributed Kalman filtering and smoothing, IEEE Trans. Autom. Control 56 (9) (2010) 2069–2084.

[6] Y. Chen, S. Kar, J.M.F. Moura, Resilient distributed field estimation, SIAM J. Control Optim. 58 (3) (2020) 1429–1456.

[7] T. Darwish, K.A. Bakar, Traffic density estimation in vehicular ad hoc networks: a review, Ad Hoc Netw. 24 (2014) 337–351.

[8] A. Das, M. Mesbahi, Distributed linear parameter estimation in sensor networks based on Laplacian dynamics consensus algorithm, in: Proceedings of the 3rd IEEE Communications Society on Sensor and Ad Hoc Communications and Networks, Reston, VA, 28, September, 2006, pp. 440–449.

[9] S. Das, J.M.F. Moura, Consensus+innovations distributed Kalman filter with optimized gains, IEEE Trans. Signal Process. 65 (2) (2017) 467–481.

[10] X. Ge, Q. Han, X. Zhang, L. Ding, F. Yang, Distributed event-triggered estimation over sensor networks: a survey, IEEE Trans. Cybern. 50 (3) (2020) 1306–1320.

[11] M.R. Gholami, M. Jansson, E.G. Ström, A.H. Sayed, Diffusion estimation over cooperative multi-agent networks with missing data, IEEE Trans. Signal Inf. Process. Netw. 2 (3) (2016) 276–289.

[12] L. Guo, Time-Varying Stochastic Systems: Stability and Adaptive Theory, second edition, Science Press, Beijing, China, 2020.

[13] X. He, Y. Xing, J. Wu, K.H. Johansson, Event-triggered distributed estimation with decaying communication rate, SIAM J. Control Optim. 60 (2) (2022) 992–1017.

[14] O. Hlinka, F. Hlawatsch, Distributed sequential estimation in asynchronous wireless sensor networks, IEEE Signal Process. Lett. 9 (2015) 997–1010.

[15] S. Kar, J.M.F. Moura, Convergence rate analysis of distributed gossip (linear parameter) estimation: fundamental limits and tradeoffs, IEEE J. Sel. Top. Signal Process. 5 (4) (2011) 674–690.

[16] S. Kar, J.M.F. Moura, Consensus+innovations distributed inference over networks: cooperation and sensing in networked systems, IEEE Signal Process. Mag. 30 (3) (2013) 99–109.

[17] S. Kar, J.M.F. Moura, Asymptotically efficient distributed estimation with exponential family statistics, IEEE Trans. Inf. Theory 60 (8) (2014) 4811–4831.

[18] S. Kar, J.M.F. Moura, K. Ramanan, Distributed parameter estimation in sensor networks: nonlinear observation models and imperfect communication, IEEE Trans. Inf. Theory 58 (6) (2012) 3575–3605.

[19] Y. Liu, C. Li, W.K.S. Tang, Z. Zhang, Distributed estimation over complex networks, Inf. Sci. 197 (2012) 91–104.

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

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

[22] M. Meng, X. Li, Distributed nonlinear estimation over unbalanced directed networks, IEEE Trans. Signal Process. 68 (2020) 6212–6223.

[23] R. Olfati-Saber, Kalman-consensus filter: optimality, stability, and performance, in: Proceedings of the 28th IEEE Conference on Decision and Control, Shangai, China, 16-18, December, 2009, pp. 7036–7042.

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

[25] B.T. Polyak, Introduction to Optimization, Optimization Software Inc., New York, 1987.

[26] H.E. Robbins, D.O. Siegmund, A convergence theorem for non negative almost supermartingales and some applications, in: J.N. Rustagi (Ed.), Optimization Methods in Statistics, Academic Press, New York, 1971, pp. 233–257.

[27] A.K. Sahu, S. Kar, J.M.F. Moura, H.V. Poor, Distributed constrained recursive nonlinear least-squares estimation: algorithms and asymptotics, IEEE Trans. Signal Inf. Process. Netw. 2 (4) (2016) 426–441.

[28] A.H. Sayed, C.G. Lopes, Distributed recursive least-squares over adaptive networks, in: Proc. 40th Asilomar Conf. Signals, Syst., Comput., Pacific Grove, CA, Oct./Nov, 2006, pp. 233–237.

[29] I.D. Schizas, G. Mateos, G.B. Giannakis, Distributed LMS for consesus-based in-network adaptive processing, IEEE Trans. Signal Process. 57 (6) (2009) 2365–2382.

[30] Y.B. Shalom, X.R. Li, T. Kirubarajan, Estimation with Applications to Tracking and Navigation, Wiley, New York, NY, USA, 2001.

[31] J. Wang, T. Li, Distributed multi-area state estimation for power systems with switching communication graphs, IEEE Trans. Smart Grid 12 (1) (2021) 787–797.

[32] J. Wang, T. Li, X. Zhang, Decentralized cooperative online estimation with random observation matrices, communication graphs and time delays, IEEE Trans. Inf. Theory 67 (6) (2021) 4035–4059.

[33] S. Xie, L. Guo, A necessary and sufficient condition for stability of LMS-based consensus adaptive filters, Automatica 93 (8) (2018) 12–19.

[34] S. Xie, L. Guo, Stability of distributed LMS under cooperative stochastic excitation, IEEE Trans. Autom. Control 63 (11) (2018) 3643–3658.
[35] S. Xie, L. Guo, Stability of distributed LMS under cooperative stochastic excitation, IEEE Trans. Autom. Control 63 (11) (2018) 3643–3658.
[36] L. Yan, X. Zhang, Z. Zhang, Y. Yang, Distributed state estimation in sensor networks with event-triggered communication, Nonlinear Dyn. 76 (2014) 169–181.
[37] J. Zhang, R. Blum, X. Lu, D. Conus, Asymptotically optimum distributed estimation in the presence of attacks, IEEE Trans. Signal Process. 63 (2015) 1086–1101.
[38] Q. Zhang, J.F. Zhang, Distributed parameter estimation over unreliable networks with Markovian switching topologies, IEEE Trans. Autom. Control 57 (10) (2012) 2545–2560.