Distributed adaptive identification for stochastic large models with infinite unknown parameters

Die Gan[1], Yuqi Xu[2], Siyu Xie[2]
1. College of Artificial Intelligence, Nankai University, Tianjin 300350, China.
E-mail: gandie@amss.ac.cn.
2. School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu 611731, China.
Emails: xuyuqi308@163.com, syxie@uestc.edu.cn.

Abstract: This paper studies the distributed adaptive identification problem for stochastic large models. Here the large models refer to those with a large or infinite number of parameters. A novel distributed recursive least squares algorithm is proposed to estimate the unknown system parameters, where the growth rate of regressors' dimension is characterized by a nondecreasing positive function. The almost sure convergence of the proposed algorithm is analyzed under a cooperative decaying excitation condition, which incorporates the temporal information as well as the spatial information to reflect the cooperative effect among multiple agents. Notably, our theoretical results are derived without imposing independence, stationarity, or ergodicity assumptions on the regression vectors, thereby allowing for the inclusion of strongly correlated feedback signals. We employ the double-array martingale theory to address the theoretical challenges arising from time-varying order growth in large models, and further extend the results established for finite-dimensional stochastic regression models and the single-agent case. Finally, a numerical simulation is provided to demonstrate the effectiveness of our findings.

Key Words: Distributed estimation, Large model, Recursive least squares, Convergence, Infinite parameters

1 Introduction

With the rapid growth of computational resources and data availability, large models have garnered widespread attention due to their remarkable performance in various fields. These models are typically characterized by a vast or even infinite number of system parameters, enabling them to capture complex, high-dimensional spatiotemporal dynamics. However, while their architectural complexity provides unprecedented modeling capacity, it simultaneously presents intractable theoretical challenges in parameter identification.

Distributed adaptive estimation, as a prominent method in system identification, can cooperatively accomplish global tasks through local information interaction. Compared with centralized estimation algorithms where a fusion center is needed to collect and process information, distributed adaptive estimation has many advantages including reduction communication and computation costs or increasing robustness to network link failures. Therefore, numerous distributed adaptive estimation algorithms based on typical distributed strategies, such as the incremental, diffusion and consensus have been developed, see [1–5].

To achieve good estimation performance for algorithms, it is often necessary to impose certain conditions on the regression data. Some studies focus on the performance analysis of distributed parameter estimation algorithms using deterministic or time-invariant regressors [2, 6, 7], while the stability of other distributed adaptive estimation algorithms is established based on the independence, stationarity, or ergodicity of stochastic regressors [4, 8–10]. Note that for the typical models such as autoregressive moving average with exogenous input model and Hammerstein system, the regressors are often generated from the past input and output signals, making it difficult for random regressors to satisfy the traditional statistical assumptions of independence, stationarity, and ergodicity. To address these challenges, some effort has been devoted to relaxing the stringent conditions on random regression vectors in non-stationary scenarios, leading to advancements in distributed adaptive estimation and filtering algorithms [11–13]. However, almost all existing theoretical analyses of distributed estimation algorithms have been tailored to finite-order or finite-dimensional system models, which are ill-equipped to address the unique challenges posed by emerging large models with infinite parameters.

In this paper, we focus on the distributed adaptive identification for the large model described by an autoregressive model with exogenous inputs (ARX) system with infinite unknown parameters. We propose a distributed recursive least squares algorithm with increasing dimension regression vectors by using the first adaption and then combination strategy. The almost sure convergence of the proposed algorithm is established under a cooperative decaying excitation condition, which integrates both temporal and spatial information from the random regression vectors. This condition captures the cooperative effect of multiple agents, enabling the the entire multi-agent system to achieve global estimation, even if any individual agent cannot due to a lack of necessary information. Moreover, our theoretical analysis is derived without relying on the stringent conditions for regression vectors such as independence, stationarity or ergodicity, see e.g. [4, 8–10], making it applicable to stochastic feedback control systems. To the best of our knowledge, this work presents the first rigorous theoretical results for distributed identification of stochastic large models with infinite parameters, and the convergence results can be viewed as an extension of the related results for single-agent case in [14], as well as a generalization of the results for finite-dimensional random regression models in [13].

The rest of the paper is organized as follows. The problem formulation and algorithm design are given in Section 2. The

This work was supported by the Natural Science Foundation of Tianjin under Grant 24JCQNJC01930 and Sichuan Science and Technology Program under Grant 2025ZNSFSC1511. Corresponding author: Siyu Xie.

convergence results of the proposed algorithm are carried out in Section 3. A simulation example is provided in Section 4. Conclusion remarks are drawn in Section 5.

2 Problem Formulation

2.1 Some preliminaries

For a matrix $X \in \mathbb{R}^{p\times q}$, $\|X\|$ denotes its Euclidean norm, defined as $\|X\| = \sqrt{\lambda_{\max}(XX^\top)}$, where $(\cdot)^\top$ represents the transpose operator and $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue of the matrix. Corresponding, the minimum eigenvalue and trace of a matrix are denoted as $\lambda_{\min}(\cdot)$ and tr$(\cdot)$. For two symmetric matrices X and Y, $X \geq Y$ means that $X - Y$ is a positive semi-definite matrix. For two positive scalar sequences $\{\alpha_k\}$ and $\{\beta_k\}$, $\alpha_k = O(\beta_k)$ indicates that there exists a positive constant c, which is independent of k, such that $\alpha_k \leq c\beta_k$ holds for any $k \in \mathbb{N}$. and $\alpha_k = o(\beta_k)$ represents $\lim_{k\to\infty} \frac{\alpha_k}{\beta_k} = 0$.

We utilize an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \Lambda)$ to model the communication links between agents, where $\mathcal{V} = \{1, 2, \cdots, n\}$ is the agent set, the edge set $\mathcal{E}$ denotes denotes the edges that signify communication between agents and Λ is the adjacency matrix. Let a_{ij} be the element located in the i-th row and j-th column of Λ. If agent j can receive information from agent i, then $(i, j) \in \mathcal{E}$ and $a_{ij} > 0$, otherwise $(i, j) \notin \mathcal{E}$ and $a_{ij} = 0$. In this paper, we assume that the topology of the agent network is undirected, that is, $a_{ij} > 0$ if and only if $a_{ji} > 0$. The neighbor set of agent i is defined as $\mathcal{N}_i = \{j|(j, i) \in \mathcal{E}\}$. A graph is said to be connected if there exists at least one path for any two agents. The diameter of a graph is defined as the maximum value of the shortest path length between any two agents, which is denoted by $D_{\mathcal{G}}$. For ease of analysis, the adjacency matrix Λ is assumed to be symmetric and doubly stochastic, i.e., $a_{ij} = a_{ji}$ and $\sum_{i=1}^{n} a_{ij} = 1$ for all j.

2.2 Stochastic large model

Consider a network of n interconnected agents, where each agent operates under local sensing and communication. At every time instant t, each agent $i \in \{1, \cdots, n\}$ is assumed to obey the following large model with infinite unknown parameters:

$$\begin{aligned} y_{t,i} &= \sum_{k=1}^{\infty}(A_k y_{t-k,i} + B_k u_{t-k,i}) + \omega_{t,i}, \quad t \geq 0; \\ y_{t,i} &= u_{t,i} = \omega_{t,i} = 0, \ \forall t < 0, \end{aligned} \tag{1}$$

where $y_{t,i} \in \mathbb{R}^m$ and $u_{t,i} \in \mathbb{R}^p$ denote the m-dimensional output and the p-dimensional input of agent i at time instant t respectively, $w_{t,i} \in \mathbb{R}^m$ is the noise process. $A_k \in \mathbb{R}^{m\times m}, B_k \in \mathbb{R}^{m\times p}(k = 1, 2, 3 \dots)$ are the unknown matrices to be estimated and satisfy the following summability condition:

$$\sum_{k=1}^{\infty}(\|A_k\| + \|B_k\|) < \infty. \tag{2}$$

It is worth noting that the model (1) can include the cases of finite, increasing or varying system order, and in these cases, the condition (2) will be satisfied naturally because A_k and B_k will be zero for all suitably large k.

Let z be the backwards-shift operator and introduce:

$$A(z) \triangleq -\sum_{k=1}^{\infty} A_k z^k, \ (A_0 \triangleq -I), \ A_k \in \mathbb{R}^{m\times m}, \tag{3}$$

$$B(z) \triangleq \sum_{k=1}^{\infty} B_k z^k, \ B_k \in \mathbb{R}^{m\times p}. \tag{4}$$

Denote the "transfer function" matrix associated with (1) as $G(z) = [A(z), B(z)]$.

In this paper, we aim to design a distributed adaptive estimation algorithm where all the agents cooperatively estimate or approximate $G(z)$ by the local observed process $\{y_{t,l}, u_{t,l}\}_{l\in\mathcal{N}_i}$ and the incresing lag (denoted by h_t). Then the following norm for measuring the accuracy of the transfer function approximation is used:

$$\|F(z)\|_\infty = \text{ess sup}_{x\in[0,2\pi]} \left\{\lambda_{\max}\left[F(e^{ix})F^*(e^{ix})\right]\right\}^{\frac{1}{2}}, \tag{5}$$

where (5) is the H^∞-norm of a complex matrix $F(z)$ analytic in $|z| < 1$ and bounded almost everywhere on he unit circle.

2.3 Algorithm design

Let $\{h_t\}$ be any non-decreasing sequence of positive integers, $h_t \leq t, \forall t$. Set

$$\theta^\top(t) = [A_1, \cdots, A_{h_t}, B_1, \cdots, B_{h_t}]. \tag{6}$$

Inspired by [13], we propose a distributed recursive least squares algorithm with increasing lag h_t to estimate $\theta(t)$. The details can be seen in Algorithm 1.

Algorithm 1 is designed using the ATC strategy. To be specific, for any fixed time t, we first choose the regression vectors $\varphi_{k,i}(t)$ whose order is only related to h_t (or t). Then in Step 1, each agent i uses the recursive least squares to update the middle estimate $\bar{\theta}_{k+1,i}(t)$ and matrix $\bar{P}_{k+1,i}(t)$. Finally in Step 2, the covariance intersection fusion rule (cf., [15, 16]) is used to diffuse the information vector $\bar{P}^{-1}_{k+1,j}(t)\bar{\theta}_{k+1,j}(t)$ and information matrix $\bar{P}^{-1}_{k+1,j}(t)$ in the form of convex combination for neighbor agents to obtain the matrix $P^{-1}_{k+1,i}(t)$ and estimate $\theta_{k+1,i}(t)$. Unlike the conventional distributed least squares algorithm tailored for finite-order systems (cf. [13]), this paper focuses on a large model characterized by infinite parameters. Consequently, the variables in Algorithm 1 exhibit time-varying increasing order. This situation poses significant challenges in conducting a rigorous convergence analysis. To simplify analysis, we take the initial positive definite matrix $P_{0,i}(t) = \beta_i I(t)$, where $I(t)$ is a square identity matrix of appropriate dimensions and $\beta_i > 0$ for all i.

Let us then write $\hat{\theta}_i(t)$ in its component form

$$\hat{\theta}_i^\top(t) = [A_{1,i}(t), \cdots, A_{h_t,i}(t), B_{1,i}(t), \cdots, B_{h_t,i}(t)], \tag{12}$$

and set

$$\hat{A}_{t,i}(z) = I - \sum_{k=1}^{h_t} A_{k,i}(t)z^k, \ \ \hat{B}_{t,i}(z) = \sum_{k=1}^{h_t} B_{k,i}(t)z^k.$$

Then the estimate $\hat{G}_{t,i}(z)$ for $G(z)$ at time t and agent i can be formed as

$$\hat{G}_{t,i}(z) = [\hat{A}_{t,i}(z), \hat{B}_{t,i}(z)]. \tag{13}$$

Algorithm 1 Distributed recursive least squares algorithm with increasing lag

Initialization. For each agent $i \in \{1, \cdots, n\}$ and any $t > 0$, define the following regressors:

$$\varphi_{k,i}^{\top}(t)=[y_{k,i}^{\top}, \cdots, y_{k-h_t+1,i}^{\top}, u_{k,i}^{\top}, \cdots, u_{k-h_t+1,i}^{\top}], 1 \leq k \leq t,$$

and begin with an initial vector $\hat{\theta}_{0,i}(t)$ and an initial positive definite matrix $P_{0,i}(t) > 0$.

for each time $k = 0, 1, \cdots, t-1$ **do**

for each agent $i = 1, \cdots, n$ **do**

Step 1. Generate $\bar{\theta}_{k+1,i}(t)$ and $\bar{P}_{k+1,i}(t)$ based on $\hat{\theta}_{k,i}(t)$, $P_{k,i}(t)$, $\varphi_{k,i}(t)$:

$$b_{k,i}(t) = [1 + \varphi_{k,i}^{\top}(t) P_{k,i}(t) \varphi_{k,i}(t)]^{-1} \tag{7}$$

$$\bar{\theta}_{k+1,i}(t) = \hat{\theta}_{k,i}(t) + b_{k,i}(t) P_{k,i}(t) \varphi_{k,i}(t) (y_{k+1,i}^{\top} - \varphi_{k,i}^{\top}(t) \hat{\theta}_{k,i}(t)), \tag{8}$$

$$\bar{P}_{k+1,i}(t) = P_{k,i}(t) - b_{k,i}(t) P_{k,i}(t) \varphi_{k,i}(t) \varphi_{k,i}^{\top}(t) P_{k,i}(t), \tag{9}$$

Step 2. Generate $P_{k+1,i}^{-1}(t)$ and $\hat{\theta}_{k+1,i}(t)$ by a convex combination of $\bar{\theta}_{k+1,j}(t)$ and $\bar{P}_{k+1,j}^{-1}(t)$:

$$P_{k+1,i}^{-1}(t) = \sum_{j \in \mathcal{N}_i} a_{ij} \bar{P}_{k+1,j}^{-1}(t), \tag{10}$$

$$\hat{\theta}_{k+1,i}(t) = P_{k+1,i}(t) \sum_{j \in \mathcal{N}_i} a_{ij} \bar{P}_{k+1,j}^{-1}(t) \bar{\theta}_{k+1,j}(t). \tag{11}$$

Output. $\hat{\theta}_i(t) \triangleq \hat{\theta}_{t,i}(t)$.

3 Convergence Results

Assumption 1 *The undirected graph $\mathcal{G}$ is connected.*

Remark 1 *Denote $\Lambda^l = \underbrace{\Lambda \cdots \Lambda}_{l} = [a_{ij}^{(l)}]$, i.e., $a_{ij}^{(l)}$ is the i-th row, j-th column entry of the matrix Λ^l. According to the theory of the product of stochastic matrices, if Assumption 1 hold and l is at least as large as the diameter of the graph, then $a_{ij}^{(l)} \geq a_{\min} > 0$ for all i and j.*

Assumption 2 *For any $i \in \{1, \cdots, n\}$, the input $u_{t,i}$ is an $\mathscr{F}_t$-measurable input, and the noise sequence $\{\omega_{t,i}, \mathscr{F}_t\}$ is a martingale difference sequence satisfying*

$$\sup_{t \geq 0} \mathbb{E}[\|w_{t+1,i}\|^2 | \mathscr{F}_t] < \infty, \ \liminf_{t \to \infty} \frac{1}{t} \sum_{k=0}^{t-1} \|\omega_{k,i}\|^2 \neq 0 \ \text{ a.s.}$$

and $\|w_{t,i}\|^2 = o(d_i(t))$, a.s., where $\{\mathscr{F}_t\}$ is a non-decreasing σ-algebras family and $\{d_i(t)\}$ is a positive, deterministic, nondecreasing sequence and satisfies

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

In Assumption 2, the term "$\|w_{t,i}\|^2 = o(d_i(t))$" characterizes the growth rate of the noise. This implies that the double array martingale estimation theory (Lemma 1 below) can be employed to handle the cumulative effect of the noise in the form of $\max_{1 \leq m \leq h_t} \left\| \sum_{k=1}^{t} f_k(m) w_{k+1} \right\|$. It is straightforward to verify that commonly used bounded or white Gaussian noises can satisfy this assumption.

Lemma 1 [14] *Let $\{v_t, \mathscr{F}_t\}$ be an s-dimensional martingale difference sequence satisfying $\|v_t\| = o(\rho(t))$ a.s., $\sup_t E(\|v_{t+1}\| | \mathscr{F}_t) < \infty$ a.s., where the properties of $\rho(t)$ is described as same as $d_i(t)$ in Assumption 2. Assume that $f_t(m), t, m = 1, 2, ...,$ 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 $h_t = O([\log t]^{\alpha})$ $(\alpha > 0)$, the following property holds as $t \to \infty$,*

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

Lemma 2 *Let $\{v_{k,i}\}$ and $\{\psi_{k,i}(t)\}$ be any m- and h_t-dimensional random sequences respectively. Then we have*

$$\left\| Q_{t,i}^{-\frac{1}{2}}(t) \sum_{j=1}^{n} \sum_{k=0}^{t-1} a_{ij}^{(t-k)} \psi_{k,j}(t) v_{k,j}^{\top} \right\|^2 \leq \sum_{j=1}^{n} \sum_{k=0}^{t-1} \|v_{k,j}(t)\|^2,$$

where $Q_{t,i}(t) = \sum_{j=1}^{n} \sum_{k=0}^{t-1} a_{ij}^{(t-k)} \psi_{k,j}(t) \psi_{k,j}^{\top}(t) + \gamma I(t)$ with $\gamma > 0$ being a positive constant.

The proof of this lemma can be derived by following the proof line of Lemma 4.7 in [14] and Lemma 3.3 in [17], it is complicated and we omit it here due to the limit of the space.

In the following, we establish the convergence of Algorithm 1.

Theorem 1 *Consider the model (1)-(2) and Algorithm 1 with $h_t = O(\log^{\alpha} t)$, $\alpha > 0$. Under Assumptions 1 and 2, then as $t \to \infty$, we have the following estimate for all i:*

$$\|\hat{\theta}_i(t) - \theta(t)\|^2 = O\left(\frac{1}{\lambda_{\min}(t)} \left[h_t \log r_t + \delta_t r_t + o\left([d_t \log\log(t)]^2 \right) \right] \right),$$

where $r_t, \delta_t, \lambda_{\min}(t)$ and $d(t)$ are defined by

$$r_t \triangleq 1 + \sum_{i=1}^{n} \sum_{k=0}^{t-1} (\|y_{k,i}\|^2 + \|u_{k,i}\|^2),$$

$$\delta_t \triangleq \left(\sum_{k=h_t+1}^{\infty} \|A_k\| \right)^2 + \left(\sum_{k=h_t+1}^{\infty} \|B_k\| \right)^2,$$

$$\lambda_{\min}(t) \triangleq \lambda_{\min} \left(\sum_{i=1}^{n} \sum_{k=0}^{t-D_{\mathcal{G}}} \varphi_{k,i}(t) \varphi_{k,i}^{\top}(t) + \sum_{i=1}^{n} \frac{1}{\beta_i} I(t) \right),$$

$$d(t) \triangleq \left(\sum_{i=1}^{n} d_i^2(t) \right)^{\frac{1}{2}}.$$

Proof of sketch. Set

$$\epsilon_{k,i}(t) = \sum_{l=h_t+1}^{\infty} [A_l y_{k+1-l,i} + B_l u_{k+1-l,i}]. \tag{14}$$

From Algorithm 1 and [3], we have

$$\hat{\theta}_i(t) = P_{t,i}(t)\left(\sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)}\varphi_{k,j}(t)y_{k+1,j}^{\top}\right), \quad (15)$$

$$P_{t,i}^{-1}(t) = \sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)}\varphi_{k,j}(t)\varphi_{k,j}^{\top}(t) + \sum_{j=1}^{n} a_{ij}^{(t)}P_{0,j}^{-1}(t). \quad (16)$$

Hence by (14)-(16) we can obtain the following equation:

$$\begin{aligned}
&\|\hat{\theta}_i(t) - \theta(t)\|^2 \\
=&\left\|P_{t,i}(t)\left(\sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)}\varphi_{k,j}(t)y_{k+1,j}^{\top}\right) - \theta(t)\right\|^2 \\
=&\Bigg\|P_{t,i}(t)\Bigg(\sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)}\varphi_{k,j}(t)\left[y_{k+1,j}^{\top} - \varphi_{k,j}^{\top}(t)\theta(t)\right] \\
&- \sum_{j=1}^{n} a_{ij}^{(t)}P_{0,j}^{-1}(t)\theta(t)\Bigg)\Bigg\|^2 \\
=&\Bigg\|P_{t,i}(t)\Bigg(\sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)}\varphi_{k,j}(t)\left[\omega_{k,j}^{\top} + \epsilon_{k,j}^{\top}(t)\right] \\
&- \sum_{j=1}^{n} a_{ij}^{(t)}P_{0,j}^{-1}(t)\theta(t)\Bigg)\Bigg\|^2.
\end{aligned}$$

Furthermore by C_r-inequality, it can be obtained that

$$\begin{aligned}
&\|\hat{\theta}_i(t) - \theta(t)\|^2 \\
\le&3\left\|P_{t,i}^{\frac{1}{2}}(t)\right\|^2\Bigg\{\left\|P_{t,i}^{\frac{1}{2}}(t)\sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)}\varphi_{k,j}(t)\omega_{k,j}^{\top}\right\|^2 \\
&+\left\|P_{t,i}^{\frac{1}{2}}(t)\sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)}\varphi_{k,j}(t)\epsilon_{k,j}^{\top}(t)\right\|^2\Bigg\} \\
&+\left\|P_{t,i}^{\frac{1}{2}}(t)\sum_{j=1}^{n} a_{ij}^{(t)}P_{0,j}^{-1}(t)\theta(t)\right\|^2. \quad (17)
\end{aligned}$$

Now let us estimate the three terms on the right-hand side of (17). Thus we denote

$$M_{t,i}(1) \triangleq \left\|P_{t,i}^{\frac{1}{2}}(t)\sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)}\varphi_{k,j}(t)\omega_{k,j}^{\top}\right\|^2,$$

$$M_{t,i}(2) \triangleq \left\|P_{t,i}^{\frac{1}{2}}(t)\sum_{j=1}^{n}\sum_{k=0}^{t-1} a_{ij}^{(t-k)}\varphi_{k,j}(t)\epsilon_{k,j}^{\top}(t)\right\|^2,$$

$$M_{t,i}(3) \triangleq \left\|P_{t,i}^{\frac{1}{2}}(t)\sum_{j=1}^{n} a_{ij}^{(t)}P_{0,j}^{-1}(t)\theta(t)\right\|^2.$$

For $M_{t,i}(1)$, by (16) and following the proof of Lemma 3.3 in [17] and using Lemma 1 we have

$$\begin{aligned}
M_{t,i}(1) =&O\left(h_t\log^+\lambda_{\max}(P_{t,i}^{-1}(t))\right) + o\left([d_t\log\log(t)]^2\right) \\
=&O\left(h_t\log^+(h_t r_t + h_t)\right) + o\left([d_t\log\log(t)]^2\right) \\
=&O(h_t\log r_t) + o\left([d_t\log\log(t)]^2\right), \quad (18)
\end{aligned}$$

where the last inequality holds due to the fact that $\lim_{t\to\infty} r_t/t > 0$ a.s. and $h_t < t$.

For $M_{t,i}(2)$, by Lemma 2 and the definition of $\epsilon_{k,i}(t)$ in (14), we have the following inequality

$$\begin{aligned}
&M_{t,i}(2) \le \sum_{j=1}^{n}\sum_{k=0}^{t-1}\|\epsilon_{k,j}(t)\|^2 \\
=&\sum_{j=1}^{n}\sum_{k=0}^{t-1}\left\|\sum_{l=h_t+1}^{\infty}[A_l y_{k+1-l,j} + B_l u_{k+1-l,j}]\right\|^2 \\
\le&2\sum_{j=1}^{n}\sum_{k=0}^{t-1}\Big(\Big\|\sum_{l=h_t+1}^{\infty}A_l y_{k+1-l,j}\Big\|^2+\Big\|\sum_{l=h_t+1}^{\infty}B_l u_{k+1-l,j}\Big\|^2\Big) \\
\le&2\sum_{l=h_t+1}^{\infty}\|A_l\|\cdot\sum_{l=h_t+1}^{\infty}\left(\|A_l\|\sum_{j=1}^{n}\sum_{k=0}^{t-1}\|y_{k+1-l,j}\|^2\right) \\
&+2\sum_{l=h_t+1}^{\infty}\|B_l\|\cdot\sum_{l=h_t+1}^{\infty}\left(\|B_l\|\sum_{j=1}^{n}\sum_{k=0}^{t-1}\|u_{k+1-l,j}\|^2\right) \\
\le&2\left(\sum_{l=h_t+1}^{\infty}\|A_l\|\right)^2\left(\sum_{j=1}^{n}\sum_{k=0}^{t-1}\|y_{k,j}\|^2\right) \\
&+2\left(\sum_{l=h_t+1}^{\infty}\|B_l\|\right)^2\left(\sum_{j=1}^{n}\sum_{k=0}^{t-1}\|u_{k,j}\|^2\right) \le 2\delta_t r_t. \quad (19)
\end{aligned}$$

For $M_{t,i}(3)$, by the summability condition (2), we have

$$\begin{aligned}
\sup_t\|\theta(t)\|^2 =& \sup_t\lambda_{\max}\left\{\sum_{k=1}^{h_t}(A_kA_k^{\top} + B_kB_k^{\top})\right\} \\
\le& \sup_t\sum_{k=1}^{h_t}(\|A_k\|^2 + \|B_k\|^2) \\
\le& \sup_t\left(\sum_{k=1}^{h_t}\|A_k\| + \|B_k\|\right)^2 < \infty. \quad (20)
\end{aligned}$$

Hence by (16) and Remark 1, we can conclude that for any t, the following inequality holds with $c_0 > 0$ being a positive constant:

$$P_{t,i}^{-1}(t) \ge \sum_{j=1}^{n} a_{ij}^{(t)}P_{0,j}^{-1}(t) \ge c_0 I(t). \quad (21)$$

and $\lambda_{\min}(P_{t,i}^{-1}(t)) \ge a_{\min}\lambda_{\min}(t)$. Combining (20) with (21), we have

$$M_{t,i}(3) = O(1). \quad (22)$$

Substituting (18), (19) and (22) into (17) we get the result of Theorem 1, which completes the proof of the theorem. ■

Remark 2 *Under some additional conditions, the above result can be refined and articulated in a more concise form. For example, if the noise process $\{\omega_{k,i}\}$ is a Gaussian white noise (i.i.d.) sequence for all i and*

$$\|A_k\| + \|B_k\| = O(\lambda^k),\ 0 < \lambda < 1,\ k \ge 0$$
$$\sum_{i=1}^{n}\sum_{k=0}^{t-1}(\|y_{k,i}\|^2 + \|u_{k,i}\|^2) = O(t^b),\ \textit{for some } b > 0,$$

then by taking the non-decreasing integers $h_t = [\log^\alpha t]$ with some $\alpha > 1$, and taking account of $\omega_{k,i} = O(\log^{1/2} k)$ for all i, we see from Theorem 2 that

$$\|\hat{\theta}_i(t) - \theta(t)\|^2 = O\left(\frac{h_t \log t}{\lambda_{\min}(t)}\right). \tag{23}$$

Note that the related result in [13] is a special case of (23) when h_t is taken as a fixed upper bound for the finite order of the system.

Theorem 2 *Under the conditions of Theorem 1, then as $t \to \infty$, the following asymptotic expansions hold for all i:*

$$\|\hat{G}_{t,i}(z) - G(z)\|_\infty^2 = O\left(\frac{h_t}{\lambda_{\min}(t)}[h_t \log r_t + \delta_t r_t]\right) + o\left(\frac{h_t[d(t)\log\log t]^2}{\lambda_{\min}(t)}\right) \quad \text{a.s.},$$

where $r_t, \delta_t, \lambda_{\min}(t)$ and $d(t)$ are defined in Theorem 1.

Proof. Let us denote

$$G_t(z) = [A_t(z), B_t(z)]$$
$$A_t(z) = I - \sum_{k=1}^{h_t} A_k z^k, \; B_t(z) = \sum_{k=1}^{h_t} B_k z^k.$$

Then by (12)-(13) and C_r-inequality, we know that

$$\begin{aligned}
&\|\hat{G}_{t,i}(z) - G(z)\|_\infty^2 \\
\le& 2\|\hat{G}_{t,i}(z) - G_t(z)\|_\infty^2 + 2\|G_t(z) - G(z)\|_\infty^2 \\
=& 2\left\| \left[\sum_{k=1}^{h_t}(A_k - A_{k,i}(t))z^k, \sum_{k=1}^{h_t}(B_{k,i}(t) - B_k)z^k\right] \right\|_\infty^2 \\
&+ 2\left\| \left[\sum_{k=h_t+1}^{\infty} A_k z^k, \sum_{k=h_t+1}^{\infty} B_k z^k\right] \right\|_\infty^2 \\
\le& 2\left\{\sum_{k=1}^{h_t} \|[A_k - A_{k,i}(t), B_{k,i}(t) - B_k]\|\right\}^2 \\
&+ 2\left\{\sum_{k=h_t+1}^{\infty} (\|A_k\| + \|B_k\|)\right\}^2. \\
\le& 2h_t \sum_{k=1}^{h_t} \|[A_k - A_{k,i}(t), B_{k,i}(t) - B_k]\|^2 \\
&+ 4\left(\sum_{k=h_t+1}^{\infty} \|A_k\|\right)^2 + 4\left(\sum_{k=h_t+1}^{\infty} \|B_k\|\right)^2 \\
\le& 2h_t \text{tr}\left\{\sum_{k=1}^{h_t}[A_k - A_{k,i}(t), B_{k,i}(t) - B_k]\right. \\
&\left. [A_k - A_{k,i}(t), B_{k,i}(t) - B_k]^\top\right\} + 4\delta_t \\
=& 2h_t \text{tr}\{[\hat{\theta}_i^\top(t) - \theta(t)]^\top[\hat{\theta}_i^\top(t) - \theta(t)]\} + 4\delta_t \\
\le& 2mh_t\|\hat{\theta}_i^\top(t) - \theta(t)\|^2 + 4\delta_t.
\end{aligned} \tag{24}$$

Then by Theorem 1 and (24) we obtain the desired result of Theorem 2, which completes the proof. ■

Remark 3 *Under the conditions of Remark 2, we have $\|\hat{G}_{t,i}(z) - G(z)\|_\infty^2 = O\left(\frac{h_t^2 \log t}{\lambda_{\min}(t)}\right)$. If the cooperative decaying excitation condition*

$$\frac{h_t^2 \log t}{\lambda_{\min}(t)} \xrightarrow{t\to\infty} 0, \quad a.s., \tag{25}$$

is satisfied, then we can get the strong convergence of Algorithm 1. The condition (25) is much weaker than the excitation condition (cf.,[14]) designed for the non-cooperative case (i.e., $\mathcal{A} = I_n$): for any agent i, $\frac{h_t^2 \log t}{\lambda_{\min}\left(\sum_{k=0}^{t-1} \varphi_{k,i}(t)\varphi_{k,i}^\top(t) + \gamma I(t)\right)} \xrightarrow{t\to\infty} 0$. It also implies that even if any individual agent can not estimate the unknown parameter matrix accurately by the traditional non-cooperative algorithm, the whole multi-agent system is likely to fulfill the estimation task by using the distributed algorithm, see the example in Section 4.

From Theorems 1 and 2, it is evident that our results are derived without invoking the assumptions of independence or stationarity for the regression signals $\varphi_{k,i}(t)$. This relaxed condition significantly broadens the applicability of our findings, including their use in practical feedback systems.

4 Simulation Results

To validate the estimation performance of Algorithm 1 proposed in the paper, this section provides a simulation experiment. We consider a network consisting of $n = 6$ agents, where each agent exchanges information with its neighboring agents based on a weighted adjacency matrix Λ, which is constructed by using the following Metropolis rule:

$$a_{\ell i} = \begin{cases} 1 - \sum_{j \neq i} a_{ij}, & \text{if } \ell = i\,, \\ 1/\max\{n_i, n_\ell\}, & \text{if } \ell \in \mathcal{N}_i \setminus \{i\}\,, \end{cases}$$

where $a_{\ell i}$ is the weight of the edge connecting the agent i and the agent ℓ, and n_i is the number of neighbors of the agent i. The network topology generated by this rule is shown in Fig. 1.

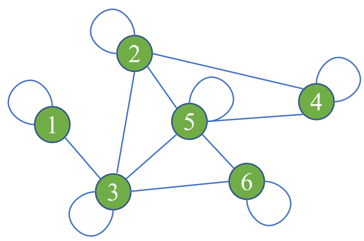

Fig. 1: Network topology of 6 agents.

In this experiment, we estimate a target parameter $\theta(t) \in \mathbb{R}^{2h_t}$ at each time instant t, where $h_t = O(\log t)$. To satisfy the summability condition (2), the parameter $\theta(t)$ is generated as follows:

$$\theta^\top(t) = \left(1, 0.5, \cdots, 0.5^{2h_t - 1}\right) \; (1 \times 2h_t)\,.$$

Also, each agent i follows the observation model:

$$y_{k,i} = \varphi_{k,i}^\top(t)\,\theta(t) + w_{k,i}, \quad 1 \le k \le t,$$

where $\varphi_{k,i}(t) \in \mathbb{R}^{2h_t}$ is the regression vector, and $w_{k,i} \sim \mathcal{N}(0,1)$ is the noise term. To ensure that different agents have distinguishable distribution characteristics and satisfy the cooperative information conditions (25), we select $\varphi_{k,i}(t)$ satisfying $\varphi_{k,i}(t) \sim \mathcal{N}(0, 0.1^i, 2h_t, 1)$ and the i-th number in the array $\varphi_{k,i}(t)$ is set to 0, which causes each agent i to fail to meet the individual excitation condition.

At time instant t, $P_{0,i}(t)$ for each agent i is chosen as the identity matrix, with initial estimates $\hat{\theta}_{0,i}(t) \sim \mathcal{N}(0,1,2h_t,1)$. The experiment is repeated for $s = 100$ trials under the same initial conditions, so that each agent i can obtain the following estimation error sequence:

$$\{\|\hat{\theta}_{i,j}(t) - \theta_j(t)\|^2, t = 1, ..., 50\}, i = 1, ..., 6, j = 1, ..., s.$$

As shown in Fig. 2, it can be observed that when each agent uses the non-cooperative algorithm to estimate the unknown parameter, i.e., $\Lambda = I$, the estimation error cannot converge to 0, indicating that these agents are unable to estimate $\theta(t)$ individually, since the agents do not satisfy the excitation condition mentioned earlier. However, when all agents cooperate with their neighbors by using Algorithm 1 with a connected network, the estimation error converges to 0 at a faster rate and remains stable, which demonstrates the superior estimation performance of Algorithm 1 proposed in this paper.

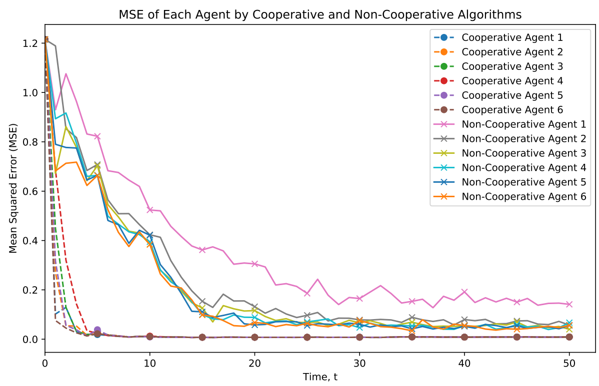

Fig. 2: Performance results for the cooperative and the non-cooperative algorithms.

5 Concluding Remarks

In this paper, we have investigated the distributed adaptive identification for large models with infinite unknown parameters. A distributed recursive least squares algorithm with increasing order is proposed to estimate the unknown invariant parameters in large models. The strong consistency of the parameter estimates in our algorithm is established under a cooperative decaying excitation condition for the random regression vectors, without requiring assumptions of independence, stationarity, or ergodicity. This relaxation enables our theoretical results applicable to the feedback systems. Future research includes the estimation of the time-varying unknown parameters in large models and the combination of the distributed adaptive estimation with the distributed control.

References

[1] A. S. Bijral, A. D. Sarwate, and N. Srebro, "Data-dependent convergence for consensus stochastic optimization," *IEEE Transactions on Automatic Control*, vol. 62, no. 9, pp. 4483–4498, 2017.

[2] Q. Yang, Z. Zhang, M. Fu, and Q. Cai, "Asymptotic convergence of a distributed weighted least squares algorithm for networked systems with vector node variables," *Systems & Control Letters*, vol. 165, p. 105265, 2022.

[3] D. Gan and Z. Liu, "Distributed sparse identification for stochastic dynamic systems under cooperative non-persistent excitation condition," *Automatica*, vol. 151, p. 110958, 2023.

[4] N. Takahashi, I. Yamada, and A. H. Sayed, "Diffusion least-mean squares with adaptive combiners: Formulation and performance analysis," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4795–4810, 2010.

[5] D. Gan and Z. Liu, "Convergence of the distributed SG algorithm under cooperative excitation condition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 5, pp. 7087–7101, 2024.

[6] A. Glushchenko and K. Lastochkin, "Robust time-varying parameters estimation based on I-DREM procedure," *IFAC-PapersOnLine*, vol. 55, no. 12, pp. 91–96, 2022.

[7] J.-G. Lee, Q. V. Tran, K.-H. Oh, P.-G. Park, and H.-S. Ahn, "Distributed object pose estimation over strongly connected networks," *Systems & Control Letters*, vol. 175, p. 105505, 2023.

[8] I. D. Schizas, G. Mateos, and G. B. Giannakis, "Distributed LMS for consensus-based in-network adaptive processing," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2365–2382, 2009.

[9] Y. Hua, F. Wan, B. Liao, Y. Zong, S. Zhu, and X. Qing, "Adaptive multitask clustering algorithm based on distributed diffusion least-mean-square estimation," *Information Sciences*, vol. 606, pp. 628–648, 2022.

[10] C. Philippenko and A. Dieuleveut, "Compressed and distributed least-squares regression: convergence rates with applications to federated learning," *Journal of Machine Learning Research*, vol. 25, pp. 1–80, 2024.

[11] S. Xie and L. Guo, "Analysis of normalized least mean squares-based consensus adaptive filters under a general information condition," *SIAM Journal on Control and Optimization*, vol. 56, no. 5, pp. 3404–3431, 2018.

[12] D. Gan, S. Xie, Z. Liu, and J. Lü, "Stability of FFLS-based diffusion adaptive filter under cooperative excitation condition," *IEEE Transactions on Automatic Control*, vol. 69, no. 11, pp. 7479–7492, 2024.

[13] S. Xie, Y. Zhang, and L. Guo, "Convergence of a distributed least squares," *IEEE Transactions on Automatic Control*, vol. 66, no. 10, pp. 4952–4959, 2021.

[14] H.-F. Chen and L. Guo, *Identification and Stochastic Adaptive Control*. Birkhäuser, Boston, Massachusetts, 1991.

[15] Q. Liu, Z. Wang, X. He, and D. H. Zhou, "On Kalman-consensus filtering with random link failures over sensor networks," *IEEE Transactions on Automatic Control*, vol. 63, no. 8, pp. 2701–2708, 2018.

[16] G. Battistelli, L. Chisci, G. Mugnai, A. Farina, and A. Graziano, "Consensus-based linear and nonlinear filtering," *IEEE Transactions on Automatic Control*, vol. 60, no. 5, pp. 1410–1415, 2015.

[17] D. Gan and Z. Liu, "Distributed order estimation of ARX model under cooperative excitation condition," *SIAM Journal on Control and Optimization*, vol. 60, no. 3, pp. 1519–1545, 2022.