SIAM J. CONTROL OPTIM.
Vol. 60, No. 3, pp. 1519–1545

DISTRIBUTED ORDER ESTIMATION OF ARX MODEL UNDER COOPERATIVE EXCITATION CONDITION*

DIE GAN† AND ZHIXIN LIU‡

Abstract. In this paper, we consider the distributed estimation problem of a linear stochastic system described by an autoregressive model with exogenous inputs when both the system orders and parameters are unknown. We design distributed algorithms to estimate the unknown orders and parameters by combining the proposed local information criterion with the distributed least squares method. The simultaneous estimation for both the system orders and parameters brings challenges for the theoretical analysis. Some analysis techniques, such as double array martingale limit theory, stochastic Lyapunov functions, and martingale convergence theorems are employed. For the case where the upper bounds of the true orders are available, we introduce a cooperative excitation condition, under which the strong consistency of the estimation for the orders and parameters is established. Moreover, for the case where the upper bounds of true orders are unknown, a similar distributed algorithm is proposed to estimate both the orders and parameters, and the corresponding convergence analysis for the proposed algorithm is provided. We remark that our results are obtained without relying on the independency or stationarity assumptions of regression vectors, and the cooperative excitation conditions can show that all sensors can cooperate to fulfill the estimation task even though any individual sensor cannot.

Key words. distributed order estimation, cooperative excitation condition, distributed least squares, convergence

AMS subject classifications. 68W15, 93B30, 93E24

DOI. 10.1137/21M1421362

1. Introduction.

Statistical models are widely used in almost every field of engineering and science, and how to choose or identify an appropriate statistical model to fit observations is an important issue. The order estimation of statistical models is one of the key steps to construct the models. The investigation of the order estimation has many applications in engineering systems, such as radar [1], power systems [2], real seismic traces [3], and physiological systems [4].

In order to estimate the order of statistical models, some criterions are proposed including Akaike information criterion [5], BIC (Bayesian information criterion) [6], CIC [7] (the first "C" emphasizes that the criterion is designed for feedback control systems), and their variants [8]. Based on these information criteria, considerable progress has been made on the order estimation in time series analysis and adaptive estimation and control (e.g., [9], [10], [11], [12], [13]). Some theoretical results are also obtained for the order estimation problem. For example, Hannan and Kavalieris in

*Received by the editors May 20, 2021; accepted for publication (in revised form) January 30, 2022; published electronically May 31, 2022.
https://doi.org/10.1137/21M1421362

Funding: This work was supported by the National Key Research and Development Program of China under grant 2018YFA0703800, by the National Natural Science Foundation of China under grants 11688101 and U21B6001, by Natural Science Foundation of Shandong Province grant ZR2020ZD26, and by the Strategic Priority Research Program of Chinese Academy of Sciences under grant XDA27000000.

†Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, and School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing, 100190, People's Republic of China (gandie@amss.ac.cn).

‡Corresponding author. Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, and School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing, 100190, People's Republic of China (lzx@amss.ac.cn).

[14] introduced an algorithm to estimate the model orders and system parameters, and the convergence of the algorithm was obtained with a stationary input sequence. Chen and Guo in [8] introduced a modification of the BIC criterion to estimate the order of the multidimensional autoregressive model with exogenous inputs (ARX) system, where the true orders are assumed to belong to a known finite set. Furthermore, the relevant results for the estimation of the system orders were generalized in [15] to the case where the upper bounds of the true orders are unknown. After that, some development for the order estimation problem are provided (e.g., [16], [17], [18], [19]). Recently, the genetic algorithm [20] and neural networks [21] were developed for model order estimation problem with good performance. However, the effectiveness of the proposed algorithms in [20] and [21] was verified by some simulation examples without rigorous theoretical analysis.

Over the past decade, with the development of communication technology and computer science, wireless sensor networks have attracted increasing attention from researchers due primarily to their practical applications in engineering systems, such as intelligent transportation and machine health monitoring [22]. We know that in a sensor network, each sensor can only measure partial information of the system due to its limited sensing capacity. In order to estimate the unknown states and system parameters by using data from sensor networks, centralized and distributed methods are two common schemes, where the latter is gaining increasing popularity because of scalability, privacy, and robustness against node and link failures. In distributed algorithms, each sensor only needs to communicate with its neighboring sensors in a certain domain. Some strategies including incremental strategies [23], consensus strategies [24], diffusion strategies [25], and combinations of them [26] are proposed to construct the distributed algorithms. Based on these strategies, the performance analysis of the distributed estimation algorithms are investigated, for example, the consensus-based least mean squares (LMS) (e.g., [27], [28]), the diffusion stochastic gradient descent algorithm [29], the diffusion Kalman filter (e.g., [30], [31]), the diffusion least squares (LS) (e.g., [32], [33], [34]), the diffusion forgetting factor recursive [35]. Most of the corresponding theoretical results are established by requiring the independency, stationarity, or Gaussian assumptions for the regression vectors due to the mathematical difficulty in analyzing the product of random matrices. However, these requirements are hard to satisfy since the regression signals may be correlated due to the multi path effect or feedback. In order to avoid using the independency and stationarity conditions of the regressors, some attempts are made for some distributed estimation algorithms. For the time-invariant unknown parameter, Xie, Zhang, and Guo studied the diffusion LS algorithm and established the convergence result in [36]. For time-varying unknown parameters, they investigated the consensus-based and diffusion LMS algorithm, and proposed the corresponding cooperative information condition to guarantee the stability of the algorithm (e.g., [24], [37]). For the diffusion Kalman filter algorithm, we introduced the collective random observability condition and provided the stability analysis of the distributed Kalman filter algorithm in [31]. We see that the analysis of all these results is established with known system orders. How to construct and analyze the distributed algorithms when the system orders are unknown brings challenges for us.

In this paper, we investigate the distributed estimation problem of linear stochastic systems described by an ARX model with unknown system orders and parameters. The estimates for the orders of each sensor are obtained by minimizing the proposed local information criterion (LIC), and the estimates for unknown system parameters are derived by the distributed LS method where the system orders are replaced by the estimates of orders. The challenges in the theoretical analysis focus on the cumulative effect caused by the system noises and the coupled relationship between the estimates of system orders and parameters. We introduce some mathematical tools including the double array martingale limit theory, martingale convergence theorems, and the stochastic Lyapunov functions to study the convergence of the proposed distributed algorithms. The main contributions of this paper are summarized as follows.

- For the case where the true orders have known upper bounds, we design a distributed algorithm to simultaneously estimate both the system orders and parameters by minimizing the proposed LIC and using the distributed LS method. A cooperative excitation condition is introduced to reflect the joint effect of multiple sensors: the estimation task can still be completed by the cooperation of the sensor networks even if any individual sensor cannot. Under the cooperative excitation condition, the strong consistency of the estimates for both system orders and parameters is established.
- For the case where the upper bounds of true orders are unknown, a similar distributed algorithm is proposed where the growth rate for the upper bounds of the system orders is characterized by a nondecreasing positive function. We employ the double array martingale limit theory to deal with the difficulty arising in analyzing the cumulative effect of the system noises. The convergence analysis for system orders and parameters can also be provided.
- The theoretical results obtained in this paper do not require the assumptions of the independency and stationarity of the regression signals as used in almost all theoretical analysis of the distributed algorithms, which makes it possible to have applications to the stochastic feedback systems.

The rest of this paper is organized as follows. We introduce some preliminaries including graph theory and the observation model in section 2. In section 3, we establish the convergence results when the upper bounds of the true orders are available. The case where the upper bounds of the true orders are unknown is investigated in section 4. A simulation example is given in section 5 to illustrate our theoretical results. We present the conclusion of the paper in section 6.

2. Problem formulation.

2.1. Some preliminaries. In this paper, we use $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ to denote an $m\times n$-dimensional real matrix. For a matrix $\boldsymbol{A}$, we use $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ to denote the largest and smallest eigenvalues of the matrix. $\|\boldsymbol{A}\|$ denotes the Euclidean norm, i.e., $\|\boldsymbol{A}\| = (\lambda_{max}(\boldsymbol{A}\boldsymbol{A}^T))^{\frac{1}{2}}$, where the notation T denotes the transpose operator. We use $det(\cdot)$ to denote the determinant of the corresponding matrix. For a symmetric matrix $\boldsymbol{A}$, if all eigenvalues of $\boldsymbol{A}$ are positive (or nonnegative), then it is a positive definite (semipositive) matrix. Suppose that $\boldsymbol{A} \in \mathbb{R}^{n\times n}$and $\boldsymbol{B} \in \mathbb{R}^{m\times m}$ are two symmetric matrices, and $\boldsymbol{C}$ is an $n \times m$-dimensional matrix. Then by the Rayleigh quotient of the symmetric matrix, we can easily obtain the following inequality:

$$\lambda_{\min}\begin{pmatrix} \boldsymbol{A} & \boldsymbol{C} \\ \boldsymbol{C}^T & \boldsymbol{B} \end{pmatrix} \le \lambda_{\min}(\boldsymbol{A}). \tag{2.1}$$

The matrix inversion formula is used in our analysis, and we list it here.

LEMMA 2.1 ([38]). *For any matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ with suitable dimensions, the following formula,*

$$(\boldsymbol{A}+\boldsymbol{B}\boldsymbol{D}\boldsymbol{C})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D}^{-1} + \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1},$$

holds, provided that the relevant matrices are invertible.

If all elements of a matrix $\boldsymbol{A} = \{a_{ij}\} \in \mathbb{R}^{n\times n}$ are nonnegative, then it is a nonnegative matrix, and furthermore if $\sum_{j=1}^{n} a_{ij} = 1$ holds for all $i \in \{1, \cdots, n\}$, then it is called a stochastic matrix.

Let $\{\boldsymbol{A}_k\}$ be a matrix sequence and $\{b_k\}$ be a positive scalar sequence. Then by $\boldsymbol{A}_k = O(b_k)$ we mean that there exists a constant $C > 0$ such that $\|\boldsymbol{A}_k\| \leq C b_k \ \forall k \geq 0$, and by $\boldsymbol{A}_k = o(b_k)$ we mean that $\lim_{k\to\infty} \|\boldsymbol{A}_k\|/b_k = 0$.

In this paper, our purpose is to design distributed algorithms to estimate both system orders and parameters in a distributed way and establish the corresponding convergence results. We use an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to describe the relationship between sensors, where $\mathcal{V}$ is the set of sensors and $\mathcal{E}$ is the edge set. The adjacency matrix $\mathcal{A} = \{a_{ij}\} \in \mathbb{R}^{n\times n}$ is introduced to reflect the weight of the corresponding edge. The elements of $\mathcal{A}$ satisfy $a_{ij} > 0$ if $(i, j) \in \mathcal{E}$ and $a_{ij} = 0$ otherwise. The set of neighbors of the sensor i is denoted as $N_i = \{j \in \mathcal{V} | (i, j) \in \mathcal{E}\}$, and we assume that sensor i belongs to N_i. A path of length ℓ is a sequence of $\ell + 1$ sensors such that the subsequent senors are connected. The graph $\mathcal{G}$ is called connected if for any two sensors i and j, there is a path connecting them. The diameter $D_{\mathcal{G}}$ of the graph $\mathcal{G}$ is defined as the maximum shortest length of the path between any two sensors. For simplicity of analysis, the convergence of the estimates in this paper is considered under the condition that the weighted adjacency matrix $\mathcal{A}$ is symmetric and stochastic. Thus, it is obvious that $\mathcal{A}$ is doubly stochastic.

2.2. Observation model. We consider a network composed of n sensors. At each time instant t $(t = 0, 1, 2, \ldots)$, the input signal $u_{t,i} \in \mathbb{R}$ and the output signal $y_{t,i} \in \mathbb{R}$ of sensor $i \in \{1, ..., n\}$ are assumed to obey the following linear stochastic ARX model,

$$
\begin{aligned}
y_{t+1,i} = b_1 y_{t,i} + \cdots + b_{p_0} y_{t+1-p_0,i} + c_1 u_{t,i} + \cdots + c_{q_0} u_{t+1-q_0,i} + w_{t+1,i},\\
y_{t,i} = 0, \quad u_{t,i} = 0, \quad \text{for} \quad t \leq 0,
\end{aligned}
\tag{2.2}
$$

where $\{w_{t,i}\}$ is a noise process, p_0, q_0 are unknown true orders $(b_{p_0} \neq 0, c_{q_0} \neq 0)$, and $b_1, \ldots, b_{p_0}$, $c_1, \ldots, c_{q_0}$ are unknown parameters.

Denote the unknown parameter vector $\boldsymbol{\theta}(p, q)$ and the corresponding regression vector $\boldsymbol{\varphi}_{t,i}(p, q)$ as

$$
\boldsymbol{\theta}(p, q) = [b_1, \ldots, b_p, c_1, \ldots, c_q]^T, \tag{2.3}
$$

$$
\boldsymbol{\varphi}_{t,i}(p, q) = [y_{t,i}, \ldots, y_{t+1-p,i}, u_{t,i}, \ldots, u_{t+1-q,i}]^T. \tag{2.4}
$$

If $p > p_0$, then $b_j = 0$ for $p_0 < j \leq p$, and if $q > q_0$, then $c_m = 0$ for $q_0 < m \leq q$. The regression model (2.2) can be rewritten as

$$
y_{t+1,i} = \boldsymbol{\theta}^T(p, q)\boldsymbol{\varphi}_{t,i}(p, q) + w_{t+1,i} \quad (\text{for} \quad p \geq p_0 \quad \text{and} \quad q \geq q_0) \tag{2.5}
$$

$$
\quad\quad = \boldsymbol{\theta}^T(p_0, q_0)\boldsymbol{\varphi}_{t,i}(p_0, q_0) + w_{t+1,i}. \tag{2.6}
$$

The purpose of this paper is to design the distributed algorithm for each sensor by using the local information from its neighbors to estimate both the system orders p_0, q_0 and the parameter vector $\boldsymbol{\theta}(p_0, q_0)$. We know that for the case where the system orders p_0, q_0 are known, the distributed LS algorithm is one of the most basic algorithms to estimate the unknown parameters, and it has wide applications because of its fast convergence rate, e.g., in the area of cloud technologies (e.g., [39]). The details of the distributed LS algorithm can be found in the following Algorithm 2.1 (see [36]).

Algorithm 2.1 Distributed LS algorithm.

For any given $i \in \{1, \ldots, n\}$ and given system order (p, q), begin with an initial estimate $\boldsymbol{\theta}_{0,i}(p,q)$, and an initial positive definite matrix $\boldsymbol{P}_{0,i}(p,q)$. The distributed LS algorithm is recursively defined at time instant $t \geq 0$ as follows.

1: Adaptation.

$$\bar{\boldsymbol{\theta}}_{t+1,i}(p,q) = \boldsymbol{\theta}_{t,i}(p,q) + d_{t,i}(p,q)\boldsymbol{P}_{t,i}(p,q)\boldsymbol{\varphi}_{t,i}(p,q) \cdot (y_{t+1,i} - \boldsymbol{\varphi}_{t,i}^T(p,q)\boldsymbol{\theta}_{t,i}(p,q)), \tag{2.7}$$

$$\bar{\boldsymbol{P}}_{t+1,i}(p,q) = \boldsymbol{P}_{t,i}(p,q) - d_{t,i}(p,q)\boldsymbol{P}_{t,i}(p,q)\boldsymbol{\varphi}_{t,i}(p,q)\boldsymbol{\varphi}_{t,i}^T(p,q)\boldsymbol{P}_{t,i}(p,q), \tag{2.8}$$

$$d_{t,i}(p,q) = [1 + \boldsymbol{\varphi}_{t,i}^T(p,q)\boldsymbol{P}_{t,i}(p,q)\boldsymbol{\varphi}_{t,i}(p,q)]^{-1}. \tag{2.9}$$

2: Diffusion.

$$\boldsymbol{P}_{t+1,i}^{-1}(p,q) = \sum_{j \in N_i} a_{ij} \bar{\boldsymbol{P}}_{t+1,j}^{-1}(p,q), \tag{2.10}$$

$$\boldsymbol{\theta}_{t+1,i}(p,q) = \boldsymbol{P}_{t+1,i}(p,q) \sum_{j \in N_i} a_{ij} \bar{\boldsymbol{P}}_{t+1,j}^{-1}(p,q)\bar{\boldsymbol{\theta}}_{t+1,j}(p,q). \tag{2.11}$$

In this section, for given (p,q), the estimation error between the true parameter and the estimate obtained by Algorithm 2.1 is denoted as $\widetilde{\boldsymbol{\theta}}_{t,i}(p,q)$,

$$\widetilde{\boldsymbol{\theta}}_{t,i}(p,q) = [b_1 - b_{1,t}^i, \ldots, b_p - b_{p,t}^i, c_1 - c_{1,t}^i, \ldots, c_q - c_{q,t}^i]^T, \tag{2.12}$$

where $\{b_{j,t}^i\}_{j=1}^p$ and $\{c_{r,t}^i\}_{r=1}^q$ are denoted as the estimates of the corresponding components of $\boldsymbol{\theta}_{t,i}(p,q)$ obtained by Algorithm 2.1.

We have the following result on the estimation error $\widetilde{\boldsymbol{\theta}}_{t,i}(p,q)$, which will be helpful for the subsequent theoretical analysis.

LEMMA 2.2. *For $p \geq p_0$ and $q \geq q_0$, the following equation holds:*

$$\boldsymbol{P}_{t+1,i}^{-1}(p,q)\widetilde{\boldsymbol{\theta}}_{t+1,i}(p,q) = \sum_{j \in N_i} a_{ij}\boldsymbol{P}_{t,j}^{-1}(p,q)\widetilde{\boldsymbol{\theta}}_{t,j}(p,q) - \sum_{j \in N_i} a_{ij}\boldsymbol{\varphi}_{t,j}(p,q)w_{t+1,j}. \tag{2.13}$$

Proof. For simplicity of expression, we use $d_{t,i}$, $\boldsymbol{\varphi}_{t,i}$, $\boldsymbol{P}_{t,i}$, $\bar{\boldsymbol{P}}_{t+1,i}$, $\bar{\boldsymbol{\theta}}_{t+1,i}$, $\widetilde{\boldsymbol{\theta}}_{t,i}$, and $\boldsymbol{\theta}_{t+1,i}$ to denote $d_{t,i}(p,q)$, $\boldsymbol{\varphi}_{t,i}(p,q)$, $\boldsymbol{P}_{t,i}(p,q)$, $\bar{\boldsymbol{P}}_{t+1,i}(p,q)$, $\bar{\boldsymbol{\theta}}_{t+1,i}(p,q)$, $\widetilde{\boldsymbol{\theta}}_{t,i}(p,q)$, and $\boldsymbol{\theta}_{t+1,i}(p,q)$. By (2.9), we have

$$d_{t,i} = 1 - d_{t,i}\boldsymbol{\varphi}_{t,i}^T\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}. \tag{2.14}$$

Combining this with (2.7) and (2.8), we have

$$\begin{aligned}
\bar{\boldsymbol{\theta}}_{t+1,i} &= (\boldsymbol{I} - d_{t,i}\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}\boldsymbol{\varphi}_{t,i}^T)\boldsymbol{\theta}_{t,i} + d_{t,i}\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}y_{t+1,i} \\
&= (\boldsymbol{I} - d_{t,i}\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}\boldsymbol{\varphi}_{t,i}^T)\boldsymbol{\theta}_{t,i} + \boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}(1 - d_{t,i}\boldsymbol{\varphi}_{t,i}^T\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i})y_{t+1,i} \\
&= (\boldsymbol{P}_{t,i} - d_{t,i}\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}\boldsymbol{\varphi}_{t,i}^T\boldsymbol{P}_{t,i})\boldsymbol{P}_{t,i}^{-1}\boldsymbol{\theta}_{t,i} + (\boldsymbol{P}_{t,i} - d_{t,i}\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}\boldsymbol{\varphi}_{t,i}^T\boldsymbol{P}_{t,i})\boldsymbol{\varphi}_{t,i}y_{t+1,i} \\
&= \bar{\boldsymbol{P}}_{t+1,i}\boldsymbol{P}_{t,i}^{-1}\boldsymbol{\theta}_{t,i} + \bar{\boldsymbol{P}}_{t+1,i}\boldsymbol{\varphi}_{t,i}y_{t+1,i}.
\end{aligned} \tag{2.15}$$

Hence we have

$$\bar{\boldsymbol{P}}_{t+1,i}^{-1}\bar{\boldsymbol{\theta}}_{t+1,i}=\boldsymbol{P}_{t,i}^{-1}\boldsymbol{\theta}_{t,i}+\boldsymbol{\varphi}_{t,i}y_{t+1,i}.$$

Substituting this equation into (2.11) yields

$$\boldsymbol{P}_{t+1,i}^{-1}\boldsymbol{\theta}_{t+1,i}=\sum_{j\in N_i}a_{ij}(\boldsymbol{P}_{t,j}^{-1}\boldsymbol{\theta}_{t,j}+\boldsymbol{\varphi}_{t,j}y_{t+1,j}). \tag{2.16}$$

By (2.8) and Lemma 2.1, we have

$$\bar{\boldsymbol{P}}_{t+1,i}^{-1}=\boldsymbol{P}_{t,i}^{-1}+\boldsymbol{\varphi}_{t,i}\boldsymbol{\varphi}_{t,i}^{T}. \tag{2.17}$$

Hence by (2.10), (2.16), and (2.17), we have

$$\begin{aligned}
&\boldsymbol{P}_{t+1,i}^{-1}\widetilde{\boldsymbol{\theta}}_{t+1,i}=\boldsymbol{P}_{t+1,i}^{-1}\boldsymbol{\theta}-\boldsymbol{P}_{t+1,i}^{-1}\boldsymbol{\theta}_{t+1,i}\\
&=\sum_{j\in N_i}a_{ij}\bar{\boldsymbol{P}}_{t+1,j}^{-1}\boldsymbol{\theta}-\sum_{j\in N_i}a_{ij}(\boldsymbol{P}_{t,j}^{-1}\boldsymbol{\theta}_{t,j}+\boldsymbol{\varphi}_{t,j}y_{t+1,j})\\
&=\sum_{j\in N_i}a_{ij}(\boldsymbol{P}_{t,j}^{-1}+\boldsymbol{\varphi}_{t,j}\boldsymbol{\varphi}_{t,j}^{T})\boldsymbol{\theta}-\sum_{j\in N_i}a_{ij}(\boldsymbol{P}_{t,j}^{-1}\boldsymbol{\theta}_{t,j}+\boldsymbol{\varphi}_{t,j}\boldsymbol{\varphi}_{t,j}^{T}\boldsymbol{\theta}+\boldsymbol{\varphi}_{t,j}w_{t+1,j})\\
&=\sum_{j\in N_i}a_{ij}\boldsymbol{P}_{t,j}^{-1}\widetilde{\boldsymbol{\theta}}_{t,j}-\sum_{j\in N_i}a_{ij}\boldsymbol{\varphi}_{t,j}w_{t+1,j},
\end{aligned}$$

which completes the proof of the lemma. ∎

For the case where the system orders p_0, q_0 are known, Xie, Zhang, and Guo in [36] proved that the distributed LS algorithm can converge to the true parameters almost surely (a.s.) under a cooperative excitation condition. However, when the system orders p_0, q_0 are unknown, the estimation for both the system orders and the parameters makes the design and analysis of the distributed algorithms quite complicated. We will deal with such a problem in the following two sections.

3. Case I: The upper bounds of true orders are known.

In this section, we will first design the distributed algorithm to estimate both the system orders (p_0, q_0) and the parameter vector $\boldsymbol{\theta}(p_0, q_0)$ for the case where the system orders have known upper bounds, i.e.,

$$(p_0,q_0)\in M\triangleq\{(p,q),0\le p\le p^*,0\le q\le q^*\},$$

where p^* and q^* are known upper bounds of the system orders.

For convenience of analysis, we introduce some notations and assumptions,

$$\begin{aligned}
&\boldsymbol{d}_t(p,q)=diag\{d_{t,1}(p,q),...,d_{t,n}(p,q)\},\\
&\boldsymbol{\Phi}_t(p,q)=diag\{\boldsymbol{\varphi}_{t,1}(p,q),...,\boldsymbol{\varphi}_{t,n}(p,q)\},\\
&\boldsymbol{W}_t=[w_{t,1},...w_{t,n}]^T,\\
&\boldsymbol{P}_t(p,q)=diag\{\boldsymbol{P}_{t,1}(p,q),...,\boldsymbol{P}_{t,n}(p,q)\},\\
&\bar{\boldsymbol{P}}_t(p,q)=diag\{\bar{\boldsymbol{P}}_{t,1}(p,q),...,\bar{\boldsymbol{P}}_{t,n}(p,q)\},\\
&\widetilde{\boldsymbol{\Theta}}_t(p,q)=col\{\widetilde{\boldsymbol{\theta}}_{t,1}(p,q),...,\widetilde{\boldsymbol{\theta}}_{t,n}(p,q)\},
\end{aligned}$$

where $col(\cdot,\cdots,\cdot)$ denotes a vector stacked by the specified vectors, and $diag(\cdot,\cdots,\cdot)$ denotes a block matrix formed in a diagonal manner of the corresponding vectors or matrices.

In order to propose and further analyze the distributed algorithm used to estimate both the system order and the parameters, we introduce some assumptions on the network topology, and the observation noise and the regression vectors.

Assumption 3.1. The communication graph $\mathcal{G}$ is connected.

Remark 3.1. For $l > 1$, we denote $\mathcal{A}^l \triangleq [a_{ij}^{(l)}]$ with $\mathcal{A}$ being the weighted adjacency matrix of the graph $\mathcal{G}$, i.e., $a_{ij}^{(l)}$ is the (i, j)th element of the matrix $\mathcal{A}^l$. Under Assumption 3.1, we can easily obtain that $\mathcal{A}^l$ is a positive matrix for $l \geq D_{\mathcal{G}}$ by the theory of the product of stochastic matrices, which means that $a_{ij}^{(l)} > 0$ for any i and j.

Assumption 3.2. For any $i \in \{1, \dots, n\}$, the noise sequence $\{w_{t,i}, \mathscr{F}_t\}$ is a martingale difference sequence, where $\mathscr{F}_t$ is a sequence of nondecreasing σ-algebras generated by $\{y_{k,i}, u_{k,i}, k \leq t, i = 1, 2, , \dots, n\}$, and there exists a constant $\beta > 2$ such that

$$\sup_{t\geq 0} E[|w_{t+1,i}|^{\beta} | \mathscr{F}_t] < \infty \text{a.s.},$$

where $E[\cdot|\cdot]$ denotes the conditional expectation operator.

Assumption 3.3. (cooperative excitation condition I).There exists a sequence $\{a_t\}$ of positive real numbers satisfying $a_t \xrightarrow[t\to\infty]{} \infty$ and

$$\frac{\log r_t(p^*, q^*)}{a_t} \xrightarrow[t\to\infty]{} 0, \quad \frac{a_t}{\lambda_{\min}^{p,q}(t)} \xrightarrow[t\to\infty]{} 0 \text{ for } (p, q) \in M^* \quad \text{a.s.,} \tag{3.1}$$

where $M^* = \{(p_0, q^*), (p^*, q_0)\}$, $r_t(p, q) = \lambda_{\max}\{\boldsymbol{P}_0^{-1}(p, q)\} + \sum_{i=1}^{n} \sum_{k=0}^{t} \|\boldsymbol{\varphi}_{k,i}(p, q)\|^2$, and

$$\lambda_{\min}^{p,q}(t) = \lambda_{\min} \left\{ \sum_{j=1}^{n} \boldsymbol{P}_{0,j}^{-1}(p, q) + \sum_{j=1}^{n} \sum_{k=0}^{t-D_{\mathcal{G}+1}} \boldsymbol{\varphi}_{k,j}(p, q) \boldsymbol{\varphi}_{k,j}^T(p, q) \right\}.$$

Remark 3.2. We give an explanation for the choice of $\{a_t\}$ in Assumption 3.3 for two typical cases: (I) If the regression vectors $\boldsymbol{\varphi}_{k,i}(p^*, q^*)$ are bounded for any $i \in \{1, ..., n\}$, and satisfy the ergodicity property, i.e., there exists a matrix $\boldsymbol{H}_i$ such that $\frac{1}{t} \sum_{k=1}^{t} \boldsymbol{\varphi}_{k,i}(p^*, q^*) \boldsymbol{\varphi}_{k,i}^T(p^*, q^*) \xrightarrow[t\to\infty]{} \boldsymbol{H}_i$ with $\sum_{i=1}^{n} \boldsymbol{H}_i$ being positive definite (see, e.g., [40]), then a_t can be taken as $a_t = t^{\rho}$, $0 < \rho < 1$. (II) If there exist three positive constants s_1, s_2, and s_3 (they may depend on the sample ω) such that

$$\sum_{i=1}^{n} \sum_{k=0}^{t} (\|y_{k,i}\|^2 + \|u_{k,i}\|^2) = O(t^{s_1}) \text{ a.s.,}$$
$$\lambda_{\min}^{p,q}(t) \geq s_2 (\log t)^{1+s_3} \text{ for } (p, q) \in M^* \text{ a.s.,}$$

then Assumption 3.3 can be also satisfied by taking $a_t = (\log t) \log \log t$.

Remark 3.3. For the case where there is only one sensor ($n = 1$), Guo, Chen, and Zhang in [7] investigated the strong consistency of the order estimate under the following conditions,

$$\begin{aligned} &\frac{\log(\sum_{k=0}^{t} \|\boldsymbol{\varphi}_{k,1}(p^*, q^*)\|^2 + 1)}{a_t} \xrightarrow[t\to\infty]{} 0 \text{ a.s.,} \\ &\frac{a_t}{\lambda_{\min}(\sum_{k=0}^{t} \boldsymbol{\varphi}_{k,1}(p, q) \boldsymbol{\varphi}_{k,1}^T(p, q) + \gamma \boldsymbol{I})} \xrightarrow[t\to\infty]{} 0 \text{ for } (p, q) \in M^* \text{ a.s.,} \end{aligned} \tag{3.2}$$

where γ is a positive constant, and $\{a_t, t \geq 1\}$ is a sequence of positive numbers. Assumption 3.3 can be considered as an extension of (3.2) to the case of multiple sensors.

Remark 3.4. Cooperative excitation condition I (i.e., Assumption 3.3) can reflect the joint effect of multiple sensors to some extent: all sensors may cooperatively estimate the unknown orders and parameters under Assumption 3.3 (see Theorems 3.4 and 3.5), even though any individual sensor cannot fulfill the estimation task since the single sensor may lack adequate excitation to satisfy the condition (3.2). We give a simulation example to illustrate this point in section 5

In the following, we propose an algorithm to estimate the system orders p_0 and q_0 in a distributed way. For this propose, we introduce an LIC $L_{t,i}(p, q)$ for the sensor i at the time instant $t \geq 0$,

$$L_{t,i}(p, q) = \sigma_{t,i}(p, q, \boldsymbol{\theta}_{t,i}(p, q)) + (p + q)a_t, \tag{3.3}$$

where $\sigma_{0,i}(p, q, \boldsymbol{\beta}(p, q)) = 0$, and $\sigma_{t,i}(p, q, \boldsymbol{\beta}(p, q))$ is recursively defined for $t > 0$ as follows:

$$\sigma_{t,i}(p, q, \boldsymbol{\beta}(p, q)) = \sum_{j\in N_i} a_{ij}\big(\sigma_{t-1,j}(p, q, \boldsymbol{\beta}(p, q)) + [y_{t,j} - \boldsymbol{\beta}^T(p, q)\boldsymbol{\varphi}_{t-1,j}(p, q)]^2\big). \tag{3.4}$$

With $\sigma_{0,i}(p, q, \boldsymbol{\beta}(p, q)) = 0$, (3.4) is equivalent to the following equation:

$$\sigma_{t,i}(p, q, \boldsymbol{\beta}(p, q)) = \sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)}[y_{k+1,j} - \boldsymbol{\beta}^T(p, q)\boldsymbol{\varphi}_{k,j}(p, q)]^2. \tag{3.5}$$

When the upper bounds of orders are known, the distributed algorithm to estimate the system orders and parameters is put forward by minimizing LIC (i.e., $L_{t,i}(p, q)$) and using Algorithm 2.1. It is clear that in (3.3), the first term is used to minimize the error between the observation signals and the prediction, while the penalty term "$(p + q)a_t$" is introduced to avoid overfitting. The details of the algorithm can be found in Algorithm 3.1.

Algorithm 3.1.

For any given $i \in \{1, \dots, n\}$, the distributed estimation of the system orders and parameters can be obtained at the time instant $t \geq 1$ as follows.

Step 1: For any $(p, q) \in M$, based on $\{\boldsymbol{\varphi}_{k,j}(p, q), y_{k+1,j}\}_{k=1}^{t-1}$, $j \in N_i$, the estimate $\boldsymbol{\theta}_{t,i}(p, q)$ can be obtained by using Algorithm 2.1.

Step 2: (order estimation) With the estimates $\{\boldsymbol{\theta}_{t,i}(p, q)\}_{(p,q)\in M}$ obtained by Step 1, the estimates $(p_{t,i}, q_{t,i})$ of system orders are given by minimizing $L_{t,i}(p, q)$, i.e.,

$$(p_{t,i}, q_{t,i}) = \arg\min_{(p,q)\in M} L_{t,i}(p, q). \tag{3.6}$$

Step 3: (parameter estimation) The estimate $\boldsymbol{\theta}_{t,i}(p_{t,i}, q_{t,i})$ for the unknown parameter $\boldsymbol{\theta}(p_0, q_0)$ can be obtained by using Algorithm 2.1, where the orders (p, q) are replaced by the estimates $(p_{t,i}, q_{t,i})$ obtained in Step 2.

Repeating the above steps, we obtain the order estimates $p_{t,i}, q_{t,i}$ and parameter estimates $\boldsymbol{\theta}_{t,i}(p_{t,i}, q_{t,i})$ for $t \geq 0$ and $i = 1, 2, \dots, n$.

In fact, the estimates of unknown parameters for each pair (p, q) in Step 1 of the above Algorithm 3.1 are obtained by minimizing the first term of $L_{t,i}(p, q)$ in (3.3), which makes the estimates of orders not allowed to be smaller than the true orders. Moreover, the penalty term (i.e., $(p+q)a_t$) makes the estimates of orders no greater than the true orders. According to these intuitive explanations, we can obtain the convergence results of the estimates of orders by properly choosing a_t.

Remark 3.5. In Algorithm 3.1, if the estimates $(p_{t,i}, q_{t,i})$ for the system order (p_0, q_0) are obtained by the following step,

$$p_{t,i} = \arg\min_{1\le p\le p^*} L_{t,i}(p, q^*), \quad q_{t,i} = \arg\min_{1\le q\le q^*} L_{t,i}(p^*, q), \tag{3.7}$$

then the sequence $(p_{t,i}, q_{t,i})$ can also converge to the true order (p_0, q_0) by a similar argument as that used in the following Theorem 3.4. At each time instant t, we need to run $p^* \times q^*$ instances to find the minimum of the function $L_{t,i}(p, q)$ in (3.6), while in (3.7) we just need to run at most $p^* + q^*$ instances.

In the following, we will analyze the convergence of the estimation for system orders and parameters obtained in Algorithm 3.1. To this end, we first introduce some preliminary lemmas.

LEMMA 3.1 ([36]). *In Algorithm* 2.1*, for any fixed* p, q*, and* $t \ge 1$*, we have*

$$\lambda_{\max}\{\boldsymbol{d}_t(p,q)\boldsymbol{\Phi}_t^T(p,q)\boldsymbol{P}_t(p,q)\boldsymbol{\Phi}_t(p,q)\} \le \frac{det(\boldsymbol{P}_{t+1}^{-1}(p,q)) - det(\boldsymbol{P}_t^{-1}(p,q))}{det(\boldsymbol{P}_{t+1}^{-1}(p,q))} \le 1.$$

LEMMA 3.2 ([36]). *Under Assumptions* 3.1 *and* 3.2*, we have for* $p \ge p_0$ *and* $q \ge q_0$*,*

$$\sum_{i=1}^{n} \widetilde{\boldsymbol{\theta}}_{t,i}^T(p,q)\boldsymbol{P}_{t,i}^{-1}(p,q)\widetilde{\boldsymbol{\theta}}_{t,i}(p,q) = O(\log r_t(p,q)),$$

where $r_t(p,q)$ *is defined in Assumption* 3.3*.*

How to deal with the effect of the noises is a crucial step for the convergence analysis of Algorithm 3.1, and the following lemma provides an upper bound for the cumulative summation of the noises.

LEMMA 3.3. *Under Assumptions* 3.1 *and* 3.2*, we have for any fixed* p, q*,*

$$\boldsymbol{S}_{t+1,i}^T(p,q)\boldsymbol{P}_{t+1,i}(p,q)\boldsymbol{S}_{t+1,i}(p,q) = O(\log r_t(p,q)),$$

where $\boldsymbol{S}_{t+1,i}(p,q) = \sum_{j=1}^{n}\sum_{k=0}^{t} a_{ij}^{(t+1-k)}\boldsymbol{\varphi}_{k,j}(p,q)w_{k+1,j}$*, and* $a_{ij}^{(t+1-k)}$ *is the* i*th row,* j*th column entry of the weight matrix* $\mathcal{A}^{t+1-k}$*.*

Proof. For the convenience of expression, we use $\boldsymbol{S}_{t,i}, \boldsymbol{P}_k, \boldsymbol{\Phi}_k$, and $\boldsymbol{d}_k$ to denote $\boldsymbol{S}_{t,i}(p,q)$, $\boldsymbol{P}_k(p,q)$, $\boldsymbol{\Phi}_k(p,q)$, and $\boldsymbol{d}_k(p,q)$.

Set $\boldsymbol{S}_0 = 0$ and $\boldsymbol{S}_t = col\{\boldsymbol{S}_{t,1}, ..., \boldsymbol{S}_{t,n}\}$. Then we have

$$\boldsymbol{S}_{k+1} = \sum_{l=0}^{k} \mathscr{A}^{k+1-l}\boldsymbol{\Phi}_l\boldsymbol{W}_{l+1} = \mathscr{A}(\boldsymbol{S}_k + \boldsymbol{\Phi}_k\boldsymbol{W}_{k+1}).$$

By (2.8) and Lemma 4.2 in [36], we have

$$\begin{aligned}
&\boldsymbol{S}_{k+1}^T\boldsymbol{P}_{k+1}\boldsymbol{S}_{k+1} = (\boldsymbol{S}_k^T + \boldsymbol{W}_{k+1}^T\boldsymbol{\Phi}_k^T)\mathscr{A}\boldsymbol{P}_{k+1}\mathscr{A}(\boldsymbol{S}_k + \boldsymbol{\Phi}_k\boldsymbol{W}_{k+1})\\
&\le (\boldsymbol{S}_k^T + \boldsymbol{W}_{k+1}^T\boldsymbol{\Phi}_k^T)\bar{\boldsymbol{P}}_{k+1}(\boldsymbol{S}_k + \boldsymbol{\Phi}_k\boldsymbol{W}_{k+1})\\
&= (\boldsymbol{S}_k^T + \boldsymbol{W}_{k+1}^T\boldsymbol{\Phi}_k^T)(\boldsymbol{P}_k - \boldsymbol{P}_k\boldsymbol{\Phi}_k\boldsymbol{d}_k\boldsymbol{\Phi}_k^T\boldsymbol{P}_k)(\boldsymbol{S}_k + \boldsymbol{\Phi}_k\boldsymbol{W}_{k+1})\\
&= \boldsymbol{S}_k^T\boldsymbol{P}_k\boldsymbol{S}_k + 2\boldsymbol{W}_{k+1}^T\boldsymbol{\Phi}_k^T\boldsymbol{P}_k\boldsymbol{S}_k + \boldsymbol{W}_{k+1}^T\boldsymbol{\Phi}_k^T\boldsymbol{P}_k\boldsymbol{\Phi}_k\boldsymbol{W}_{k+1} - \boldsymbol{S}_k^T\boldsymbol{P}_k\boldsymbol{\Phi}_k\boldsymbol{d}_k\boldsymbol{\Phi}_k^T\boldsymbol{P}_k\boldsymbol{S}_k\\
&\quad - 2\boldsymbol{S}_k^T\boldsymbol{P}_k\boldsymbol{\Phi}_k\boldsymbol{d}_k\boldsymbol{\Phi}_k^T\boldsymbol{P}_k\boldsymbol{\Phi}_k\boldsymbol{W}_{k+1} - \boldsymbol{W}_{k+1}^T\boldsymbol{\Phi}_k^T\boldsymbol{P}_k\boldsymbol{\Phi}_k\boldsymbol{d}_k\boldsymbol{\Phi}_k^T\boldsymbol{P}_k\boldsymbol{\Phi}_k\boldsymbol{W}_{k+1}.
\end{aligned} \tag{3.8}$$

Moreover, by the definition of $\boldsymbol{d}_k$ and (2.14), we have

$$\boldsymbol{d}_k \boldsymbol{\Phi}_k^T \boldsymbol{P}_k \boldsymbol{\Phi}_k = \boldsymbol{I} - \boldsymbol{d}_k. \tag{3.9}$$

By (2.8) and (3.9), we derive that

$$\begin{aligned} \bar{\boldsymbol{P}}_{k+1}\boldsymbol{\Phi}_k &= \boldsymbol{P}_k\boldsymbol{\Phi}_k - \boldsymbol{P}_k\boldsymbol{\Phi}_k\boldsymbol{d}_k\boldsymbol{\Phi}_k^T\boldsymbol{P}_k\boldsymbol{\Phi}_k \\ &= \boldsymbol{P}_k\boldsymbol{\Phi}_k - \boldsymbol{P}_k\boldsymbol{\Phi}_k(\boldsymbol{I}-\boldsymbol{d}_k) = \boldsymbol{P}_k\boldsymbol{\Phi}_k\boldsymbol{d}_k. \end{aligned} \tag{3.10}$$

Substituting (3.9) into (3.8), we have by (3.10)

$$\begin{aligned} \boldsymbol{S}_{k+1}^T\boldsymbol{P}_{k+1}\boldsymbol{S}_{k+1} \le\ & \boldsymbol{S}_k^T\boldsymbol{P}_k\boldsymbol{S}_k + 2\boldsymbol{S}_k^T\boldsymbol{P}_k\boldsymbol{\Phi}_k\boldsymbol{d}_k\boldsymbol{W}_{k+1} - \boldsymbol{S}_k^T\boldsymbol{P}_k\boldsymbol{\Phi}_k\boldsymbol{d}_k\boldsymbol{\Phi}_k^T\boldsymbol{P}_k\boldsymbol{S}_k \\ & + \boldsymbol{W}_{k+1}^T\boldsymbol{\Phi}_k^T\boldsymbol{P}_k\boldsymbol{\Phi}_k\boldsymbol{d}_k\boldsymbol{W}_{k+1} \\ =\ & \boldsymbol{S}_k^T\boldsymbol{P}_k\boldsymbol{S}_k + 2\boldsymbol{S}_k^T\bar{\boldsymbol{P}}_{k+1}\boldsymbol{\Phi}_k\boldsymbol{W}_{k+1} - \boldsymbol{S}_k^T\bar{\boldsymbol{P}}_{k+1}\boldsymbol{\Phi}_k\boldsymbol{d}_k^{-1}\boldsymbol{\Phi}_k^T\bar{\boldsymbol{P}}_{k+1}\boldsymbol{S}_k \\ & + \boldsymbol{W}_{k+1}^T\boldsymbol{\Phi}_k^T\boldsymbol{P}_k\boldsymbol{\Phi}_k\boldsymbol{d}_k\boldsymbol{W}_{k+1} \\ \le\ & \boldsymbol{S}_k^T\boldsymbol{P}_k\boldsymbol{S}_k + 2\boldsymbol{S}_k^T\bar{\boldsymbol{P}}_{k+1}\boldsymbol{\Phi}_k\boldsymbol{W}_{k+1} - \boldsymbol{S}_k^T\bar{\boldsymbol{P}}_{k+1}\boldsymbol{\Phi}_k\boldsymbol{\Phi}_k^T\bar{\boldsymbol{P}}_{k+1}\boldsymbol{S}_k \\ & + \boldsymbol{W}_{k+1}^T\boldsymbol{\Phi}_k^T\boldsymbol{P}_k\boldsymbol{\Phi}_k\boldsymbol{d}_k\boldsymbol{W}_{k+1}. \end{aligned}$$

By the summation of both sides of the above inequality, we have

$$\begin{aligned} & \boldsymbol{S}_{t+1}^T\boldsymbol{P}_{t+1}\boldsymbol{S}_{t+1} + \sum_{k=0}^{t}\|\boldsymbol{S}_k^T\bar{\boldsymbol{P}}_{k+1}\boldsymbol{\Phi}_k\|^2 \\ & \le 2\sum_{k=0}^{t}\boldsymbol{S}_k^T\bar{\boldsymbol{P}}_{k+1}\boldsymbol{\Phi}_k\boldsymbol{W}_{k+1} + \sum_{k=0}^{t}\boldsymbol{W}_{k+1}^T\boldsymbol{\Phi}_k^T\boldsymbol{P}_k\boldsymbol{\Phi}_k\boldsymbol{d}_k\boldsymbol{W}_{k+1}. \end{aligned} \tag{3.11}$$

Next, we estimate the two terms on the right-hand side of (3.11) separately. By Assumption 3.2 and the martingale estimation theorem (see, e.g., [41]), we can get the following inequality:

$$\sum_{k=0}^{t}\boldsymbol{S}_k^T\bar{\boldsymbol{P}}_{k+1}\boldsymbol{\Phi}_k\boldsymbol{W}_{k+1} = O(1) + o\left(\sum_{k=0}^{t}\|\boldsymbol{S}_k^T\bar{\boldsymbol{P}}_{k+1}\boldsymbol{\Phi}_k\|^2\right). \tag{3.12}$$

Then by the proof of Lemma 4.4 in [36], we obtain

$$\sum_{k=0}^{t}\boldsymbol{W}_{k+1}^T\boldsymbol{\Phi}_k^T\boldsymbol{P}_k\boldsymbol{\Phi}_k\boldsymbol{d}_k\boldsymbol{W}_{k+1} = \sum_{k=0}^{t}\boldsymbol{W}_{k+1}^T\boldsymbol{d}_k\boldsymbol{\Phi}_k^T\boldsymbol{P}_k\boldsymbol{\Phi}_k\boldsymbol{W}_{k+1} = O(\log r_t). \tag{3.13}$$

Substituting (3.12) and (3.13) into (3.11) yields

$$\boldsymbol{S}_{t+1}^T\boldsymbol{P}_{t+1}\boldsymbol{S}_{t+1} = O(\log r_t),$$

which completes the proof of the lemma. ∎

Remark 3.6. If Assumption 3.2 is relaxed to the following weaker noise condition

$$\sup_{t\ge 0} E[|w_{t+1,i}|^2|\mathscr{F}_t] < \infty \text{ a.s.}, \tag{3.14}$$

then under Assumption 3.1 similar results as those of Lemmas 3.2 and 3.3 can also be obtained, save that the term "$\log r_t(p,q)$" in Lemmas 3.2 and 3.3 is replaced by "$\log r_t(p,q)\ (\log\log r_t(p,q))^\tau$ (for some $\tau > 1$)."

Remark 3.7. For the directed communication graph case, if the graph is strongly connected and balanced (i.e., both the inflow and outflow of each sensor are equal to 1), we can obtain a similar result as that of the above lemma by just replacing "$\mathscr{A}\boldsymbol{P}_{k+1}\mathscr{A}$" with "$\mathscr{A}^T\boldsymbol{P}_{k+1}\mathscr{A}$" in (3.8).

Now, we present the main results concerning the convergence of the order estimates obtained by Algorithm 3.1.

THEOREM 3.4. *Under Assumptions* 3.1–3.3, *the order estimate sequence* $(p_{t,i}, q_{t,i})$ *given by (*3.6*) converges to the true order* (p_0, q_0) *a.s., i.e.,*

$$(p_{t,i}, q_{t,i}) \xrightarrow[t\to\infty]{} (p_0, q_0) \quad \text{a.s.} \quad \textit{for} \quad i \in \{1, \dots, n\}.$$

Proof. For $i \in \{1, \dots, n\}$, we need to show that the sequence $(p_{t,i}, q_{t,i})$ has only one limit point (p_0, q_0). Let $(p_i', q_i') \in M$ be a limit point of $(p_{t,i}, q_{t,i})$, i.e., there is a subsequence $\{t_k\}$ such that

$$(p_{t_k,i}, q_{t_k,i}) \xrightarrow[k\to\infty]{} (p_i', q_i'). \tag{3.15}$$

In order to prove $(p_{t,i}, q_{t,i}) \xrightarrow[t\to\infty]{} (p_0, q_0)$, we just need to show the impossibility of the following two situations:

(i) $p_i' \geq p_0$, $q_i' \geq q_0$, and $p_i' + q_i' > p_0 + q_0$;

(ii) $p_i' < p_0$ or $q_i' < q_0$.

Noting that both $p_{t_k,i}$ and $q_{t_k,i}$ are integers, by (3.15) we have $(p_{t_k,i}, q_{t_k,i}) \equiv (p_i', q_i')$ for sufficiently large k. We first show that the situation (i) will not happen by reduction to absurdity.

Suppose that (i) holds. By (2.5) and (3.5), we see that $\sigma_{t_k,i}(p_i', q_i', \boldsymbol{\theta}_{t_k,i}(p_i', q_i'))$ can be calculated by the following equation:

$$\begin{aligned}
&\sigma_{t_k,i}(p_i', q_i', \boldsymbol{\theta}_{t_k,i}(p_i', q_i')) \\
&= \sum_{j=1}^{n} \sum_{l=0}^{t_k-1} a_{ij}^{(t_k-l)} [y_{l+1,j} - \boldsymbol{\theta}_{t_k,i}^T(p_i', q_i')\boldsymbol{\varphi}_{l,j}(p_i', q_i')]^2 \\
&= \sum_{j=1}^{n} \sum_{l=0}^{t_k-1} a_{ij}^{(t_k-l)} [\widetilde{\boldsymbol{\theta}}_{t_k,i}^T(p_i', q_i')\boldsymbol{\varphi}_{l,j}(p_i', q_i') + w_{l+1,j}]^2 \\
&= \widetilde{\boldsymbol{\theta}}_{t_k,i}^T(p_i', q_i') \left(\sum_{j=1}^{n} \sum_{l=0}^{t_k-1} a_{ij}^{(t_k-l)} \boldsymbol{\varphi}_{l,j}(p_i', q_i')\boldsymbol{\varphi}_{l,j}^T(p_i', q_i') \right) \widetilde{\boldsymbol{\theta}}_{t_k,i}(p_i', q_i') \\
&+ 2\widetilde{\boldsymbol{\theta}}_{t_k,i}^T(p_i', q_i') \left(\sum_{j=1}^{n} \sum_{l=0}^{t_k-1} a_{ij}^{(t_k-l)} \boldsymbol{\varphi}_{l,j}(p_i', q_i') w_{l+1,j} \right) + \sum_{j=1}^{n} \sum_{l=0}^{t_k-1} a_{ij}^{(t_k-l)} w_{l+1,j}^2.
\end{aligned} \tag{3.16}$$

By Lemmas 3.2 and 3.3, we have the following relationship:

$$\begin{aligned}
&\left| \widetilde{\boldsymbol{\theta}}_{t_k,i}^T(p_i', q_i') \left(\sum_{j=1}^{n} \sum_{l=0}^{t_k-1} a_{ij}^{(t_k-l)} \boldsymbol{\varphi}_{l,j}(p_i', q_i') w_{l+1,j} \right) \right| \\
&\leq \left\| \widetilde{\boldsymbol{\theta}}_{t_k,i}^T(p_i', q_i') \boldsymbol{P}_{t_k,i}^{-\frac{1}{2}}(p_i', q_i') \right\| \left\| \boldsymbol{P}_{t_k,i}^{\frac{1}{2}}(p_i', q_i') \left(\sum_{j=1}^{n} \sum_{l=0}^{t_k-1} a_{ij}^{(t_k-l)} \boldsymbol{\varphi}_{l,j}(p_i', q_i') w_{l+1,j} \right) \right\| \\
&= O(\log(r_{t_k}(p_i', q_i'))) = O(\log(r_{t_k}(p^*, q^*))).
\end{aligned} \tag{3.17}$$

By (2.10) and (2.17), we have for any p and q

$$\boldsymbol{P}_{t_k,i}^{-1}(p,q)=\sum_{j=1}^{n}\sum_{l=0}^{t_k-1}a_{ij}^{(t_k-l)}\boldsymbol{\varphi}_{l,j}(p,q)\boldsymbol{\varphi}_{l,j}^T(p,q)+\sum_{j=1}^{n}a_{ij}^{(t_k)}\boldsymbol{P}_{0,j}^{-1}(p,q). \tag{3.18}$$

By this equation and Lemma 3.2, we can easily obtain that

$$\begin{aligned}&\widetilde{\boldsymbol{\theta}}_{t_k,i}^T(p_i',q_i')\sum_{j=1}^{n}\sum_{l=0}^{t_k-1}a_{ij}^{(t_k-l)}\boldsymbol{\varphi}_{l,j}(p,q)\boldsymbol{\varphi}_{l,j}^T(p_i',q_i')\widetilde{\boldsymbol{\theta}}_{t_k,i}(p_i',q_i')\\ &\quad=O(\log r_{t_k}(p_i',q_i'))=O(\log(r_{t_k}(p^*,q^*))).\end{aligned} \tag{3.19}$$

Substituting (3.17) and (3.19) into (3.16), we see that there exists a positive constant C_1 satisfying

$$\sigma_{t_k,i}(p_i',q_i',\boldsymbol{\theta}_{t_k,i}(p_i',q_i'))-\sum_{j=1}^{n}\sum_{l=0}^{t_k-1}a_{ij}^{(t_k-l)}w_{l+1,j}^2\geq -C_1\log(r_{t_k}(p^*,q^*)). \tag{3.20}$$

Now, we will consider $\sigma_{t_k,i}(p_0,q_0,\boldsymbol{\theta}_{t_k,i}(p_0,q_0))$. By Lemma 2.2, we have for $p\geq p_0$ and $q\geq q_0$,

$$\begin{aligned}&\sum_{j=1}^{n}\sum_{l=0}^{t_k-1}a_{ij}^{(t_k-l)}\boldsymbol{\varphi}_{l,j}(p,q)w_{l+1,j}\\ &\quad=\sum_{j=1}^{n}a_{ij}^{(t_k)}\boldsymbol{P}_{0,j}^{-1}(p,q)\widetilde{\boldsymbol{\theta}}_{0,j}(p,q)-\boldsymbol{P}_{t_k,i}^{-1}(p,q)\widetilde{\boldsymbol{\theta}}_{t_k,i}(p,q).\end{aligned} \tag{3.21}$$

In a similar way to that used in (3.16), we obtain

$$\begin{aligned}&\sigma_{t_k,i}(p_0,q_0,\boldsymbol{\theta}_{t_k,i}(p_0,q_0))-\sum_{j=1}^{n}\sum_{l=0}^{t_k-1}a_{ij}^{(t_k-l)}w_{l+1,j}^2\\ &=\widetilde{\boldsymbol{\theta}}_{t_k,i}^T(p_0,q_0)\left(\sum_{j=1}^{n}\sum_{l=0}^{t_k-1}a_{ij}^{(t_k-l)}\boldsymbol{\varphi}_{l,j}(p_0,q_0)\boldsymbol{\varphi}_{l,j}^T(p_0,q_0)\right)\widetilde{\boldsymbol{\theta}}_{t_k,i}(p_0,q_0)\\ &\quad+2\widetilde{\boldsymbol{\theta}}_{t_k,i}^T(p_0,q_0)\left(\sum_{j=1}^{n}\sum_{l=0}^{t_k-1}a_{ij}^{(t_k-l)}\boldsymbol{\varphi}_{l,j}(p_0,q_0)w_{l+1,j}\right).\end{aligned}$$

Combining this with (3.18) and (3.21) yieldsCombining this with

(3.22)

$$\begin{aligned}&\sigma_{t_k,i}(p_0,q_0,\boldsymbol{\theta}_{t_k,i}(p_0,q_0))-\sum_{j=1}^{n}\sum_{l=0}^{t_k-1}a_{ij}^{(t_k-l)}w_{l+1,j}^2\\ &\quad=\widetilde{\boldsymbol{\theta}}_{t_k,i}^T(p_0,q_0)\boldsymbol{P}_{t_k,i}^{-1}(p_0,q_0)\widetilde{\boldsymbol{\theta}}_{t_k,i}(p_0,q_0)\\ &\qquad-\widetilde{\boldsymbol{\theta}}_{t_k,i}^T(p_0,q_0)\left(\sum_{j=1}^{n}a_{ij}^{(t_k)}\boldsymbol{P}_{0,j}^{-1}(p_0,q_0)\right)\widetilde{\boldsymbol{\theta}}_{t_k,i}(p_0,q_0)\end{aligned}$$

$$
\begin{aligned}
&+ 2\widetilde{\boldsymbol{\theta}}^T_{t_k,i}(p_0,q_0)\left(\sum_{j=1}^n a_{ij}^{(t_k)}\boldsymbol{P}_{0,j}^{-1}(p_0,q_0)\widetilde{\boldsymbol{\theta}}_{0,j}(p_0,q_0) - \boldsymbol{P}_{t_k,i}^{-1}(p_0,q_0)\widetilde{\boldsymbol{\theta}}_{t_k,i}(p_0,q_0)\right)\\
&\le -\widetilde{\boldsymbol{\theta}}^T_{t_k,i}(p_0,q_0)\left(\sum_{j=1}^n a_{ij}^{(t_k)}\boldsymbol{P}_{0,j}^{-1}(p_0,q_0)\right)\widetilde{\boldsymbol{\theta}}_{t_k,i}(p_0,q_0)\\
&+\ 2\widetilde{\boldsymbol{\theta}}^T_{t_k,i}(p_0,q_0)\left(\sum_{j=1}^n a_{ij}^{(t_k)}\boldsymbol{P}_{0,j}^{-1}(p_0,q_0)\widetilde{\boldsymbol{\theta}}_{0,j}(p_0,q_0)\right)\\
&\le \left(\sum_{j=1}^n a_{ij}^{(t_k)}\widetilde{\boldsymbol{\theta}}^T_{0,j}(p_0,q_0)\boldsymbol{P}_{0,j}^{-1}(p_0,q_0)\widetilde{\boldsymbol{\theta}}_{0,j}(p_0,q_0)\right) = O(1),
\end{aligned}
$$

where the last inequality is obtained by

$$
2\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{y} \le \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + \boldsymbol{y}^T\boldsymbol{A}\boldsymbol{y} \quad for \quad \boldsymbol{A}\ge 0. \tag{3.23}
$$

From (3.20) and (3.22), we have

$$
\sigma_{t_k,i}(p_i',q_i',\boldsymbol{\theta}_{t_k,i}(p_i',q_i')) - \sigma_{t_k,i}(p_0,q_0,\boldsymbol{\theta}_{t_k,i}(p_0,q_0)) \ge -C_1\log r_{t_k}(p^*,q^*) - C_2,
$$

where C_2 is a positive constant. Note that $(p_{t_k,i},q_{t_k,i}) = \arg\min_{p,q\in M} L_{t_k,i}(p,q)$. By Assumption 3.3, we have

$$
\begin{aligned}
0 &\ge L_{t_k,i}(p_{t_k,i},q_{t_k,i}) - L_{t_k,i}(p_0,q_0) = L_{t_k,i}(p_i',q_i') - L_{t_k,i}(p_0,q_0)\\
&= \sigma_{t_k,i}(p_i',q_i',\boldsymbol{\theta}_{t_k,i}(p_i',q_i')) - \sigma_{t_k,i}(p_0,q_0,\boldsymbol{\theta}_{t_k,i}(p_0,q_0)) + (p_i'+q_i'-p_0-q_0)a_{t_k}\\
&\ge -C_1\log r_{t_k}(p^*,q^*) - C_2 + (p_i'+q_i'-p_0-q_0)a_{t_k}\\
&= a_{t_k}\left(\frac{-C_1\log r_{t_k}(p^*,q^*)}{a_{t_k}} + (p_i'+q_i'-p_0-q_0)\right) - C_2 \to \infty \text{as } k\to\infty,
\end{aligned}
$$

which leads to the contradiction. Thus, situation (i) will not happen.

In the following, we will show the impossibility of situation (ii) by reduction to absurdity. Suppose that (ii) holds, i.e., $p_i' < p_0$ or $q_i' < q_0$. In order to analyze the properties of the estimate error, we introduce the following (s_i+v_i)-dimensional vector with $s_i = \max\{p_0,p_i'\}$, $v_i = \max\{q_0,q_i'\}$,

$$
\boldsymbol{\theta}_{t_k,i}(s_i,v_i) = [b^i_{1,t_k},\dots,b^i_{s_i,t_k},c^i_{1,t_k},\dots,c^i_{v_i,t_k}]^T.
$$

If $p_i' < p_0$, then $b^i_{m,t_k} \triangleq 0$ for $p_i' < m \le p_0$; and if $q_i' < q_0$, then $c^i_{m,t_k} \triangleq 0$ for $q_i' < m \le q_0$.

Denote $\widetilde{\boldsymbol{\theta}}_{t_k,i}(s_i,v_i) = \boldsymbol{\theta}(s_i,v_i) - \boldsymbol{\theta}_{t_k,i}(s_i,v_i)$. It is clear that

$$
\|\widetilde{\boldsymbol{\theta}}_{t_k,i}(s_i,v_i)\|^2 \ge \min\{|b_{p_0}|^2,|c_{q_0}|^2\} \triangleq \alpha_0 > 0. \tag{3.24}
$$

Then by (2.6), we have

$$
\begin{aligned}
&\sigma_{t_k,i}(p_i',q_i',\boldsymbol{\theta}_{t_k,i}(p_i',q_i'))\\
&= \sum_{j=1}^n\sum_{l=0}^{t_k-1} a_{ij}^{(t_k-l)}[\boldsymbol{\theta}^T(p_0,q_0)\boldsymbol{\varphi}_{l,j}(p_0,q_0) - \boldsymbol{\theta}^T_{t_k,i}(p_i',q_i')\boldsymbol{\varphi}_{l,j}(p_i',q_i') + w_{l+1,j}]^2.
\end{aligned}
$$

Hence combining this equation and the definition $\widetilde{\boldsymbol{\theta}}_{t_k,i}(s_i, v_i)$, we obtain

$$\begin{aligned}
&\sigma_{t_k,i}(p'_i, q'_i, \boldsymbol{\theta}_{t_k,i}(p'_i, q'_i)) - \sum_{j=1}^{n}\sum_{l=0}^{t_k-1} a_{ij}^{(t_k-l)} w_{l+1,j}^2 \\
&= \widetilde{\boldsymbol{\theta}}_{t_k,i}^T(s_i, v_i)\boldsymbol{P}_{t_k,i}^{-1}(s_i, v_i)\widetilde{\boldsymbol{\theta}}_{t_k,i}(s_i, v_i) - \widetilde{\boldsymbol{\theta}}_{t_k,i}^T(s_i, v_i)\left(\sum_{j=1}^{n} a_{ij}^{(t)}\boldsymbol{P}_{0,j}^{-1}(s_i, v_i)\right)\widetilde{\boldsymbol{\theta}}_{t_k,i}(s_i, v_i) \\
&\quad + 2\widetilde{\boldsymbol{\theta}}_{t_k,i}^T(s_i, v_i)\left(\sum_{j=1}^{n}\sum_{l=0}^{t_k-1} a_{ij}^{(t_k-l)}\boldsymbol{\varphi}_{l,j}(s_i, v_i) w_{l+1,j}\right) \\
&\triangleq \widetilde{\boldsymbol{\theta}}_{t_k,i}^T(s_i, v_i)\boldsymbol{P}_{t_k,i}^{-1}(s_i, v_i)\widetilde{\boldsymbol{\theta}}_{t_k,i}(s_i, v_i) - M_1 + M_2. \qquad (3.25)
\end{aligned}$$

By (3.18) and Remark 3.1, we have for any p and q

$$\lambda_{\min}(\boldsymbol{P}_{t_k,i}^{-1}(p, q)) \geq a_{\min}\lambda_{\min}^{p,q}(t_k), \qquad (3.26)$$

where $a_{\min} = \min_{i,j\in\mathcal{V}} a_{ij}^{(D_{\mathcal{G}})} > 0$. Hence, by (3.26) and Lemma 3.2, we have for $p \geq p_0$ and $q \geq q_0$,

$$\sum_{i=1}^{n} \|\widetilde{\boldsymbol{\theta}}_{t+1,i}(p, q)\|^2 = O\left(\frac{\log r_t(p, q)}{\lambda_{\min}^{p,q}(t)}\right). \qquad (3.27)$$

When $p'_i < p_0$ (as does the case $q'_i < q_0$), we can use (3.27) in the first p'_i components of $\widetilde{\boldsymbol{\theta}}_{t_k,i}(s_i, v_i)$. Then by (2.1) and Assumption 3.3, we obtain $\|\widetilde{\boldsymbol{\theta}}_{t_k,i}(s_i, v_i)\| = O(1)$, hence we have

$$M_1 \leq \lambda_{\max}\left(\sum_{j=1}^{n} a_{ij}^{(t)}\boldsymbol{P}_{0,j}^{-1}(s_i, v_i)\right)\|\widetilde{\boldsymbol{\theta}}_{t_k,i}(s_i, v_i)\|^2 = O(1). \qquad (3.28)$$

In the following, we will analyze M_2. Similarly to the analysis of (3.17), by Lemma 3.3, we have

$$\begin{aligned}
|M_2| &\leq \left\|\widetilde{\boldsymbol{\theta}}_{t_k,i}^T(s_i, v_i)\boldsymbol{P}_{t_k,i}^{-\frac{1}{2}}(s_i, v_i)\right\| \cdot \left\|\boldsymbol{P}_{t_k,i}^{\frac{1}{2}}(s_i, v_i)\left(\sum_{j=1}^{n}\sum_{l=0}^{t_k-1} a_{ij}^{(t_k-l)}\boldsymbol{\varphi}_{l,j}(s_i, v_i) w_{l+1,j}\right)\right\| \\
&= O\left\{\left\|\widetilde{\boldsymbol{\theta}}_{t_k,i}^T(s_i, v_i)\boldsymbol{P}_{t_k,i}^{-\frac{1}{2}}(s_i, v_i)\right\| \cdot \log^{\frac{1}{2}}(r_{t_k}(s_i, v_i))\right\} \\
&= O\left\{\left\|\widetilde{\boldsymbol{\theta}}_{t_k,i}^T(s_i, v_i)\boldsymbol{P}_{t_k,i}^{-\frac{1}{2}}(s_i, v_i)\right\| \cdot \log^{\frac{1}{2}}(r_{t_k}(p^*, q^*))\right\}. \qquad (3.29)
\end{aligned}$$

Therefore, by (3.25)–(3.29), we see that there exist two positive constants C_3 and C_4 such that

$$\begin{aligned}
&\sigma_{t_k,i}(p'_i, q'_i, \boldsymbol{\theta}_{t_k,i}(p'_i, q'_i)) - \sum_{j=1}^{n}\sum_{l=0}^{t_k-1} a_{ij}^{(t_k-l)} w_{l+1,j}^2 \\
&\geq \widetilde{\boldsymbol{\theta}}_{t_k,i}^T(s_i, v_i)\boldsymbol{P}_{t_k,i}^{-1}(s_i, v_i)\widetilde{\boldsymbol{\theta}}_{t_k,i}(s_i, v_i) - C_3 \\
&\quad - C_4\left\|\widetilde{\boldsymbol{\theta}}_{t_k,i}^T(s_i, v_i)\boldsymbol{P}_{t_k,i}^{-\frac{1}{2}}(s_i, v_i)\right\| \cdot \log^{\frac{1}{2}}(r_{t_k}(p^*, q^*)). \qquad (3.30)
\end{aligned}$$

By Assumption 3.3, (3.24), and (3.26), we have

$$\left\|\widetilde{\boldsymbol{\theta}}_{t_k,i}^T(s_i,v_i)\boldsymbol{P}_{t_k,i}^{-\frac{1}{2}}(s_i,v_i)\right\|\cdot\log^{\frac{1}{2}}(r_{t_k}(p^*,q^*))=o(\widetilde{\boldsymbol{\theta}}_{t_k,i}^T(s_i,v_i)\boldsymbol{P}_{t_k,i}^{-1}(s_i,v_i)\widetilde{\boldsymbol{\theta}}_{t_k,i}(s_i,v_i)).$$

Furthermore, by (3.24) and (3.26), we have

$$\begin{aligned}&\sigma_{t_k,i}(p_i',q_i',\boldsymbol{\theta}_{t_k,i}(p_i',q_i'))-\sum_{j=1}^{n}\sum_{l=0}^{t_k-1}a_{ij}^{(t_k-l)}w_{l+1,j}^2\\&=a_{\min}\alpha_0\lambda_{\min}^{s_i,v_i}(t_k)(1+o(1))-C_3\\&\geq\frac{a_{\min}\alpha_0\min\{\lambda_{\min}^{p_0,q^*}(t_k),\lambda_{\min}^{p^*,q_0}(t_k)\}}{2}-C_3,\end{aligned}\tag{3.31}$$

where (2.1) is used in the last inequality.

By (3.3), (3.22), (3.31), and Assumption 3.3, for large k and some positive constant C_5, we have the following inequality for $i\in\{1,2,\cdots,n\}$,

$$\begin{aligned}0&\geq L_{t_k,i}(p_{t_k,i},q_{t_k,i})-L_{t_k,i}(p_0,q_0)=L_{t_k,i}(p_i',q_i')-L_{t_k,i}(p_0,q_0)\\&=\sigma_{t_k,i}(p_i',q_i',\boldsymbol{\theta}_{t_k,i}(p_i',q_i'))-\sigma_{t_k,i}(p_0,q_0,\boldsymbol{\theta}_{t_k,i}(p_0,q_0))+(p_i'+q_i'-p_0-q_0)a_{t_k}\\&\geq\frac{a_{\min}\alpha_0\min\{\lambda_{\min}^{p_0,q^*}(t_k),\lambda_{\min}^{p^*,q_0}(t_k)\}}{2}-C_5+(p_i'+q_i'-p_0-q_0)a_{t_k}\\&\geq\frac{a_{\min}\alpha_0\min\{\lambda_{\min}^{p_0,q^*}(t_k),\lambda_{\min}^{p^*,q_0}(t_k)\}}{2}\Big(1+\frac{2(p_i'+q_i'-p_0-q_0)a_{t_k}}{a_{\min}\alpha_0\min\{\lambda_{\min}^{p_0,q^*}(t_k),\lambda_{\min}^{p^*,q_0}(t_k)\}}\Big)-C_5\\&\geq\frac{a_{\min}\alpha_0\min\{\lambda_{\min}^{p_0,q^*}(t_k),\lambda_{\min}^{p^*,q_0}(t_k)\}}{4}-C_5\to\infty,\end{aligned}$$

which leads to a contradiction. The proof of the theorem is complete. ∎

Remark 3.8. Under Assumption 3.1 and the weaker noise condition (3.14), by Remark 3.6, we can verify that the result of Theorem 3.4 still holds by taking the sequence $\{a_t\}$ in Assumption 3.3 to satisfy the following conditions,

$$\frac{\log r_t(p^*,q^*)(\log\log r_t(p^*,q^*))^{\tau}}{a_t}\xrightarrow[t\to\infty]{}0,\quad\frac{a_t}{\lambda_{\min}^{p,q}(t)}\xrightarrow[t\to\infty]{}0\qquad\text{a.s.},$$

where $(p,q)\in M^*$.

Note that both the estimates $(p_{t,i},q_{t,i})$ and the true orders (p_0,q_0) are integers; from Theorem 3.4, we see that there exists a large enough T such that $p_{t,i}=p_0$ and $q_{t,i}=q_0$ for $t\geq T$. Thus, from (3.27) and Assumption 3.3, we have the following consistent estimation of the parameter vector $\boldsymbol{\theta}(p_0,q_0)$.

THEOREM 3.5. *Under the conditions of Theorem* 3.4 *for any* $i\in\{1,\cdots,n\}$, *we have*

$$\boldsymbol{\theta}_{t,i}(p_{t,i},q_{t,i})\xrightarrow[t\to\infty]{}\boldsymbol{\theta}(p_0,q_0)\quad\text{a.s.},$$

where $\boldsymbol{\theta}_{t,i}(p_{t,i},q_{t,i})$ *is obtained by Algorithm* 3.1.

4. Case II: The upper bounds of true orders are unknown. In this section, we consider a general case where the upper bounds of the system orders are unknown. We first give some assumptions on the system signals and the noise.

Assumption 4.1. For $i \in \{1, \cdots, n\}$, the noise sequence $\{w_{t,i}, \mathscr{F}_t\}$ is a martingale difference sequence satisfying

$$\sup_{t\geq 0} E[|w_{t+1,i}|^2|\mathscr{F}_t] < \infty, \quad \|w_{t,i}\| = O(\eta_i(t)) \ \text{ a.s.},$$

where $\mathscr{F}_t$ is defined in Assumption 3.2 and $\eta_i(t)$ is a positive, deterministic, nondecreasing function satisfying

$$\sup_t \eta_i(e^{t+1})/\eta_i(e^t) < \infty.$$

The term "$\|w_{t,i}\| = O(\eta_i(t))$" in Assumption 4.1 describes the growth rate of the noise, which means the double array martingale estimation theory (Lemma 4.1) can be used to deal with the cumulative effect of the noise. It can be easily verified that the commonly used bounded or white Gaussian noises can satisfy this assumption.

In order to simplify the analysis of the estimation error, we need to introduce an assumption on the input and output signals which implies that the system is not explosive. This assumption is commonly used in the stability analysis of the closed-loop feedback control systems for a single sensor case (see, e.g., [7, 8, 15]).

Assumption 4.2. There exists a positive constant b such that the input and output signals satisfy

$$\sum_{i=1}^{n}\sum_{k=0}^{t-1}(\|y_{k,i}\|^2 + \|u_{k,i}\|^2) = O(t^b) \ \text{ a.s.} \tag{4.1}$$

Similarly to Assumption 3.3 in section 3, we introduce the following cooperative excitation condition which can be considered as an extension of the excitation condition used in [15] for a single sensor to the distributed order estimation algorithm when the upper bounds of true orders are unknown. This condition can also reflect the joint effect of multiple sensors as illustrated in Remark 3.4.

Assumption 4.3. (cooperative excitation condition II). A sequence $\{\bar{a}_t\}$ of positive real numbers can be found such that $\bar{a}_t \xrightarrow[t\to\infty]{} \infty$ and

$$\frac{h_t \log t + [\eta(t)\log\log t]^2}{\bar{a}_t} \xrightarrow[t\to\infty]{} 0, \qquad \frac{\bar{a}_t}{\lambda^0_{\min}(t)} \xrightarrow[t\to\infty]{} 0, \tag{4.2}$$

hold a.s., where

$$\lambda^0_{\min}(t) = \lambda_{\min}\left\{\sum_{j=1}^{n}\boldsymbol{P}^{-1}_{0,j}(m_0, m_0) + \sum_{i=1}^{n}\sum_{k=0}^{t-D_{\mathcal{G}}}\boldsymbol{\varphi}^0_{k,i}(\boldsymbol{\varphi}^0_{k,i})^T\right\}$$

with $\eta(t) \triangleq (\sum_{i=1}^n \eta_i^2(t))^{\frac{1}{2}}$, $\boldsymbol{\varphi}^0_{t,i} = [y_{t,i}, \ldots, y_{t-m_0+1,i}, u_{t,i}, \ldots, u_{t-m_0+1,i},]^T$, $m_0 \triangleq \max\{p_0, q_0\}$, and the regression lag h_t is chosen as $h_t = O((\log t)^\alpha)(\alpha > 1)$, and $\log t = o(h_t)$.

We will now construct the algorithm to estimate both the system orders and parameters in a distributed way when the upper bounds of orders are unknown. For estimating the unknown orders (p_0, q_0), we introduce the following LIC $\bar{L}_{t,i}(p,q)$ for the sensor i,

$$\bar{L}_{t,i}(p,q) = \sigma_{t,i}(p,q,\boldsymbol{\theta}_{t,i}(p,q)) + (p+q)\bar{a}_t, \tag{4.3}$$

where $\sigma_{t,i}(p,q,\boldsymbol{\beta}(p,q))$ is recursively defined in (3.4).

By minimizing the LIC (4.3) and using Algorithm 2.1, we obtain the following distributed algorithm.

Algorithm 4.1.

For any given sensor $i \in \{1, \dots, n\}$, the distributed algorithm for the estimation of the system orders and the parameters is defined at the time instant $t \geq 1$ as follows.

Step 1: For any $0 \leq s \leq [\log t]$, based on $\{\boldsymbol{\varphi}_{k,j}(s,s), y_{k+1,j}\}_{k=1}^{t-1}$, $j \in N_i$, the estimate $\boldsymbol{\theta}_{t,i}(s,s)$ can obtained by using Algorithm 2.1, where the orders (p,q) are replaced by (s,s) $(0 \leq s \leq [\log t])$.

Step 2: (order estimation) With the estimates $\{\boldsymbol{\theta}_{t,i}(s,s)\}_{s=0}^{[\log t]}$ obtained by Step 1,
take $\hat{m}_{t,i}$ by minimizing $\bar{L}_{t,i}(s,s)$ for $0 \leq s \leq [\log t]$;
take $\hat{p}_{t,i}$ by minimizing $\bar{L}_{t,i}(p, \hat{m}_{t,i})$ for $0 \leq p \leq \hat{m}_{t,i}$;
take $\hat{q}_{t,i}$ by minimizing $\bar{L}_{t,i}(\hat{p}_{t,i}, q)$ for $0 \leq q \leq \hat{m}_{t,i}$.

Step 3: (parameter estimation) The estimate $\boldsymbol{\theta}_{t,i}(\hat{p}_{t,i}, \hat{q}_{t,i})$ for the unknown parameter vector $\boldsymbol{\theta}(p_0, q_0)$ is obtained by using Algorithm 2.1, where the orders (p,q) are replaced by the estimates $(\hat{p}_{t,i}, \hat{q}_{t,i})$ obtained by Step 2.

Output: $\hat{p}_{t,i}, \hat{q}_{t,i}$ and $\boldsymbol{\theta}_{t,i}(\hat{p}_{t,i}, \hat{q}_{t,i})$.

Remark 4.1. In Step 2 of the above Algorithm 4.1, we first estimate the maximum value m_0 of true orders whose upper bound is characterized by the function $\log t$. Then the true orders p_0, q_0 are obtained by searching among at most $2\hat{m}_{t,i}$ points at each time instant t.

In the following, we will provide the consistency analysis of Algorithm 4.1 when the upper bounds of orders are unknown, in which a crucial step is to prove that for any i, $\hat{m}_{t,i} \to m_0$ as $t \to \infty$. Then by the order estimation procedure in Algorithm 4.1, the convergence of the estimates for the system orders and parameters can be obtained by a similar analysis as those in section 3. To this end, we need to introduce the following double array martingale estimation lemma to deal with the noise effect in the form of $\max_{1 \leq m \leq h_t} \left\| \sum_{k=1}^{t} f_k(m) w_{k+1} \right\|$.

LEMMA 4.1 ([15]). *Let $\{v_t, \mathscr{F}_t\}$ be an s'-dimensional martingale difference sequence satisfying $\|v_t\| = o(\rho(t))$ a.s., $\sup_t E(\|v_{t+1}\|^2 | \mathscr{F}_t) < \infty$ a.s., where the properties of $\rho(t)$ are described as the same as $\eta_i(t)$ in Assumption* 4.1. *Assume that $f_t(m), t, m = 1, 2, \ldots,$ is an $\mathscr{F}_t$-measurable, $r' \times s'$-dimensional random matrix satisfying $\|f_t(m)\| \leq C < \infty$ a.s. for all t, m, and some deterministic constant C. Then for $h_t = O([\log t]^\alpha)$ $(\alpha > 1)$, the following property holds as $t \to \infty$:*

$$\max_{1 \leq m \leq h_t} \max_{1 \leq j \leq t} \left\| \sum_{k=1}^{j} f_k(m) v_{k+1} \right\| = O\left(\max_{1 \leq m \leq h_t} \sum_{k=1}^{t} \|f_k(m)\|^2 \right) + o(\rho(t) \log \log t) \quad \text{a.s.}$$

In order to simplify the expression of the following lemmas and theorems, we write (l) for (l,l) in $\boldsymbol{\theta}_{t,i}, \boldsymbol{\varphi}_{t,i}$ and $\boldsymbol{P}_{t,i}$ when $p = q = l$.

LEMMA 4.2. *Let $V_t(l) = \widetilde{\boldsymbol{\Theta}}_t^T(l) \boldsymbol{P}_t^{-1}(l) \widetilde{\boldsymbol{\Theta}}_t(l)$. Then under Assumptions* 3.1 *and* 4.1–4.3, *we have*

$$\max_{m_0 \leq l \leq h_t} V_{t+1}(l) = O(h_t \log t) + o(\eta^2(t) \log \log t),$$

where h_t and $\eta(t)$ are defined in Assumption 4.3.

Proof. By the proof of Lemma 4.4 in [36], we have for $l \geq m_0$

$$V_{t+1}(l) = O(1) + \sum_{k=0}^{t} \boldsymbol{W}_{k+1}^T \boldsymbol{d}_k(l) \boldsymbol{\Phi}_k^T(l) \boldsymbol{P}_k(l) \boldsymbol{\Phi}_k(l) \boldsymbol{W}_{k+1}. \tag{4.4}$$

By (3.18) and Lemma 3.1, we have

$$\begin{aligned} &\max_{1\leq l\leq h_t} \sum_{k=0}^{t} \lambda_{\max}\{\boldsymbol{d}_k(l)\boldsymbol{\Phi}_k^T(l)\boldsymbol{P}_k(l)\boldsymbol{\Phi}_k(l)\} \\ &\leq \max_{1\leq l\leq h_t} [\log det(\boldsymbol{P}_{t+1}^{-1}(l)) - \log det(\boldsymbol{P}_0^{-1}(l))] \\ &= O\left\{ \max_{1\leq l\leq h_t} \left(l \cdot \log \left(\lambda_{\max} \boldsymbol{P}_0^{-1}(l) + \sum_{j=1}^{n} \sum_{k=0}^{t} \|\boldsymbol{\varphi}_{k,j}(l)\|^2 \right) \right) \right\} = O(h_t \log t), \quad (4.5) \end{aligned}$$

where Assumption 4.2 is used in the last equation. By Assumptions 4.1, 3.1 and 4.1, we have

$$\begin{aligned} &\max_{1\leq l\leq h_t} \sum_{k=0}^{t} \lambda_{\max}\{\boldsymbol{d}_k(l)\boldsymbol{\Phi}_k^T(l)\boldsymbol{P}_k(l)\boldsymbol{\Phi}_k(l)\}(\|\boldsymbol{W}_{k+1}\|^2 - E(\|\boldsymbol{W}_{k+1}\|^2|\mathscr{F}_k)) \\ &= o(\eta^2(t)\log\log t) + O\left(\max_{1\leq l\leq h_t} \sum_{k=0}^{t} \lambda_{\max}\{\boldsymbol{d}_k(l)\boldsymbol{\Phi}_k^T(l)\boldsymbol{P}_k(l)\boldsymbol{\Phi}_k(l)\} \right). \end{aligned}$$

Hence by (4.5) and Assumption 4.1, we have

$$\begin{aligned} &\max_{1\leq l\leq h_t} \sum_{k=0}^{t} \boldsymbol{W}_{k+1}^T \boldsymbol{d}_k(l)\boldsymbol{\Phi}_k^T(l)\boldsymbol{P}_k(l)\boldsymbol{\Phi}_k(l)\boldsymbol{W}_{k+1} \\ &\leq \max_{1\leq l\leq h_t} \sum_{k=0}^{t} \lambda_{\max}\{\boldsymbol{d}_k(l)\boldsymbol{\Phi}_k^T(l)\boldsymbol{P}_k(l)\boldsymbol{\Phi}_k(l)\}\|\boldsymbol{W}_{k+1}\|^2 \\ &\leq \max_{1\leq l\leq h_t} \sum_{k=0}^{t} \lambda_{\max}\{\boldsymbol{d}_k(l)\boldsymbol{\Phi}_k^T(l)\boldsymbol{P}_k(l)\boldsymbol{\Phi}_k(l)\}(\|\boldsymbol{W}_{k+1}\|^2 - E(\|\boldsymbol{W}_{k+1}\|^2|\mathscr{F}_k)) \\ &\quad + \max_{1\leq l\leq h_t} \sum_{k=0}^{t} \lambda_{\max}\{\boldsymbol{d}_k(l)\boldsymbol{\Phi}_k^T(l)\boldsymbol{P}_k(l)\boldsymbol{\Phi}_k(l)\}E(\|\boldsymbol{W}_{k+1}\|^2|\mathscr{F}_k) \\ &= o(\eta^2(t)\log\log t) + O(h_t \log t). \quad (4.6) \end{aligned}$$

Substituting (4.6) into (4.4) yields the result of the lemma. ∎

Compared with [15], we need to deal with the cooperative effect of cumulative noises of multiple sensors which is shown by the following lemma. This lemma can be derived by following the proof line of Lemma 3.3, and we omit it here.

LEMMA 4.3. *Under Assumptions* 3.1 *and* 4.1–4.3*, for any* $i \in \{1, ..., n\}$*, we have*

$$\max_{1\leq l\leq h_t} \left\{ \boldsymbol{S}_{t,i}^T(l)\boldsymbol{P}_{t,i}(l)\boldsymbol{S}_{t,i}(l) \right\} = O(h_t \log t) + o(\{\eta(t)\log\log t\}^2),$$

where $\boldsymbol{S}_{t,i}(l) = \left(\sum_{j=1}^{n} \sum_{k=0}^{t-1} a_{ij}^{(t-k)} \boldsymbol{\varphi}_{k,j}(l) w_{k+1,j} \right)$*, and* h_t *and* $\eta(t)$ *are defined in Assumption* 4.3.

The following theorem will establish the convergence of the estimates $\hat{m}_{t,i}, \hat{p}_{t,i}, \hat{q}_{t,i}$ and $\boldsymbol{\theta}_{t,i}(\hat{p}_{t,i}, \hat{q}_{t,i})$ given by Algorithm 4.1 to the true values.

THEOREM 4.4. *Under Assumptions* 3.1 *and* 4.1–4.3*, we have for any* $i \in \{1,\ldots,n\}$,

$$\hat{m}_{t,i} \xrightarrow[t\to\infty]{} m_0 \quad \text{a.s.,} \tag{4.7}$$

$$(\hat{p}_{t,i}, \hat{q}_{t,i}) \xrightarrow[t\to\infty]{} (p_0, q_0) \quad \text{a.s.,} \tag{4.8}$$

$$\boldsymbol{\theta}_{t,i}(\hat{p}_{t,i}, \hat{q}_{t,i}) \xrightarrow[t\to\infty]{} \boldsymbol{\theta}(p_0, q_0) \quad \text{a.s.} \tag{4.9}$$

Proof. We first show that $\limsup_{t\to\infty} \hat{m}_i(t) \le m_0$ a.s. For $p > p_0$, $q > q_0$, set

$$\boldsymbol{\theta}(p,q) = [b_1,\ldots,b_p,c_1,\ldots,c_q]^T,$$
$$\widetilde{\boldsymbol{\theta}}_{t,i}(p,q) = \boldsymbol{\theta}(p,q) - \boldsymbol{\theta}_{t,i}(p,q),$$

where $b_p = 0, p > p_0$, $c_q = 0, q > q_0$, $\boldsymbol{\theta}_{t,i}(p,q)$ is obtained by Algorithm 2.1.

Then by (3.5), for $l \ge m_0$, we have

$$\begin{aligned}
\sigma_{t,i}(l,l,\boldsymbol{\theta}_{t,i}(l,l)) &= \sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)}[\widetilde{\boldsymbol{\theta}}_{t,i}^T(l)\boldsymbol{\varphi}_{k,j}(l) + w_{k+1,j}]^2 \\
&= \widetilde{\boldsymbol{\theta}}_{t,i}^T(l)\left(\sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)}\boldsymbol{\varphi}_{k,j}(l)\boldsymbol{\varphi}_{k,j}^T(l)\right)\widetilde{\boldsymbol{\theta}}_{t,i}(l) \\
&\quad + 2\widetilde{\boldsymbol{\theta}}_{t,i}^T(l)\left(\sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)}\boldsymbol{\varphi}_{k,j}(l)w_{k+1,j}\right) + \sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)}w_{k+1,j}^2 \\
&\triangleq I_1 + I_2 + \sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)}w_{k+1,j}^2.
\end{aligned} \tag{4.10}$$

In the following, we estimate I_1, I_2 separately.

On the one hand, by (3.18) (3.21) and (3.23), we have

$$\begin{aligned}
I_1 + I_2 &= \widetilde{\boldsymbol{\theta}}_{t,i}^T(l)\left(\boldsymbol{P}_{t,i}^{-1}(l) - \sum_{j=1}^{n} a_{ij}^{(t)}\boldsymbol{P}_{0,j}^{-1}(l)\right)\widetilde{\boldsymbol{\theta}}_{t,i}(l) \\
&\quad + 2\widetilde{\boldsymbol{\theta}}_{t,i}^T(l)\left(\sum_{j=1}^{n} a_{ij}^{(t)}\boldsymbol{P}_{0,j}^{-1}(l)\widetilde{\boldsymbol{\theta}}_{0,j}(l) - \boldsymbol{P}_{t,i}^{-1}(l)\widetilde{\boldsymbol{\theta}}_{t,i}(l)\right) \\
&= -\widetilde{\boldsymbol{\theta}}_{t,i}^T(l)\boldsymbol{P}_{t,i}^{-1}(l)\widetilde{\boldsymbol{\theta}}_{t,i}(l) + 2\widetilde{\boldsymbol{\theta}}_{t,i}^T(l)\left(\sum_{j=1}^{n} a_{ij}^{(t)}\boldsymbol{P}_{0,j}^{-1}(l)\widetilde{\boldsymbol{\theta}}_{0,j}(l)\right) \\
&\quad - \widetilde{\boldsymbol{\theta}}_{t,i}^T(l)\left(\sum_{j=1}^{n} a_{ij}^{(t)}\boldsymbol{P}_{0,j}^{-1}(l)\right)\widetilde{\boldsymbol{\theta}}_{t,i}(l) \\
&\le -\widetilde{\boldsymbol{\theta}}_{t,i}^T(l)\boldsymbol{P}_{t,i}^{-1}(l)\widetilde{\boldsymbol{\theta}}_{t,i}(l) + \sum_{j=1}^{n}\left(a_{ij}^{(t)}\widetilde{\boldsymbol{\theta}}_{0,j}^T(l)\boldsymbol{P}_{0,j}^{-1}(l)\widetilde{\boldsymbol{\theta}}_{0,j}(l)\right).
\end{aligned} \tag{4.11}$$

Hence by (4.10) and (4.11), we have for $l \geq m_0$,

$$\sigma_{t,i}(l,l,\boldsymbol{\theta}_{t,i}(l,l)) - \sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)} w_{k+1,j}^2 \leq -\widetilde{\boldsymbol{\theta}}_{t,i}^T(l)\boldsymbol{P}_{t,i}^{-1}(l)\widetilde{\boldsymbol{\theta}}_{t,i}(l) + O(1). \tag{4.12}$$

On the other hand, by Lemma 4.3, we have

$$\begin{aligned}|I_2| &\leq 2\left\|\widetilde{\boldsymbol{\theta}}_{t,i}^T(l)\boldsymbol{P}_{t,i}^{-\frac{1}{2}}(l)\right\| \cdot \left\|\boldsymbol{P}_{t,i}^{\frac{1}{2}}(l)\left(\sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)}\boldsymbol{\varphi}_{k,j}(l)w_{k+1,j}\right)\right\| \\ &\leq O\left\{\left\|\widetilde{\boldsymbol{\theta}}_{t,i}^T(l)\boldsymbol{P}_{t,i}^{-\frac{1}{2}}(l)\right\| \cdot \{o([\eta(t)\log\log t]^2) + O(h_t \log t)\}^{\frac{1}{2}}\right\}.\end{aligned}$$

Then for $l \geq m_0$ and sufficiently large t, by (3.18), (4.10), and Assumption 4.3 , we have for some positive constant C_6,

$$\begin{aligned}&\sigma_{t,i}(l,l,\boldsymbol{\theta}_{t,i}(l,l)) - \sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)} w_{k+1,j}^2 \\ \geq\ & \frac{1}{2}\widetilde{\boldsymbol{\theta}}_{t,i}^T(l)\boldsymbol{P}_{t,i}^{-1}(l)\widetilde{\boldsymbol{\theta}}_{t,i}(l) \\ &- C_6\left\{\left\|\widetilde{\boldsymbol{\theta}}_{t,i}^T(l)\boldsymbol{P}_{t,i}^{-\frac{1}{2}}(l)\right\| \cdot \{o([\eta(t)\log\log t]^2) + h_t \log t\}^{\frac{1}{2}}\right\}.\end{aligned}$$

Hence by (4.3) and Lemma 4.2, we have for sufficiently large t,

$$\begin{aligned}&\max_{m_0 < l \leq \log t}\{\bar{L}_{t,i}(m_0,m_0) - \bar{L}_{t,i}(l,l)\} \\ =\ & \max_{m_0 < l \leq \log t}\Big\{\sigma_{t,i}(m_0,m_0,\boldsymbol{\theta}_{t,i}(m_0,m_0)) - \sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)} w_{k+1,j}^2 \\ &- \sigma_{t,i}(l,l,\boldsymbol{\theta}_{t,i}(l,l)) + \sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)} w_{k+1,j}^2 - 2(l-m_0)\bar{a}_t\Big\} \\ \leq\ & \max_{m_0 < l \leq \log t} O\left\{\left\|\widetilde{\boldsymbol{\theta}}_{t,i}^T(l)\boldsymbol{P}_{t,i}^{-\frac{1}{2}}(l)\right\|\{o([\eta(t)\log\log t]^2) + O(h_t \log t)\}^{\frac{1}{2}}\right\} + O(1) - 2\bar{a}_t \\ =\ & o([\eta(t)\log\log t]^2) + O(h_t\log t) + O(1) - 2\bar{a}_t < 0.\end{aligned}$$

By the above equation, we have

$$\bar{L}_{t,i}(m_0,m_0) < \min_{m_0 < l \leq \log t} \bar{L}_{t,i}(l,l),$$

which implies that $\limsup_{t\to\infty} \hat{m}_{t,i} \leq m_0$.

We now show that $\liminf_{t\to\infty} \hat{m}_{t,i} \geq m_0$ holds a.s. For any $l \leq m_0$, let us write $\boldsymbol{\theta}_{t,i}(l)$ given by Algorithm 2.1 in its component form:

$$\boldsymbol{\theta}_{t,i}(l) = [b_{1,t}^i, \ldots, b_{l,t}^i, c_{1,t}^i, \ldots, c_{l,t}^i]^T \in \mathbb{R}^{2l}.$$

In order to avoid confusion, for any $l \leq m_0$, we denote the following m_0-dimensional vector,

$$\boldsymbol{\theta}_{t,i}(m_0) = [b_{1,t}^i, \ldots, b_{m_0,t}^i, c_{1,t}^i, \ldots, c_{m_0,t}^i]^T \in \mathbb{R}^{2m_0},$$

where $b^i_{j,t}=0, c^i_{j,t}=0$ for $m_0>j>l$.

For any $l\le m_0$, we have

$$\begin{aligned} &y_{k+1,j}-\boldsymbol{\theta}^T_{t,i}(l)\boldsymbol{\varphi}_{k,j}(l)=y_{k+1,j}-\boldsymbol{\theta}^T_{t,i}(m_0)\boldsymbol{\varphi}_{k,j}(m_0)\\ &=y_{k+1,j}-\boldsymbol{\theta}^T(m_0)\boldsymbol{\varphi}_{k,j}(m_0)+[\boldsymbol{\theta}^T(m_0)-\boldsymbol{\theta}^T_{t,i}(m_0)]\boldsymbol{\varphi}_{k,j}(m_0)\\ &=w_{k+1,j}+\widetilde{\boldsymbol{\theta}}^T_{t,i}(m_0)\boldsymbol{\varphi}_{k,j}(m_0), \end{aligned}$$

where $\widetilde{\boldsymbol{\theta}}_{t,i}(m_0)=\boldsymbol{\theta}(m_0)-\boldsymbol{\theta}_{t,i}(m_0)\in\mathbb{R}^{2m_0}$.

Hence by (3.5), we have for any $l\le m_0$

$$\sigma_{t,i}(l,l,\boldsymbol{\theta}_{t,i}(l,l))=\sum_{j=1}^n\sum_{k=0}^{t-1}a_{ij}^{(t-k)}(w_{k+1,j}+\widetilde{\boldsymbol{\theta}}^T_{t,i}(m_0)\boldsymbol{\varphi}_{k,j}(m_0))^2.$$

Thus, we have for any $l\le m_0$

$$\begin{aligned} &\sigma_{t,i}(l,l,\boldsymbol{\theta}_{t,i}(l,l))-\sum_{j=1}^n\sum_{k=0}^{t-1}a_{ij}^{(t-k)}w^2_{k+1,j}\\ &=\widetilde{\boldsymbol{\theta}}^T_{t,i}(m_0)\left(\sum_{j=1}^n\sum_{k=0}^{t-1}a_{ij}^{(t-k)}\boldsymbol{\varphi}_{k,j}(m_0)\boldsymbol{\varphi}^T_{k,j}(m_0)\right)\widetilde{\boldsymbol{\theta}}_{t,i}(m_0)\\ &\quad+2\widetilde{\boldsymbol{\theta}}^T_{t,i}(m_0)\left(\sum_{j=1}^n\sum_{k=0}^{t-1}a_{ij}^{(t-k)}\boldsymbol{\varphi}_{k,j}(m_0)w_{k+1,j}\right)\triangleq J_1+J_2. \end{aligned}\tag{4.13}$$

In the following, we estimate J_1,J_2. For $l<m_0$, by the definition of $\widetilde{\boldsymbol{\theta}}_{t,i}(m_0)$, we have

$$\|\widetilde{\boldsymbol{\theta}}_{t,i}(m_0)\|^2\ge\min\{|b_{p_0}|^2,|c_{q_0}|^2\}=\alpha_0>0.\tag{4.14}$$

Then by (3.18), we have

$$\widetilde{\boldsymbol{\theta}}^T_{t,i}(m_0)\boldsymbol{P}^{-1}_{t,i}(m_0)\widetilde{\boldsymbol{\theta}}_{t,i}(m_0)\ge a_{\min}\lambda^0_{\min}(t)\alpha_0.\tag{4.15}$$

Moreover, by (2.1), Assumption 4.3, and Lemma 4.2, we have

$$\begin{aligned} &\widetilde{\boldsymbol{\theta}}^T_{t,i}(m_0)\left(\sum_{j=1}^n a_{ij}^{(t)}\boldsymbol{P}^{-1}_{0,j}(m_0)\right)\widetilde{\boldsymbol{\theta}}_{t,i}(m_0)\\ &\le\lambda_{\max}\left(\sum_{j=1}^n a_{ij}^{(t)}\boldsymbol{P}^{-1}_{0,j}(m_0)\right)\|\widetilde{\boldsymbol{\theta}}_{t,i}(m_0)\|^2=O(1). \end{aligned}$$

Then for $l<m_0$, by (3.18), we obtain for some positive constant C_7,

$$\begin{aligned} J_1&=\widetilde{\boldsymbol{\theta}}^T_{t,i}(m_0)\boldsymbol{P}^{-1}_{t,i}(m_0)\widetilde{\boldsymbol{\theta}}_{t,i}(m_0)-\widetilde{\boldsymbol{\theta}}^T_{t,i}(m_0)\left(\sum_{j=1}^n a_{ij}^{(t)}\boldsymbol{P}^{-1}_{0,j}(m_0)\right)\widetilde{\boldsymbol{\theta}}_{t,i}(m_0)\\ &\ge\widetilde{\boldsymbol{\theta}}^T_{t,i}(m_0)\boldsymbol{P}^{-1}_{t,i}(m_0)\widetilde{\boldsymbol{\theta}}_{t,i}(m_0)-C_7. \end{aligned}\tag{4.16}$$

By Lemma 4.3, we have

$$|J_2| \le 2\left\|\widetilde{\boldsymbol{\theta}}^T_{t,i}(m_0)\boldsymbol{P}^{-\frac{1}{2}}_{t,i}(m_0)\right\| \cdot \left\|\boldsymbol{P}^{\frac{1}{2}}_{t,i}(m_0)\left(\sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)}\boldsymbol{\varphi}_{k,j}(m_0)w_{k+1,j}\right)\right\|$$

$$\le O\left\{\left\|\widetilde{\boldsymbol{\theta}}^T_{t,i}(m_0)\boldsymbol{P}^{-\frac{1}{2}}_{t,i}(m_0)\right\| \cdot \{o([\eta(t)\log\log t]^2) + O(h_t \log t)\}^{\frac{1}{2}}\right\}. \tag{4.17}$$

Then by (4.15)–(4.17) and Assumption 4.3, we have for large t

$$\begin{aligned} J_1 + J_2 &\ge \widetilde{\boldsymbol{\theta}}^T_{t,i}(m_0)\boldsymbol{P}^{-1}_{t,i}(m_0)\widetilde{\boldsymbol{\theta}}_{t,i}(m_0) - C_8\Big\{\left\|\widetilde{\boldsymbol{\theta}}^T_{t,i}(m_0)\boldsymbol{P}^{-\frac{1}{2}}_{t,i}(m_0)\right\| \\ &\quad \cdot\{o([\eta(t)\log\log t]^2) + O(h_t\log t)\}^{\frac{1}{2}}\Big\} - C_7 \\ &\ge a_{\min}\alpha_0\lambda^0_{\min}(t)(1+o(1)) \text{ a.s.,} \end{aligned}$$

where C_8 is a positive constant.

Hence by (4.13), we have for any $l < m_0$,

$$\sigma_{t,i}(l,l,\boldsymbol{\theta}_{t,i}(l,l)) \ge a_{\min}\alpha_0\lambda^0_{\min}(t)(1+o(1)) + \sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)} w^2_{k+1,j}. \tag{4.18}$$

For $l = m_0$, by (4.12) and Lemma 4.2, we have

$$\begin{aligned} &\sigma_{t,i}(m_0, m_0, \boldsymbol{\theta}_{t,i}(m_0,m_0)) \\ &\le -\widetilde{\boldsymbol{\theta}}^T_{t,i}(m_0)\boldsymbol{P}^{-1}_{t,i}(m_0)\widetilde{\boldsymbol{\theta}}_{t,i}(m_0) + \sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)} w^2_{k+1,j} + O(1) \\ &\le O(h_t\log t) + o([\eta(t)\log\log t]^2) + \sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)} w^2_{k+1,j} + O(1). \qquad (4.19) \end{aligned}$$

For any $l < m_0$, by (4.18)–(4.19) and Assumption 4.3, we have

$$\begin{aligned} &\bar{L}_{t,i}(l,l) - \bar{L}_{t,i}(m_0,m_0) \\ &= \sigma_{t,i}(l,l,\boldsymbol{\theta}_{t,i}(l,l)) - \sigma_{t,i}(m_0,m_0,\boldsymbol{\theta}_{t,i}(m_0,m_0)) + 2(l-m_0)\bar{a}_t \\ &\ge a_{\min}\alpha_0\lambda^0_{\min}(t)(1+o(1)) - C_9((h_t\log t) + o([\eta(t)\log\log t]^2)) + C_{10} - C_{11}\bar{a}_t \\ &= a_{\min}\lambda^0_{\min}(t)(\alpha_0 + o(1)) > 0 \text{ as } t\to\infty, \end{aligned}$$

where C_9, C_{10}, C_{11} are positive constants. Hence we have

$$\bar{L}_{t,i}(m_0,m_0) < \min_{1\le l<m_0} \bar{L}_{t,i}(l,l),$$

which implies that $\liminf_{t\to\infty} \hat{m}_{t,i} \ge m_0$ a.s. Thus the first assertion (4.7) has been proved.

By $\hat{m}_{t,i} \xrightarrow[t\to\infty]{} m_0$, the proof of (4.8) can be carried out by a similar argument as that used in section 3.

Note that both the estimates $(\hat{p}_{t,i}, \hat{q}_{t,i})$ and the true orders (p_0, q_0) are integers, from (4.8); we see that there exists a large enough T such that $\hat{p}_{t,i} = p_0$ and $\hat{q}_{t,i} = q_0$ for $t \ge T$. By the proof of Lemma 4.2, we have

$$V_t(p_0,q_0) = O(h_t\log t) + o(\eta^2(t)\log\log t).$$

Therefore,

$$\|\boldsymbol{\theta}_{t,i}(p_0,q_0)-\boldsymbol{\theta}(p_0,q_0)\|^2=\frac{O(h_t\log t)+o(\eta^2(t)\log\log t)}{\lambda^0_{\min}(t)}.$$

The convergence of the parameters can be obtained by Assumption 4.3. This completes the proof of the theorem. ∎

Remark 4.2. From Theorem 4.4 (also Theorems 3.4 and 3.5), we see that the convergence of the estimates for both the system orders and parameters are derived without using the independency or stationarity assumptions on the regression vectors, which makes it possible to apply our distributed algorithms to practical feedback control systems.

Remark 4.3. We note that in the area of stochastic adaptive control and optimization (see e.g., [42, 43]), the dither signals are often introduced into the controller design to deal with the fundamental conflict between parameter estimation and control performance, i.e., the adaptive controller can be taken as the form $u_{t,i}=u^0_{t,i}+v_{t,i}$, where $u^0_{t,i}$ is the desired controller and $v_{t,i}$ is the dither signal to enhance the excitation condition in the controller. By following similar analyses as those used in [42] and [43], we can show that cooperative excitation conditions I and II can be satisfied when the controllers are designed according to the above manner, and thus the convergence of the distributed algorithms can be obtained.

5. A simulation example. In this section, we provide an example to demonstrate the theoretical results obtained in this paper. For convenience, we just consider the first case where the true orders have known upper bounds, and the results for the case where the upper bounds of the true orders are unknown are almost the same.

Example 5.1. Consider a network composed of $n=6$ sensors whose dynamics obey the following dynamic model

$$y_{t+1,i}=\boldsymbol{\theta}^T\boldsymbol{\varphi}_{t,i}+w_{t+1,i}, \tag{5.1}$$

where both the system order p_0 and the parameter $\boldsymbol{\theta}\in\mathbb{R}^{p_0}$ are unknown. The noise sequence $\{w_{t,i}, t\geq 1, i=1,\ldots,6\}$ in (5.1) is independent and identically distributed with $w_{t,i}\sim\mathcal{N}(0,0.1)$ (Gaussian distribution with zero mean and variance 0.1). Assume that an upper bound $p^*=5$ for the unknown order is available. Let the regression vectors $\boldsymbol{\varphi}_{t,i}\in\mathbb{R}^p$ $(1\leq p\leq p^*,\ i=1,\ldots,6,\ t\geq 1)$ be generated by the following state space model,

$$\begin{aligned}\boldsymbol{x}_{t,i}&=\boldsymbol{A}_i\boldsymbol{x}_{t-1,i}+\boldsymbol{B}_i\varepsilon_{t,i},\\ \boldsymbol{\varphi}_{t,i}&=\boldsymbol{C}_i\boldsymbol{x}_{t,i},\end{aligned} \tag{5.2}$$

where $\boldsymbol{x}_{t,i}\in\mathbb{R}^p$ is the state of the above system with $\boldsymbol{x}_{0,i}=[\underbrace{1,\ldots,1}_{p}]^T$, the matrices $\boldsymbol{A}_i$, $\boldsymbol{B}_i$, and $\boldsymbol{C}_i$ $(i=1,2,\ldots,6)$ are chosen according to the following way such that the excitation condition for any individual sensor cannot be satisfied,

$$\begin{aligned}&\boldsymbol{A}_i=diag\{\underbrace{1.15,\ldots,1.15}_{p}\},\\ &\boldsymbol{B}_i=\boldsymbol{e_j}\in\mathbb{R}^p,\\ &\boldsymbol{C}_i=col\{0,\ldots,0,\underset{j^{th}}{\boldsymbol{e}_j},0,\ldots,0\}\in\mathbb{R}^{p\times p},\end{aligned}$$

where $j = \mathrm{mod}(i, p)$ and $\boldsymbol{e}_j$ $(j = 1, \ldots, m)$ is the jth column of the identity matrix $\boldsymbol{I}_p$ $(1 \le p \le p^*)$, and the noise sequence $\{\varepsilon_{t,i}, t \ge 1, i = 1, \ldots, n\}$ in (5.2) is independent and identically distributed with $\varepsilon_{t,i} \sim \mathcal{N}(0, 0.2)$. All sensors will estimate the system order $p_0 = 4$ and parameter $\boldsymbol{\theta} = [1, 0.5, 3, -1.8]^T$. The initial estimate is taken as $\boldsymbol{\theta}_{0,i} = [\underbrace{2, \ldots, 2}_{p}]^T$ for $i = 1, 2, \ldots, n$. We use the Metropolis rule in [44] to construct the weights in the network topology, i.e.,

$$a_{li} = \begin{cases} 1 - \sum_{j \neq i} a_{ij} & \text{if } \; l = i, \\ 1/(\max\{n_i, n_l\}) & \text{if } \; l \in N_i \setminus \{i\}, \end{cases} \tag{5.3}$$

where n_i is the degree of the node i.

It can be verified that for each sensor i $(i = 1, \ldots, 6)$, the regression signals $\boldsymbol{\varphi}_{t,i}$ (generated by (5.2)) have no adequate excitation to estimate the unknown order and parameter, but they can cooperate to satisfy cooperative excitation condition I (i.e., Assumption 3.3) by taking $a_t = t^\alpha$ with $\alpha > 1$. We repeat the simulation 100 times with the same initial states.

Figures 1 and 2 show the simulation results for the estimation of the unknown system order and parameter generated by Algorithm 3.1 and the non-cooperative

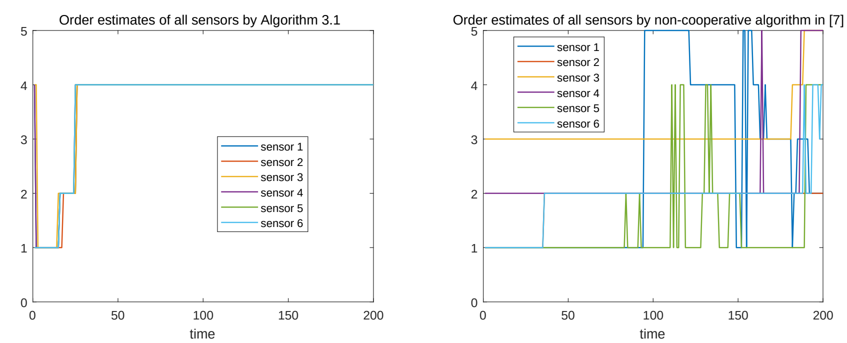

FIG. 1. *The order estimate sequences* $\{p_{t,i}\}_{t=0}^{200}$ *of all sensors by Algorithm* 3.1 *and noncooperative algorithm in* [7].

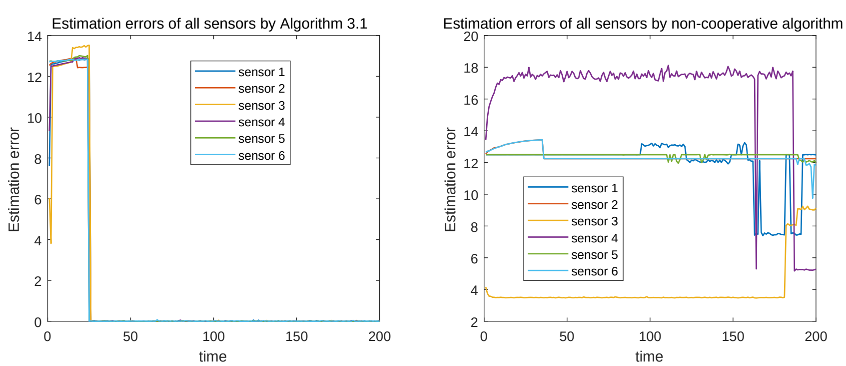

FIG. 2. *The parameter estimation errors of Algorithm* 3.1 *and noncooperative algorithm in* [7].

estimation algorithm, respectively. From Figure 1, we can see that the order estimates $\{p_{t,i}\}_{t=0}^{200}$ $(i = 1, \ldots, 6)$ generated by Algorithm 3.1 can converge to the true order, while the estimates obtained form the noncooperative estimation algorithm cannot. Moreover, from Figure 2, it is clear that the estimation error of unknown parameters $\|\widetilde{\boldsymbol{\theta}}_{t,i}(p^*)\|^2$ obtained by Algorithm 3.1 converges to zero as t increases, while the estimation error obtained by the noncooperative estimation algorithm cannot converge to zero. Therefore, the estimation task can be fulfilled through exchanging information between sensors even though any individual sensor cannot.

6. Conclusion. In this paper, we proposed distributed algorithms to simultaneously estimate both the unknown system orders and parameters by minimizing the LIC and using the distributed LS algorithm. For the case where the upper bounds of true orders are known, we show that the estimates of the parameters and the orders can converge to the true values under the cooperative excitation condition introduced in this paper. We note that the convergence results are obtained without using the independency and stationarity assumptions of regression vectors as is commonly used in most existing literature. Moreover, for the case where the upper bounds of true orders are unknown, we constructed a similar distributed algorithm to estimate both the parameters and the orders by introducing a time-varying regression lag, and obtained the strong consistency of the distributed algorithm. The cooperative excitation condition can reveal the joint effect of multiple sensors. Many interesting problems deserve to be further investigated, for example, the distributed order estimation problem of the autoregressive moving average model with exogenous inputs, the recursive distributed algorithm for the order estimation problem.

REFERENCES

[1] J. Fuchs, R. Weigel, and M. Gardill, *Model order estimation using a multi-layer perceptron for direction-of-arrival estimation in automotive radar sensors*, in Proceedings of IEEE Topical Conference on Wireless Sensors and Sensor Networks (WiSNet), Orlando, FL, IEEE, Piscataway, NJ, 2019, pp. 1–3.

[2] P. W. Pande, B. R. Kumar, S.Chakrabarti, and C. Srivastava, *Model order estimation methods for low frequency oscillations in power systems*, Int. J. Elect. Power Energy Syst., 115 (2020), 105438.

[3] H. Bellahsene and A. Taleb-ahmed, *ARMA order model detection using minimum of kurtosis: Application on seismic data*, Arab. J. Geosci. (Springer), 11 (2018), pp. 1–7.

[4] M. -B. Hossain, J. Moon, and K. H. Chon, *Estimation of ARMA model order via artificial neural network for modeling physiological systems*, IEEE Access, 8 (2020), pp. 186813–186820.

[5] H. Akaike, *A new look at the statistical model identification*, IEEE Trans. Automat. Control, 19 (1974), pp. 716–723.

[6] D. L. Weakliem, *A critique of the Bayesian information criterion for model selection*, Sociol. Methods Res., 27 (1999), pp. 359–397.

[7] L. Guo, H. F. Chen, and J. F. Zhang, *Consistent order estimation for linear stochastic feedback control system (CARMA model)*, Automatica J. IFAC, 25 (1989), pp. 147–151.

[8] H. F. Chen and L. Guo, *Consistent estimation of the order of stochastic control systems*, IEEE Trans. Automat. Control, 32 (1987), pp. 531–535.

[9] H. M. Zhang and P. Wang, *A new way to estimate orders in time series*, J. Time Series Anal., 15 (1994), pp. 545–559.

[10] E. M. Hemerly and M. H. A. Davis, *Strong consistency of the PLS criterion for order determination of autoregressive processes*, Ann. Statist., 17 (1989), pp. 941–946.

[11] H. M. Zhang, *Improved consistent estimation of the order of stochastic control systems*, IEEE Trans. Automat. Control, 35 (1990), pp. 97–100.

[12] E. J. Hannan, *The estimation of the order of an ARMA process*, Ann. Statist., 8 (1980), pp. 1071–1081.

[13] A. Al-Smadi and D. M. Wilkes, *Data-adaptive higher order ARMA model order estimation*, in Proceedings of IEEE Southeastcon '95, Raleigh, NC, IEEE, Piscataway, NJ, 1995, pp. 210–213.

[14] E. J. Hannan and L. Kavalieris, *Multivariate linear time series models*, Adv. in App. Probab., 16 (1984), pp. 492–561.

[15] D. Huang and L. Guo, *Estimation of nonstationary ARMAX models based on the Hannan-Rissanen method*, Ann. Statist., 18 (1990), pp. 1729–1756.

[16] B. Ninness, *On the CRLB for combined model and model-order estimation of stationary stochastic processes*, IEEE Signal Process. Lett., 11 (2004), pp. 293–296.

[17] H.-F. Chen and W.-X. Zhao, *New method of order estimation for ARMA/ARMAX processes*, SIAM J. Control Optim., 48 (2010), pp. 4157–4176.

[18] A. Al-Smadi and D. M.Wilkes, *Robust and accurate ARX and ARMA model order estimation of non-Gaussian processes*, IEEE Trans. Signal Process., 50 (2002), pp. 759–763.

[19] A. P. Liavas, P. A. Regalia, and J.-P. Delmas, *Blind channel approximation: Effective channel order determination*, IEEE Trans. Signal Process., 47 (1999), pp. 3336–3344.

[20] Z. S. Abo-Hammour, O. M. K. Alsmadi , A. M. Al-Smadi, M. I. Zaqout, and M. S. Saraireh, *ARMA model order and parameter estimation using genetic algorithms*, Math. Comput. Model. Dyn. Syst., 18 (2012), pp. 201–221.

[21] J. Moon, M. B. Hossain, and K. H. Chon, *AR and ARMA model order selection for time-series modeling with ImageNet classification*, Signal Process., 183 (2021), 108026.

[22] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayrici, *Wireless sensor networks: A survey*, Comput. Netw., 38 (2002), pp. 393–422.

[23] N. Bogdanović, J. Plata-Chaves, and K. Berberidis, *Distributed incremental-based LMS for node-specific adaptive parameter estimation*, IEEE Trans. Signal Process., 62 (2014), pp. 5382–5397.

[24] S. Xie and L. Guo, *Analysis of normalized least mean squares-based consensus adaptive filters under a general information condition*, SIAM J. Control Optim., 56 (2018), pp. 3404–3431.

[25] A. N. Sadigh, H. S. Yazdi, and A. Harati, *Diversity-based diffusion robust RLS using adaptive forgetting factor*, Signal Process., 182 (2021), 107950.

[26] D. Gan and Z. X. Liu, *Strong consistency of the distributed stochastic gradient algorithm*, in Proceedings of IEEE 58th Conference on Decision and Control, Nice, France, IEEE, Piscataway, NJ, 2019, pp. 5082–5087.

[27] R. Abdolee and B. Champagne, *Diffusion LMS strategies in sensor networks with noisy input data*, IEEE/ACM Trans. Network., 24 (2016), pp. 3–14.

[28] A. Khalili, M. A. Tinati, A. Rastegarnia, and J. A. Chanbers, *Steady-state analysis of diffusion LMS adaptive networks with noisy links*, Signal Process., 60 (2012), pp. 974–979.

[29] F. Barani, A. Savadi, and H. S. Yazdi, *Convergence behavior of diffusion stochastic gradient descent algorithm*, IEEE Trans. Signal Process., 183 (2021), 108014.

[30] G. Q. Wang, N. Li, and Y. G. Zhang, *Diffusion distributed Kalman filter over sensor networks without exchanging raw measurements*, Signal Process., 132 (2017), pp. 1–7.

[31] D. Gan, S. Y. Xie, and Z. X. Liu, *Stability of the distributed Kalman filter using general random coefficients*, Sci. China Inf. Sci., 64 (2021), 172204.

[32] R. Arablouei, K. Dogancay, S. Werner, and Y.-F. Huang, *Adaptive distributed estimation based on recursive least-squares and partial diffusion*, IEEE Trans. Signal Process., 62 (2014), pp. 3510–3522.

[33] Y. Yu, H. Zhao, R. C. de Lamare, Y. Zakharov, and L. Lu, *Robust distributed diffusion recursive least squares algorithms with side information for adaptive networks*, IEEE Trans. Signal Process., 67 (2019), pp. 1566–1581.

[34] V. Vahidpour, A. Rastegarnia, A. Khalili, and S. Sanei, *Analysis of partial diffusion recursive least squares adaptation over noisy links*, IET Signal Processing, 11 (2017), pp. 749–757.

[35] L. Zhang, Y. Cai, C. Li, and R. C. de Lamare, *Variable forgetting factor mechanisms for diffusion recursive least squares algorithm in sensor networks*, EURASIP J. Adv. Signal Process., 57 (2017), https://doi.org/10.1186/s13634-017-0490-z.

[36] S. Y. Xie, Y. Q. Zhang, and L. Guo, *Convergence of a distributed least squares*, IEEE Trans. Automt. Control, 66 (2021), pp. 4952–4959.

[37] S. Y. Xie and L. Guo, *Analysis of distributed adaptive filters based on diffusion strategies over sensor networks*, IEEE Trans. Automat. Control, 63 (2018), pp. 3643–3658.

[38] L. Guo, *Time-varying Stochastic Systems: Stability and Adaptive Theory*, 2nd ed., Science Press, Beijing, 2020.

[39] V. Breschi, A. Bemporad, and I. V. Kolmanovsky, *Cooperative constrained parameter estimation by ADMM-RLS*, Automatica J. IFAC, 121 (2020), pp. 1–14.

[40] J. L. Lei and H. F. Chen, *Distributed estimation for parameter in heterogeneous linear time-varying models with observations at network sensors*, Commun. Info. Syst., 15 (2015), pp. 423–451.

[41] H. F. Chen and L. Guo, *Identification and Stochastic Adaptive Control*, Birkhäsuser, Boston, 1991.

[42] L. Guo and H. F. Chen, *Convergence rate of least squares identification and adaptive control for stochastic systems*, Intern. J. Control, 44 (1986), pp. 1459–1476.

[43] S. Y. Xie, S. Liang, L. Y. Wang, G. Yin, and W. Chen, *Stochastic adaptive optimization with dithers*, IEEE Trans. Automat. Control, doi:10.1109/TAC.2021.3050438, 2021.

[44] L. Xiao, S. Boyd, and S. Lall, *A scheme for robust distributed sensor fusion based on average consensus*, in Proceedings of the 4th Fourth International Symposium on Information Processing in Sensor Networks, Boise, ID, IEEE, Piscataway, NJ, 2005, pp. 63–70.