Automatica 183 (2026) 112658

Contents lists available at ScienceDirect

Automatica

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

Distributed estimation with quantized measurements and communication over Markovian switching topologies☆

Ying Wang [a,b], Jian Guo [c], Yanlong Zhao [a,d,*], Ji-Feng Zhang [e,a,d]

[a] State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, PR China
[b] Division of Decision and Control Systems, KTH Royal Institute of Technology, Stockholm 11428, Sweden
[c] Department of Applied Mathematics, The Hong Kong Polytechnic University, HongKong, PR China
[d] School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, PR China
[e] School of Automation and Electrical Engineering, Zhongyuan University of Technology, Zhengzhou 450007, PR China

ARTICLE INFO

Article history:
Received 20 November 2024
Received in revised form 28 June 2025
Accepted 16 September 2025

Keywords:
Distributed estimation
Quantized measurements
Quantized communication
Markovian switching topologies
Stochastic approximation

ABSTRACT

This paper addresses distributed parameter estimation in stochastic dynamic systems with quantized measurements, constrained by quantized communication and Markovian switching directed topologies. To enable accurate recovery of the original signal from quantized communication signal, a persistent excitation-compliant linear compression encoding method is introduced. Leveraging this encoding, this paper proposes an estimation-fusion type quantized distributed identification algorithm under a stochastic approximation framework. The algorithm operates in two phases: first, it estimates neighboring estimates using quantized communication information, then it creates a fusion estimate by combining these estimates through a consensus-based distributed stochastic approximation approach. To tackle the difficulty caused by the coupling between these two estimates, two combined Lyapunov functions are constructed to analyze the convergence performance. Specifically, the mean-square convergence of the estimates is established under a conditional expectation-type cooperative excitation condition and the union topology containing a spanning tree. Besides, the convergence rate is derived to match the step size's order under suitable step-size coefficients. Furthermore, the impact of communication uncertainties including stochastic communication noise and Markov-switching rate is analyzed on the convergence rate. A numerical example illustrates the theoretical findings and highlights the joint effect of sensors under quantized communication.

© 2025 Published by Elsevier Ltd.

1. Introduction

1.1. Background and motivations

Wireless sensor networks (WSNs) are extensively used in fields such as environmental monitoring, intelligent transportation, and smart agriculture due to their easy deployment, flexibility, low power consumption, high accuracy, and reliability (Akyildiz et al., 2002; Ge et al., 2020; Rawat et al., 2014; Taj & Cavallaro, 2011). Distributed parameter estimation over sensor networks is a critical theoretical issue in WSNs research and has garnered significant academic attention (Bianchi et al., 2013; Dimakis et al., 2010; Mateos & Giannakis, 2012). Various effective distributed estimation algorithms have been developed, including distributed stochastic approximation algorithms (Lei & Chen, 2020; Zhang & Zhang, 2012), distributed stochastic gradient algorithms (Gan & Liu, 2024; Swenson et al., 2020), and distributed least squares algorithms (Takahashi et al., 2010; Xie et al., 2021). These distributed approaches only rely on partial information and local data exchange due to collaboration constraints caused by the sensing and communication limitations of individual sensors. However, it is important to note that these algorithms generally assume both accurate communication and precise measurements.

Actually, WSNs consists of numerous spatially distributed sensor nodes, each powered by batteries and equipped with sensing, communication, and computational capabilities. Consequently, WSNs face significant energy and communication constraints due to limited battery life in addition to the inherent collaboration

☆ The work is supported by National Natural Science Foundation of China under Grants 62025306, 62433020, 62303452, 62588101, T2293770 and 62473040, and CAS Project for Young Scientists in Basic Research, China under Grant YSBR-008, China Postdoctoral Program for Innovation Talents under BX20230403. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Antonio Vicino under the direction of Editor Alessandro Chiuso.

* Corresponding author at: State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, PR China.

E-mail addresses: wangying96@amss.ac.cn (Y. Wang), j.guo@amss.ac.cn (J. Guo), ylzhao@amss.ac.cn (Y. Zhao), jif@iss.ac.cn (J.-F. Zhang).

https://doi.org/10.1016/j.automatica.2025.112658
0005-1098/© 2025 Published by Elsevier Ltd.

limitations (Akyildiz et al., 2002; Rawat et al., 2014). Key constraints include:

- Energy constraints (Han et al., 2023a, 2023b): Limited sensing and measurement ranges, restricted processing capabilities, small memory, and more.
- Communication constraints (Ge et al., 2020): Short communication bandwidth, time-varying network topology, network link failures, and more.

It is well-known that both data transmission and measurement consume more energy than local data processing, with communication being the most energy-intensive of the three (Pottie & Kaiser, 2000). Therefore, reducing the bandwidth required for data transmission and the precision of measurement data can significantly extend battery life. Consequently, designing distributed adaptive estimation algorithms that can cooperatively estimate unknown parameters based on quantized transmitted data and quantized measurement data has become an increasingly important research focus.

1.2. Related literature

Quantized data refers to data where only the set it belongs to is known, without precise knowledge of exact values, i.e., data with finite precision or bits (Wang et al., 2003). The existing works on distributed estimation with quantized data generally falls into two categories:

Distributed estimation with quantized measurements: Here, each sensor has quantized measurements but can communicate accurately with its neighbors (Ding et al., 2017; Fu et al., 2022; Han et al., 2023a, 2023b; Wang & Zhang, 2019; Wang et al., 2021). For example, Fu et al. (2022) investigated distributed estimation with binary-valued measurements in time-varying networks, proposing a distributed stochastic approximation algorithm with extended truncation and proving its almost sure convergence under strictly stationary signals. Besides, Wang et al. (2021) developed a Quasi-Newton type projection algorithm with a time-varying projection operator and a diffusion strategy under binary measurements and an undirected, connected topology, achieving almost sure convergence without relying on signal periodicity, independence, or stationarity.

Distributed estimation with quantized communication: In this approach, sensors have accurate measurements but can only communicate with neighbors through quantized channels (Kar et al., 2012; Li et al., 2023; Zhu et al., 2018, 2017, 2015). For instance, Kar et al. (2012) proposed a consensus+innovations distributed algorithm based on stochastic approximation idea and probabilistic quantizer-based communication, showing almost sure convergence in undirected and connected topologies. Additionally, Zhu et al. (2018, 2015) introduced running average distributed estimation algorithms within the consensus+innovation framework, using probabilistic quantized communication with only one measurement per sensor.

While these approaches are insightful, the first class addresses quantized measurement alone, without considering the complexity and energy demands of quantized communication. The second class employs probabilistic quantizers, linking quantized layers to variable size, which hinders finite-bit communication. Meanwhile, it overlooks the irreversible loss of information caused by quantized measurements. Furthermore, both approaches fail to account for the effects of unreliable communication networks, such as time-varying topologies and stochastic communication noise. The goal of this paper is to develop a distributed estimation algorithm that integrates both quantized measurements and quantized communication under Markov-switching topologies and stochastic communication noise, thereby reducing communication bandwidth and measurement precision requirements to extend sensor battery life substantially.

1.3. Main contribution

This paper investigates the distributed parameter estimation problem with both quantized measurements and quantized communication in the presence of Markov-switching directed topologies and stochastic communication noise. The primary contributions of this study, in contrast to previous work, are as follows:

- **Problem aspect**: This work is the first to address distributed parameter estimation with both quantized measurements and quantized communication, while also accounting for communication uncertainties such as Markov-switching topologies and stochastic communication noise (Ding et al., 2017; Fu et al., 2022; Kar et al., 2012; Li et al., 2023; Wang et al., 2021; Zhu et al., 2018, 2015). Additionally, it achieves one-bit communication through a linear compression encoding method, significantly reducing communication costs compared to prior methods (Kar et al., 2012; Li et al., 2023; Zhu et al., 2018, 2015).
- **Algorithm aspect**: An estimation-fusion type quantized distributed identification (EFTQDI) algorithm is proposed to handle quantized measurements and quantized communication. To counteract the information reduction and nonlinearity from quantized communication, the estimate of neighboring estimate is designed using a stochastic approximation approach. Notably, a linear communication encoding method satisfying persistent excitation condition (Wang et al., 2022) is developed, ensuring that the original information can be reconstructed from one-bit transmitted data. Then, the algorithm creates a fusion estimate by using stochastic approximation idea through a combination of these estimates and a consensus strategy.
- **Result aspect**: Two Lyapunov functions are constructed to overcome the coupling difficulty of two estimates, which are introduced to deal with quantized communications and quantized measurements. By analyzing the coupling dynamics of two Lyapunov functions, the mean-square convergence of both the fusion estimate and the estimate of neighboring estimate are established under a conditional expectation-type cooperative excitation condition which is more general than typical cooperative excitation conditions (Fu et al., 2022; Kar et al., 2012; Zhang & Zhang, 2012). Additionally, the mean-square convergence rate of the EFTQDI algorithm is shown to match the order of the step size. This work further analyzes the impact of communication uncertainties, such as stochastic communication noise and Markov-switching rates, on the convergence rate.

The remainder of this paper is organized as follows. Section 2 introduces graph preliminaries and problem formulation. Section 3 presents the proposed algorithm, while Section 4 demonstrates its convergence properties. Proofs of the main results are provided in Section 5. Section 6 includes a numerical example to illustrate the main findings, and Section 7 concludes the paper and discusses potential future work.

Notations. This paper uses $x = [x_1, \ldots, x_n]^T \in \mathbb{R}^n$ and $A \in \mathbb{R}^{m\times n}$ to denote n-dimensional vector and $m \times n$-dimensional real matrix respectively. Moreover, we denote $\|x\| = \sum_{i=1}^{n} x_i^2$ and $\|A\| = \left(\lambda_{\max}(A^TA)\right)^{\frac{1}{2}}$ as the Euclidean norm of vector and matrix respectively, where the notation T denotes the transpose operator and $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of the matrix.

Correspondingly, we use $\lambda_{\min}(\cdot)$ to denote the smallest eigenvalue of the matrix. The function $I_{\{\cdot\}}$ denotes the indicator function, whose value is 1 if its argument is true, and 0, otherwise. Let A and B be two symmetric matrices, then $A \geq B$ means that $A - B$ is a positive semi-definite matrix. The Kronecker product of two matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$ is defined as $A \otimes B \in \mathbb{R}^{mp\times nq}$. Besides, $\mathrm{col}\{u_i\}_{i\in\{1,\ldots,m\}}$ denotes the vector stacked by the specified vectors $u_1, \ldots, u_m$, $\mathrm{diag}\{A_i\}_{i\in\{1,\ldots,m\}}$ represents a block matrix arranged diagonally with the given vectors or matrices $A_1, \ldots, A_m$. Denote $1_m = [1, \ldots, 1]^T \in \mathbb{R}^m$ and $\bar{0}_m = [0, \ldots, 0]^T \in \mathbb{R}^m$.

2. Problem formulation

2.1. Graph preliminaries

In order to describe the relationship between sensor nodes, a sequence of time-varying digraphs $\mathcal{G}_k = (\mathcal{V}, \mathcal{E}_k, \mathcal{A}_k)$ are introduced here, where $\mathcal{V} = \{1, 2, \ldots, m\}$ is the set of nodes, and the node $i \in \mathcal{V}$ represents the sensor node i. $\mathcal{E}_k \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set which is used to describe the communication between agents at time k, and a directed edge $(i, j) \in \mathcal{E}_k$ if and only if there is a communication link from i to j at time k. $\mathcal{A}_k = \{a_{ij}^k\} \in \mathbb{R}^{m\times m}$ is the adjacency matrix of $\mathcal{G}_k$. The elements of the matrix $\mathcal{A}_k$ satisfy: $0 < a_{ij}^k < 1$ if and only if $(j, i) \in \mathcal{E}_k$, otherwise $a_{ij}^k = 0$. The set of the neighbors of agent i at time k is denoted as $\mathcal{N}_i^k = \{j \in \mathcal{V} | (j, i) \in \mathcal{E}_k\}$ and each agent can only exchange information with its neighbors. The Laplacian matrix of $\mathcal{G}_k$ is defined as $\mathcal{L}_k = \mathrm{diag}\{\sum_{j=1}^m a_{1j}^k, \sum_{j=1}^m a_{2j}^k, \ldots, \sum_{j=1}^m a_{mj}^k\} - \mathcal{A}_k$.

The dynamic changes of the communication topology digraph $\mathcal{G}_k$ is described with a Markovian chain $\{m_k, k = 1, 2, \ldots\}$ with a finite state space $\mathcal{S} = \{1, 2, \ldots, s\}$ and transition probability $p_{uv} = \mathbb{P}(m_{k+1} = v \mid m_k = u)$, where u, v belong to $\mathcal{S}$. Then the corresponding communication topology digraph set is $\{\mathcal{G}^{(1)}, \ldots, \mathcal{G}^{(s)}\}$, where $\mathcal{G}^{(u)} = (\mathcal{V}, \mathcal{E}^{(u)}, \mathcal{A}^{(u)})$ is a digraph, $\mathcal{E}^{(u)}$, $\mathcal{A}^{(u)} = \{a_{ij}^{(u)}\}$ and $\mathcal{L}^{(u)}$ are the corresponding edge set, adjacency matrix and Laplacian matrix for $u \in S$. We say $m_k = u$ if and only if $\mathcal{G}_k = \mathcal{G}^{(u)}$. The union topology of $\mathcal{G}^{(1)}, \ldots, \mathcal{G}^{(s)}$ is denoted by $\mathcal{G}_\mathcal{C} = (\mathcal{V}, \mathcal{E}_\mathcal{C}, \mathcal{A}_\mathcal{C})$, where $\mathcal{E}_\mathcal{C} = \bigcup_{u=1}^s \mathcal{E}^{(u)}$ and $\mathcal{A}_\mathcal{C} = \{a_{ij}^\mathcal{C}\}$. A directed tree is a digraph, where each node except the root has exactly one parent node. A spanning tree of $\mathcal{G}_\mathcal{C}$ is a directed tree whose node set is $\mathcal{V}$ and whose edge set is a subset of $\mathcal{E}_\mathcal{C}$. Besides, $\mathcal{G}_\mathcal{C}$ is called a balanced digraph if $\sum_{j=1}^m a_{ij}^\mathcal{C} = \sum_{j=1}^m a_{ji}^\mathcal{C}$ for all $i \in \mathcal{V}$. $\mathcal{G}_\mathcal{C}$ is called an undirected graph, if $\mathcal{A}_\mathcal{C}$ is symmetric. The mirror graph $\hat{\mathcal{G}}_\mathcal{C}$ of the digraph $\mathcal{G}_\mathcal{C}$ is an undirected graph, denoted by $\hat{\mathcal{G}}_\mathcal{C} = (\mathcal{V}, \hat{\mathcal{E}}_\mathcal{C}, \hat{\mathcal{A}}_\mathcal{C})$ with $\hat{\mathcal{A}}_\mathcal{C} = \{\hat{a}_{ij}^\mathcal{C}\}$, $\hat{a}_{ij}^\mathcal{C} = \hat{a}_{ji}^\mathcal{C} = (a_{ij}^\mathcal{C} + a_{ji}^\mathcal{C})/2$.

Remark 2.1. Markovian chain modeling effectively captures the stochastic and time-correlated nature of communication channels, as shown in Zhang and Zhang (2012) and Huang et al. (2012). It describes channel state evolution via state transition probabilities and accounts for dynamic factors like multipath propagation, shadow fading, link failures and random packet loss. Another advantage of Markovian chain modeling is its ability to simplify link failures and reconstructions into a finite-state Markovian chain m_k, where the m_f failure-prone links (with $m_f \leq m(m-1)$) yield at most 2^{m_f} configurations of the WSN. Then, we use m_k to represent these link failures at time k, where m_k is an m_f-dimensional vector with each component taking a value of 0 or 1 to indicate whether a link is absent or present, and its state space $\mathcal{S}$ is finite. Thus, we transform the link failures and reconstructions into the Markovian switching of topologies.

2.2. System description

Consider a multi-agent network consisting of m sensors, the dynamic of the ith ($i = 1, \ldots, m$) sensor at kth times is a linear stochastic model with quantized measurements as follows:

$$\begin{cases} y_{k,i} = \phi_{k,i}^T \theta + d_{k,i}, \\ s_{k,i} = I_{\{y_{k,i} \leq C_i\}}, \end{cases} \tag{1}$$

where $\phi_{k,i} \in \mathbb{R}^n$ is an n-dimensional regressor, $d_{k,i}$ is a stochastic noise and $\theta \in \mathbb{R}^n$ is an unknown parameter vector to be estimated. And $y_{k,i}$ is a scalar measurement of sensor i, which cannot be exactly measured. What can be measured is only the quantized information $s_{k,i}$, where C_i is the fixed threshold of the quantized sensor i.

The information communication between two adjacent sensors is as follows: at the sender side of the communication channel $(j, i) \in \mathcal{E}_k$, the sensor j sends its estimate $\theta_{k-1,j}$ to the sensor i. Due to the uncertainties of communication channels, at the receiver side of the channel (j, i), the sensor i receives the encoding value of the estimate of $\theta_{k-1,j}$ from the sender sensor j as follows

$$\begin{cases} y_{k,ij} = Q(\theta_{k-1,j}) + \omega_{k,ij}, \\ z_{k,ij} = I_{\{y_{k,ij} \leq C_{ij}\}}, \end{cases} \tag{2}$$

where $Q(\cdot) : \mathbb{R}^n \to \mathbb{R}$ is a compression encoding function to be designed, which could transform the vector to the scalar; $\omega_{k,ij}$ is the stochastic communication noise, which can model the thermal noise and channel fading (Li & Zhang, 2010); $z_{k,ij}$ is the one-bit quantized data that the sensor i collects from its neighbor j, where C_{ij} is the fixed threshold of the channel (j, i).

2.3. Assumptions

In order to proceed our analysis, we introduce some assumptions concerning the topologies, regressors, priori information of the unknown parameter and noises. Define the σ-algebra $\mathcal{F}_k = \{\theta_{0,i}, m_l, d_{l,i}, \phi_{l+1,i}, \omega_{l,ij}$ for $i \in \mathcal{V}$ and $j \in \mathcal{N}_i^l, 0 \leq l \leq k\}$.

Assumption 1. The digraph $\mathcal{G}^{(u)}$ is balanced for $u = 1, \ldots, s$ and their union topology $\mathcal{G}_\mathcal{C}$ contains a spanning tree. Besides, the Markov chain m_k is homogeneous and ergodic with a station distribution $\pi_u = \lim_{k\to\infty} \mathbb{P}(m_k = u)$ for all $u \in \mathcal{S}$, where $\sum_{u=1}^s \pi_u = 1$.

Remark 2.2. Define $p_{vu} = \mathbb{P}(m_{k+1} = v \mid m_k = u)$, $q_{ij,k} = \mathbb{P}((i, j) \in \mathcal{E}_k)$ and $P_{u,k} = \mathbb{P}(\mathcal{G}_k = \mathcal{G}^{(u)})$. Then, it can be seen $P_{u,k+1} = \sum_{v=1}^s P_{v,k} p_{vu}$. Define $G_{ij} = \{u \in \mathcal{S} | (i, j) \in \mathcal{E}^{(u)}\}$. Then, we can get $q_{ij,k} = \sum_{u\in G_{ij}} P_{u,k}$ for any $k \geq 1$. From Theorem 1.2 on Seneta (2006), there exists $\lambda_q \in (0, 1)$ such that $q_{ij,k} = \sum_{u\in G_{ij}} \pi_u + O(\lambda_q^k)$. Moreover, if Assumption 1 hold, then it can be seen

$$\underline{\pi} \triangleq \min_{(i,j)\in\mathcal{E}_\mathcal{C}} \sum_{u\in G_{ij}} \pi_u > 0. \tag{3}$$

Assumption 2 (*Conditional Expectation Type Cooperative Excitation Condition*)**.** The regressor $\{\phi_{k,i}, i = 1, \ldots, m\}$ satisfy $\|\phi_{k,i}\| \leq \bar{\phi} < \infty$ and there exist a positive integer h and a positive number δ_ϕ such that

$$\frac{1}{mh} \mathbb{E}\left[\sum_{i=1}^m \sum_{l=k}^{k+h-1} \phi_{l,i}\phi_{l,i}^T \middle| \mathcal{F}_{k-1}\right] \geq \delta_\phi^2 I_n, \forall k \geq 1.$$

Remark 2.3. Assumption 2 is the conditional expectation type cooperative excitation condition, which is more general than cooperative excitation condition (i.e., $\mathbb{E}[\sum_{i=1}^m \phi_{k,i}\phi_{k,i}^T] > 0$) in Fu

et al. (2022), Kar et al. (2012), Lei and Chen (2020). Moreover, it includes the periodical full-rank condition (Wang et al., 2018, 2003; Zhao et al., 2023) and the deterministic persistent excitation condition (Guo & Zhao, 2013; Wang et al., 2022) in the existing quantized identification work.

Assumption 3. There is a known convex compact set $\Omega \subset \mathbb{R}^n$ such that the unknown parameter $\theta \in \Omega$. And denote $\bar{\theta} = \sup_{\eta\in\Omega} \|\eta\|$.

Assumption 4. For the observation noise sequence $\{d_{k,i}\}$, its conditional probability distribution function given $\mathcal{F}_{k-1}$ is denoted by $F_{k,i}(\cdot)$. Besides, the communication noise $\omega_{k,ij}$ is mutually independent with respect to i and j, and the conditional probability distribution function of $\omega_{k,ij}$ given $\mathcal{F}_{k-1}$ is denoted by $G_{k,ij}(\cdot)$.

Remark 2.4. The prior information condition on the unknown parameter (i.e., Assumption 3) is a common assumption of quantized identification studies (Guo & Zhao, 2013; Wang et al., 2021, 2022; Zhang et al., 2021, 2022), primarily used to ensure the boundedness of the estimate. In contrast to previous works (Fu et al., 2022; Guo & Zhao, 2013; Kar et al., 2012; Wang et al., 2021, 2022; You, 2015), Assumption 4 no longer requires the noise to be independently and identically distributed.

The goal of this paper is to develop a distributed estimation algorithm and a suitable quantized communication mechanism to estimate the unknown parameter based on quantized measurement and quantized communication under Markov-switching digraphs.

3. Algorithm design

For the quantized system (1)–(2), the main challenge in distributed estimation lies in handling dual quantized information. Existing approaches quantized communication (Kar et al., 2012; Zhu et al., 2018, 2015), leverage probabilistic quantizers to manage quantized errors as bounded noise with zero mean and finite variance, preserving information in expectation. However, these methods require communication bit counts proportional to data dimension, making low-bit communication difficult. In contrast, this paper focuses on achieving one-bit communication, where the main difficulty lies in balancing the task of estimation while enabling one-bit communication and quantized measurements.

This paper addresses this challenge by designing an estimator based on quantized communication to recover the neighboring estimates. To accurately reconstruct the original estimates, the communication rule must satisfy an excitation condition, where we employ the persistent excitation condition (Guo & Zhao, 2013; Wang et al., 2022) for simplicity. Using these estimated values in place of the transmitted neighboring estimates that are restricted from being transmitted, a consensus strategy is developed to construct a distributed identification algorithm. The resulting estimation-fusion type quantized distributed identification (EFTQDI) algorithm is presented as follows.

Remark 3.1. The core reason this algorithm successfully achieves distributed identification under quantized communication lies in its distinctive design, which requires only the quantized transmission of local estimate information rather than detailed regressor data. First, the applied distributed stochastic approximation-based algorithm relies solely on exchanging local estimates (Lei & Chen, 2020; Zhang & Zhang, 2012). This approach contrasts with distributed least squares or distributed stochastic gradient methods, which require extra sharing of regressor data among sensors (Gan & Liu, 2024; Wang et al., 2021; Xie et al., 2021).

Algorithm 1 EFTQDI algorithm

For any given sensor $i \in \{1, 2, \ldots, m\}$, begin with initial estimates $\hat{\theta}_{0,i} \in \mathbb{R}^n$ and $\theta_{0,i} \in \mathbb{R}^n$. The algorithm is recursively defined as follows:

1: Design of communication rule (design the compression recoding function $Q(\cdot)$ as a linear rule):

$$Q(\theta_{k-1,j}) = \psi_k^T \theta_{k-1,j}, j = 1, \ldots, m,$$

where the encoding rule $\{\psi_k\}$ satisfy $\|\psi_k\| \leq \bar{\psi} < \infty$ and there exists a positive constant δ_ψ such that $\frac{1}{h}\sum_{l=k}^{k+h-1} \psi_l \psi_l^T \geq \delta_\psi^2 I_n$ for all $k \geq 1$. Then, the quantized communication mechanism (2) is

$$z_{k,ij} = I_{\{\psi_k^T \theta_{k-1,j} + \omega_{k,ij} \leq C_{ij}\}}, \forall j \in \mathcal{N}_i^k. \tag{4}$$

2: Estimation (establish the estimate of neighboring estimate based on quantized communication (4)):

$$\hat{\theta}_{k,ij} = \begin{cases} \Pi_\Omega\{\hat{\theta}_{k-1,ij} + \gamma b_k \psi_k (\hat{G}_{k,ij} - z_{k,ij})\}, & \text{if } j \in \mathcal{N}_i^k; \\ \hat{\theta}_{k-1,ij}, & \text{if } j \in \mathcal{N}_i^{\mathcal{C}} \setminus \mathcal{N}_i^k, \end{cases} \tag{5}$$

where the projection $\Pi_\Omega(\cdot)$ is defined as $\Pi_\Omega(x) = \arg\min_{z\in\Omega} \|x - z\|$ for all $x \in \mathbb{R}^n$, $b_k > 0$ is a step size, $\gamma > 0$ is its step coefficient to be designed, $\hat{G}_{k,ij} = G_{k,ij}(C_{ij} - \psi_k^T \hat{\theta}_{k-1,ij})$, and $\mathcal{N}_i^{\mathcal{C}}$ is the neighbor set of the sensor i in the union topology $\mathcal{G}_{\mathcal{C}}$.

3: Fusion (generate $\theta_{k,i}$ based on the estimates of neighboring estimate $\hat{\theta}_{k-1,ij}$ and a consensus strategy):

$$\theta_{k,i} = \Pi_\Omega \Big\{ \theta_{k-1,i} + \beta b_k \Big(\phi_{k,i}(\hat{F}_{k,i} - s_{k,i}) + \sum_{j\in\mathcal{N}_i^k} a_{ij}^k (\hat{\theta}_{k-1,ij} - \theta_{k-1,i}) \Big) \Big\}, \tag{6}$$

where $\beta > 0$ is a step coefficient to be designed and $\hat{F}_{k,i} = F_{k,i}(C_i - \phi_{k,i}^T \theta_{k-1,i})$.

Second, the design of the communication rule ensures the recoverability of the slowly time-varying data before quantized process. These two features significantly reduce communication costs, making the algorithm both highly efficient and well-suited for scenarios with constrained communication resources.

Remark 3.2. The linear encoding coefficients ψ_k is designed to satisfy the persistent excitation condition (Guo & Zhao, 2013; Wang et al., 2022), which is used to ensure that the original vector estimate can be recovered through the quantized communication information. A simple example meeting this encoding-rule is that periodically select ψ_k from the set $\{e_1, e_2, \ldots, e_n\}$, where $e_i = [0, \ldots, 0, 1, 0, \ldots, 0]^T \in \mathbb{R}^n$ is the standard unit basis vectors of Euclidean space $\mathbb{R}^n$. In fact, the condition of ψ_k can be generalized to a conditional expectation-based persistent excitation condition. Specifically, there exists a positive integer h_ψ and a positive number δ_ψ such that $\frac{1}{h_\psi}\mathbb{E}\big[\sum_{l=k}^{k+h_\psi-1} \psi_l \psi_l^T \big| \mathcal{F}_{k-1}\big] \geq \delta_\psi^2 I_n$, where h_ψ could be different with h in Assumption 2.

Remark 3.3. The adjacency matrix $\mathcal{A}_k = \{a_{ij}^k\}$ of the communication topology $\mathcal{G}_k$ has the following properties under Assumption 1: there exists $\lambda_a \in (0, 1)$ such that $\mathbb{E}a_{ij}^k = \sum_{u=1}^s \pi_u a_{ij}^{(u)} + O(\lambda_a^k)$, which can be derived by Remark 2.2 and Theorem 1.2 on Seneta (2006). Therefore, its adjacency matrix satisfies $\mathbb{E}\mathcal{A}_k = \sum_{u=1}^s \pi_u \mathcal{A}^{(u)} + O(\lambda_a^k)$ and its Laplacian matrix $\mathcal{L}_k$ satisfies $\mathbb{E}\mathcal{L}_k = \sum_{u=1}^s \pi_u \mathcal{L}^{(u)} + O(\lambda_a^k)$.

Next, we give the design requirement about the step size b_k in the EFTQDI algorithm as follows.

Assumption 5. The step size b_k satisfies: $\sum_{k=0}^{\infty} b_k = \infty$, $\lim_{k\to\infty} b_k = 0$ and $b_k = O(b_{k+1})$.

Remark 3.4. The conditions of $\sum_{k=0}^{\infty} b_k = \infty$ and $\lim_{k\to\infty} b_k = 0$ are common requirements for stochastic approximation type algorithms (Lei & Chen, 2020; Sinha & Griscik, 1971; Zhang & Zhang, 2012). Besides, $b_k = O(b_{k+1})$ guarantees that the step sizes of the algorithm do not vary significantly over a finite number of steps, which is useful for handling the conditional expectation-type cooperative excitation condition. Moreover, Assumption 5 shows $0 < \sup_{|p-q|<h} b_p/b_q < \infty$.

4. Main results

To clearly present the findings of this paper, this section outlines the main results, including the mean square convergence and convergence rate of the proposed algorithm. Proofs of these results will be provided collectively in the following section. Specifically, Section 4.1 introduces the common notations used in the theoretical analysis, while Section 4.2 presents the key results.

4.1. Notation

For convenience of analysis, we consider the agent i within the union graph $\mathcal{G}_{\mathcal{C}}$, denote N_i as the size of its neighbor set $\mathcal{N}_i^{\mathcal{C}}$ and $N = \sum_{i=1}^{m} N_i$. Denote $a_{ij}^{\mathcal{C}} = \sum_{u=1}^{s} \pi_u a_{ij}^{(u)} \in [0, 1)$ in its adjacency matrix $\mathcal{A}_{\mathcal{C}}$. Accordingly, its Laplacian matrix is given by $\mathcal{L}_{\mathcal{C}} = \sum_{u=1}^{s} \pi_u \mathcal{L}^{(u)}$.

Let $\varepsilon_{k,ij} = \hat{\theta}_{k,ij} - \theta_{k,j}$ as the estimation error of neighboring estimate. Putting $\varepsilon_{k,ij}, j \in \mathcal{N}_i^{\mathcal{C}}$ in a given order yields the error vector ε_k. Without loss of generality, let

$$\varepsilon_k = (\varepsilon_{k,1r_1}, \dots, \varepsilon_{k,1r_{N_1}}, \varepsilon_{k,2r_{N_1+1}}, \dots, \varepsilon_{k,2r_{N_1+N_2}}, \dots, \varepsilon_{k,mr_{N_1+\cdots+N_{m-1}+1}}, \dots, \varepsilon_{k,mr_{N_1+\cdots+N_m}})^T, \tag{7}$$

where $r_1, \dots, r_{N_1} \in \mathcal{N}_1^{\mathcal{C}}$, $r_{N_1+1}, \dots, r_{N_1+N_2} \in \mathcal{N}_2^{\mathcal{C}}, \dots, r_{N_1+\cdots+N_{m-1}+1}, \dots, r_{N_1+\cdots+N_m} \in \mathcal{N}_m^{\mathcal{C}}$.

Then, we construct three matrices to establish the corresponding relation of fusion estimate and its estimates.

R_k is designed to select each neighbor of each sensor node at time k in the union graph $\mathcal{G}_{\mathcal{C}}$. Define $R_k = \mathrm{diag}\{\bar{r}_{k,1}, \dots, \bar{r}_{k,N}\} \in \mathbb{R}^{N\times N}$. Let r_l in vector ε_k represents the neighbor j of agent i, i.e., $\varepsilon_{k,ir_l} = \varepsilon_{k,ij}$, where $r_l \in \mathcal{N}_i^{\mathcal{C}}$ and $l \in \{N_1 + \cdots + N_{i-1} + 1, \dots, N_1 + \cdots + N_i\}$. Then, $\bar{r}_{k,l} = \mathrm{sign}\{a_{ir_l}^k\} = \mathrm{sign}\{a_{ij}^k\}$ for $l \in \{1, \dots, N\}$, where $\mathrm{sign}\{x\}$ is the sign of scalar x and a_{ij}^k is the element of the adjacency matrix $\mathcal{A}_k$.

P_k is designed to select the neighbor set of each node at time k. Define $P_k = [p_{k,1}, \dots, p_{k,m}]^T \in \mathbb{R}^{m\times N}$, where $p_{k,i} = [\vec{0}_{N_1+\cdots+N_{i-1}}^T, a_{ir_{N_1+\cdots+N_{i-1}+1}}^k, \dots, a_{ir_{N_1+\cdots+N_i}}^k, \vec{0}_{N_{i+1}+\cdots+N_m}^T]^T \in \mathbb{R}^N$ for $i = 1, \dots, m$.

Q is designed to select the true neighboring estimate of each node that correlates with its estimate. Define $Q = [q_{1,r_1}, \dots, q_{1,r_{N_1}}, \dots, q_{m,r_{N_1+\cdots+N_{m-1}+1}}, \dots, q_{m,r_{N_1+\cdots+N_m}}]^T \in \mathbb{R}^{N\times m}$, where $q_{i,j} = [\vec{0}_{j-1}^T, 1, \vec{0}_{m-j}^T]^T$ for $(j, i) \in \mathcal{E}^{\mathcal{C}}$.

Next, we introduce the following notations.

$$S_k = \mathrm{col}\{s_{k,i}\}_{i\in\{1,\dots,m\}} \in \mathbb{R}^m$$
$$\Phi_k = \mathrm{diag}\{\phi_{k,i}\}_{i\in\{1,\dots,m\}} \in \mathbb{R}^{mn\times m}$$
$$\Psi_k = I_N \otimes \psi_k \in \mathbb{R}^{nN\times N}$$
$$\Theta_k = \mathrm{col}\{\theta_{k,i}\}_{i\in\{1,\dots,m\}} \in \mathbb{R}^{mn}$$
$$\tilde{\Theta}_k = \mathrm{col}\{\tilde{\theta}_{k,i}\}_{i\in\{1,\dots,m\}} \in \mathbb{R}^{mn}, \text{ where } \tilde{\theta}_{k,i} = \theta_{k,i} - \theta$$
$$\hat{F}_k = \mathrm{col}\{\hat{F}_{k,i}\}_{i\in\{1,\dots,m\}} \in \mathbb{R}^m$$
$$\hat{\Theta}_k = \mathrm{col}\{\hat{\theta}_{k,ij}\}_{(i,j)\in\mathcal{R}} \in \mathbb{R}^{nN}$$
$$Z_k = \mathrm{col}\{z_{k,ij}\}_{(i,j)\in\mathcal{R}} \in \mathbb{R}^N$$
$$\hat{G}_k = \mathrm{col}\{\hat{G}_{k,ij}\}_{(i,j)\in\mathcal{R}} \in \mathbb{R}^N$$

where $\mathcal{R} = \{(1, r_1), \dots, (1, r_{N_1}), (2, r_{N_1+1}), \dots, (2, r_{N_1+N_2}), \dots, (m, r_{N_1+\cdots+N_{m-1}+1}), \dots, (m, r_{N_1+\cdots+N_m})\}$. For any $(j, i) \in \mathcal{E}_{\mathcal{C}} \setminus \mathcal{E}_k$, we set $z_{k,ij} = \hat{G}_{k,ij} = -1$ for all without loss of generality. Actually, $z_{k,ij}$ and $\hat{G}_{k,ij}$ for $(j, i) \in \mathcal{E}_{\mathcal{C}} \setminus \mathcal{E}_k$ can be assigned any constant value without affecting the algorithm, as shown in (5), which indicates that the proposed algorithm does not depend on $z_{k,ij}$ and $\hat{G}_{k,ij}$ when $(j, i) \in \mathcal{E}_{\mathcal{C}} \setminus \mathcal{E}_k$.

Based on the above notations, the EFTQDI algorithm (5)–(6) can be written as

$$\begin{cases} \hat{\Theta}_k = \mathbf{\Pi}\left\{\hat{\Theta}_{k-1} + \gamma b_k R_k \otimes I_n \Psi_k (\hat{G}_k - Z_k)\right\}, \\ \Theta_k = \mathbf{\Pi}\big\{\Theta_{k-1} + \beta b_k \Phi_k (\hat{F}_k - S_k) - \beta b_k \mathcal{L}_k \\ \qquad \otimes I_n \Theta_{k-1} + \beta b_k P_k \otimes I_n \varepsilon_{k-1}\big\}. \end{cases} \tag{8}$$

where $\mathbf{\Pi}\{\cdot\}$ is defined as $\mathbf{\Pi}\{\zeta\} \triangleq (\Pi\{\zeta_1\}, \dots, \Pi\{\zeta_n\})^T$ for $\zeta = \mathrm{col}\{\zeta_i\}_{i\in\{1,\dots,n\}}$, and $\varepsilon_k = \hat{\Theta}_k - Q \otimes I_n \Theta_k$.

4.2. Convergence properties

It is worth noticing that the fusion estimate $\theta_{k,i}$ is related to the estimates of neighboring estimates $\hat{\theta}_{k-1,ij}$, which complicates the convergence analysis of both estimates. To deal with the coupling between these two estimates, we introduce the following two Lyapunov functions,

$$U_k = \mathbb{E}[\varepsilon_k^T \varepsilon_k], \quad V_k = \mathbb{E}[\tilde{\Theta}_k^T \tilde{\Theta}_k]. \tag{9}$$

Then, we will establish the coupling relationship between these two estimates by jointly analyzing these two Lyapunov functions. First, we introduce the following conditions regarding the observation noise $d_{k,i}$ and communication noise $\omega_{k,ij}$.

Assumption 6. The corresponding condition density functions $f_{k,i}(\cdot)$ and $g_{k,ij}(\cdot)$ given $\mathcal{F}_{k-1}$ of the observation noise $d_{k,i}$ and the communication noise $\omega_{k,ij}$ satisfy

$$\underline{f} = \min_i \inf_{k\geq 1} \inf_{|C_i - x| \leq \bar{\phi}\bar{\theta}} f_{k,i}(x) > 0,$$
$$\bar{f} = \max_i \sup_{k\geq 1} \sup_{|C_i - x| \leq \bar{\phi}\bar{\theta}} f_{k,i}(x) < \infty,$$
$$\underline{g} = \min_{(j,i)\in\mathcal{E}_{\mathcal{C}}} \inf_{k\geq 1} \inf_{|C_{ij} - x| \leq \bar{\psi}\bar{\theta}} g_{k,ij}(x) > 0.$$

Then, we present the following lemma to demonstrate the relationship between the Lyapunov function U_k for the fusion estimation error and the Lyapunov function V_k for the neighboring estimation error.

Lemma 4.1. *If Assumption 1, 3–6 hold, then the Lyapunov function U_k for fusion estimation error satisfies*

$$U_k \leq \left(1 - \left(\frac{2\gamma \underline{g}\, \underline{\pi}\, \delta_\psi^2}{c_h} - \beta\left(\frac{\bar{f}\bar{\phi}^2}{\alpha_1} + 2\bar{N} + \frac{\bar{N}}{\alpha_2}\right)\right) \sum_{l=k-h}^{k-1} b_{l+1}\right) U_{k-h} + \beta(\alpha_1 \bar{f}\bar{N}\bar{\phi}^2 + \alpha_2 \lambda_m^2) \sum_{l=k-h}^{k-1} b_{l+1} V_{k-h} + O\left(b_k^2\right),$$

for all $\alpha_1, \alpha_2 \in (0, \infty)$, where $\bar{N} = \max_{i=1,\dots,m} N_i$ is the biggest size of the neighbor set in the union topology $\mathcal{G}_{\mathcal{C}}$, $\lambda_m \triangleq \max_{u=1,\dots,s} \|\mathcal{L}^{(u)}\|$, $c_h \triangleq \sup_{|p-q|<h} \frac{b_p}{b_q}$, $\underline{\pi}$, $\bar{\phi}$, $\underline{g}$, $\bar{f}$, δ_ψ are defined in (3), Assumptions 2, 6 and Algorithm 1.

Remark 4.1. Lemma 4.1 provides the iterative relationship between U_k, U_{k-h} and V_{k-h}, showing that with appropriate step coefficient design, U_k can be reduced relative to U_{k-h}, while still being influenced by the coupling term V_{k-h}. The reason we establish an iterative relationship between U_k and U_{k-h} rather than U_k and U_{k-1} is that, in the latter case (i.e., Eq. (13)), the encoding rule ψ_k does not satisfy the persistent excitation condition under one iteration. This prevents us from establishing a compression relationship between U_k and U_{k-1}, resulting in divergence due to the perturbation term V_{k-1} and making convergence analysis of U_k infeasible. However, when analyzing the relationship between U_k and U_{k-h}, the encoding rule ψ_k, under the persistent excitation condition, provides a compression coefficient (i.e., Eq. (24)), enabling a compression relationship between U_k and U_{k-h} with proper step coefficient design.

Remark 4.2. This lemma also supports the view expressed in Remark 3.2, which states that the encoding rule ψ_k can be extended to satisfy $\|\psi_k\| \leq \bar{\psi} < \infty$ and $\frac{1}{h_\psi}\mathbb{E}\left[\sum_{l=k}^{k+h_\psi-1}\psi_l\psi_l^T\middle|\mathcal{F}_{k-1}\right] \geq \delta_\psi^2 I_{n\times n}$ for all $k \geq 1$. Its proof is similar as (23)–(24) and is omitted. Under this condition, Lemma 4.1 establishes the iterative relationship between U_k, U_{k-h_ψ} and V_{k-h_ψ}.

Next, we present the following lemma to illustrate how the iteration of the Lyapunov function V_k is influenced by the Lyapunov function U_k for fusion estimation error.

Lemma 4.2. *If Assumptions 1–2 and 4–6 hold, then the Lyapunov function V_k for the neighboring estimation error satisfies that*

$$V_k \leq \left(1-\beta\left(\frac{2\sigma}{c_h}-\frac{1}{\alpha_3}\right)\sum_{l=k-h}^{k-1}b_{l+1}\right)V_{k-h} + \alpha_3\beta\bar{N}\sum_{l=k-h}^{k-1}b_{l+1}U_{k-h}+O\left(b_k^2\right),$$

for all $\alpha_3 \in (0,\infty)$, where $\sigma = \frac{hf\hat{\lambda}_2\delta_\phi^2}{2f\bar{\phi}^2+h\hat{\lambda}_2}$, $\hat{\lambda}_2$ is the smallest nonzero eigenvalue of the Laplacian matrix $\hat{\mathcal{L}}_{\mathcal{C}}$ of the mirror graph for the union graph $\mathcal{G}_{\mathcal{C}}$, c_h, δ_ϕ, $\bar{\phi}$ and $\underline{f}$ are defined in Lemma 4.1, Assumptions 2 and 6.

Remark 4.3. Lemma 4.2 presents the iterative relationship between V_k, V_{k-h} and U_{k-h}, showing that with appropriate step coefficient design, V_k can be reduced relative to V_{k-h}, while still being perturbed by the coupling term U_{k-h}. Together with Remark 4.1, this implies that we cannot analyze the convergence of U_k or V_k independently; instead, we must jointly analyze both U_k and V_k with respect to their common coupling terms.

Remark 4.4. Similar with Remark 4.1, the conditional expectation-type cooperative excitation condition of the regressor $\phi_{k,i}$ guarantees the compression relationship between V_k and V_{k-h}. Specifically, this condition, combined with the connectivity of the mirror graph of the union graph $\mathcal{G}_{\mathcal{C}}$, guarantees the compression term $\frac{2\beta\sigma}{c_h}\sum_{l=k-h}^{k-1}b_{l+1}V_{k-h}$. This implies that even without transmitting the regressor information, the collective effect of the sensors can still be achieved solely through the quantized information exchanged between sensors. In other words, while each sensor cannot estimate the global information based solely on its local measurements, the exchange of one-bit information between sensors allows each sensor to reconstruct the global information.

Building on the above lemmas, the following theorem demonstrates the mean square convergence of the fusion estimate of the unknown parameter given by (6) and the estimate of the neighboring estimate given by (5).

Theorem 4.1. *If Assumptions 1–6 hold, then the proposed EFTQDI algorithm is convergent in the mean square sense, i.e.,*

$$\lim_{k\to\infty}\mathbb{E}\|\tilde{\Theta}_k\|^2 = 0 \text{ and } \lim_{k\to\infty}\mathbb{E}\|\varepsilon_k\|^2 = 0,$$

providing that $\gamma > \frac{\beta}{2\underline{g}\underline{\pi}\delta_\psi^2}\left(2c_h\bar{N}+\sigma\left(\bar{f}^2\bar{\phi}^4+\lambda_m^2\right)+\frac{8c_h^4\bar{N}^2}{3\sigma^3}\right)$, where $\underline{\pi}$, σ, $\hat{\lambda}_2$, $\bar{N}$, λ_m, $\underline{g}$, $\bar{f}$, δ_ψ are defined in Lemmas 4.1–4.2, respectively.

Remark 4.5. The core challenge of addressing both quantized measurements and quantized communication lies in handling the coupling between their respective estimates. This paper introduces two coupled Lyapunov functions (i.e., U_k and V_k) and analyzes their joint iterative dynamics (i.e., Lemmas 4.1 and 4.2), thereby establishing their mean-square convergence. For the mean-square convergence, the step-size coefficient γ of the estimate of neighboring estimate must be in a certain proportional relationship with β, the step-size for the fusion estimate. This is because the estimate for neighboring estimate $\hat{\theta}_{k,ij}$ essentially tracks the fusion estimate $\theta_{k,j}$ generated by its neighbors. When the step-size coefficient β of the fusion estimate increases, the step-size coefficient γ of the estimate for neighboring estimate $\hat{\theta}_{k,ij}$ must also be adjusted accordingly to effectively track the slow time-varying fusion estimate $\theta_{k,j}$.

Finally, the following theorem demonstrates the mean square convergence rate for the case where the step-size is of the polynomial form $b_k = \frac{1}{k^p}$, $p \in (1/2, 1]$.

Theorem 4.2. *Under the conditions of Theorem 4.1, the proposed EFTQDI algorithm has the following mean square convergence rate:*

- *For the case of the step-size as $b_k = \frac{1}{k^p}$ $\left(\frac{1}{2} < p < 1\right)$, $\mathbb{E}\|\tilde{\Theta}_k\|^2 = O\left(\frac{1}{k^p}\right)$ and $\mathbb{E}\|\varepsilon_k\|^2 = O\left(\frac{1}{k^p}\right)$.*
- *For the case of the step-size as $b_k = \frac{1}{k}$, it can been seen that $\mathbb{E}\|\tilde{\Theta}_k\|^2 = O\left(\frac{1}{k}\right)$ and $\mathbb{E}\|\varepsilon_k\|^2 = O\left(\frac{1}{k}\right)$ providing that $\beta > \frac{2c_h}{3\sigma}$ and $\gamma > \frac{\beta}{2\underline{g}\underline{\pi}\delta_\psi^2}\left(2c_h\bar{N}+\sigma\left(\bar{f}^2\bar{\phi}^4+\lambda_m^2\right)+\frac{8\beta c_h^4\bar{N}^2}{\sigma^2(3\beta\sigma-2c_h)}+\frac{c_h}{\beta}\right)$.*

Remark 4.6. It is important to noting that when the step size is $b_k = \frac{1}{k}$, achieving the convergence rate at the same order as itself requires additional constraints on the step-size coefficient γ and β, due to the unique properties of the harmonic series $\sum\frac{1}{k}$. Specifically, the harmonic series is highly sensitive, and its coefficient τ can affect the order of the product $\prod_{k=1}^{l}\left(1-\frac{\tau}{k}\right)$ while the product $\prod_{k=1}^{l}\left(1-\frac{\tau}{k^p}\right)$ for $\frac{1}{2} < p < 1$ is not similarly influenced. Specifically, to guarantee the convergence rate as $O(b_k)$ for the case of $b_k = \frac{1}{k}$, the step-size coefficient β of the fusion estimate is in inverse proportion to the connectivity $\hat{\lambda}_2$ of the mirror graph of the union graph $\mathcal{G}_{\mathcal{C}}$ and stochastic measurement noise coefficient f. And the step-size coefficient γ for the estimate of the neighboring estimate is inversely proportional to communication uncertainties including Markov-switching rate $\underline{\pi}$ and stochastic communication noise coefficient $\underline{g}$. Moreover, this observation highlights the influence of communication uncertainties on the convergence rate. Specifically, a larger Markov-switching rate $\underline{\pi}$ leads to a slower convergence rate, and a higher stochastic communication noise coefficient $\underline{g}$ also results in slower convergence rate.

5. Proofs of the main results

In this section, some lemmas are collected and established, which are frequently used in the analysis of convergence and convergence rate.

Lemma 5.1 (Calamai & Moré, 1987). *For the bounded convex set Ω, the projection is defined as $\Pi(x) = \arg\min_{z\in\Omega} \|x-z\|$ for all $x \in \mathbb{R}^n$. Then, for all $x \in \mathbb{R}$ and $x^* \in \Omega$, it holds $\|\Pi(x) - x^*\| \le \|x - x^*\|$.*

Lemma 5.2 (Ren & Beard, 2005). *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ be a directed graph with the Laplacian matrix $\mathcal{L}$. Then, $\frac{\mathcal{L}+\mathcal{L}^T}{2}$ is the Laplacian matrix of its mirror graph $\hat{\mathcal{G}}$ if and only if $\mathcal{G}$ is a balanced graph.*

Lemma 5.3 (Xie & Guo, 2018). *Let $A_k^i \in \mathbb{R}^{n\times n}$ with $0 \le A_k^i \le bI_n$, $A_k = \mathrm{diag}\{A_k^i\}_{i\in\{1,\dots,m\}}$, where $b > 0$ is a constant. If $\mathcal{L}$ is the Laplacian matrix of the undirect and connect topology, then for all $k \ge 0$ and $h \ge 1$,*

$$\lambda_{\min}\left(\sum_{l=k}^{k+h-1}[A_l + \nu\mathcal{L}]\right) \ge \frac{\nu\lambda_2}{m(2b+\nu\lambda_2)}\lambda_{\min}\left(\sum_{i=1}^{m}\sum_{l=k}^{k+h-1}A_l^i\right),$$

where $\nu > 0$, and λ_2 is the smallest nonzero eigenvalue of the Laplacian matrix $\mathcal{L}$.

Lemma 5.4 (Polyak, 1987). *Let $\{p_k\}$, $\{q_k\}$ and $\{\alpha_k\}$ be real sequences satisfying $p_{k+1} \le (1-q_k)p_k + \alpha_k$, where $0 < q_k \le 1$, $\sum_{k=0}^{\infty} q_k = \infty$, $\alpha_k \ge 0$ and $\lim_{k\to\infty}\frac{\alpha_k}{q_k} = 0$. Then, $\limsup_{k\to\infty} p_k \le 0$. Particularly, if $p_k \ge 0$, then $\lim_{k\to\infty} p_k = 0$.*

Lemma 5.5 (Wang et al., 2024). *For $0 < b \le 1$, $a > 0$, $k_0 \ge 0$ and sufficiently large l, we have*

$$\prod_{i=l}^{k}\left(1 - \frac{a}{(i+k_0)^b}\right) \le \begin{cases} \left(\frac{l+k_0}{k+k_0}\right)^a, & b = 1; \\ e^{\frac{a}{1-b}((l+k_0)^{1-b} - (k+k_0+1)^{1-b})}, & b \in (0,1); \end{cases}$$

$$\sum_{l=1}^{k}\prod_{i=l}^{k}\left(1 - \frac{a}{(i+k_0)^b}\right)O\left(\frac{1}{(l+k_0)^{2b}}\right) = O\left(\frac{1}{k^b}\right), b \in (0,1).$$

Lemma 5.6 (Zhang et al., 2021). *For any given positive integer l and $a, b \in \mathbb{R}$, the following results hold*

$$\sum_{l=1}^{k}\prod_{i=l+1}^{k}\left(1 - \frac{a}{i}\right)\frac{1}{l^{1+b}} = \begin{cases} O\left(\frac{1}{k^a}\right), a < b, \\ O\left(\frac{\ln k}{k^a}\right), a = b, \\ O\left(\frac{1}{k^b}\right), a > b. \end{cases}$$

Lemma 5.7. *Under Assumption 1, R_k, P_k and Q defined in Section 4.1 have the following properties:*
(i) $\mathbb{E}R_k \ge \underline{\pi}I_N + O(\lambda_q^k)$; (ii) $\|P_k\| \le \sqrt{\bar{N}}$; (iii) $\|Q\| \le \sqrt{\bar{N}}$.

Proof. From the definition of R_k and (3), we have

$$\begin{aligned}\mathbb{E}\bar{r}_{k,l} =& \mathbb{P}(r_l \in \mathcal{N}_i^k) = \mathbb{P}((i, r_l) \in \mathcal{E}_k) \\ =& \sum_{u\in G_{ir_l}}\pi_u + O(\lambda_q^k) \ge \underline{\pi} + O(\lambda_q^k),\end{aligned}$$

for any $l = N_1 + \cdots + N_{i-1} + 1, \dots, N_1 + \cdots + N_i$. Thus,

$$\begin{aligned}\mathbb{E}R_k =& \mathrm{diag}\{\mathbb{E}\bar{r}_{k,1}, \dots, \mathbb{E}\bar{r}_{k,N_1}, \dots, \mathbb{E}\bar{r}_{k,N_1+\cdots+N_{m-1}+1}, \\ & \dots, \mathbb{E}\bar{r}_{k,N_1+\cdots+N_m}\} \ge \underline{\pi}I_N + O(\lambda_q^k).\end{aligned}$$

From $a_{ij}^k \in [0,1)$, the definition of P_k and Q, we have

$$\begin{aligned}P_kP_k^T =& \mathrm{diag}\{p_{k,1}^Tp_{k,1}, \dots, p_{k,m}^Tp_{k,m}\} \\ =& \mathrm{diag}\left\{\sum_{l=1}^{N_1}(a_{1r_l}^k)^2, \dots, \sum_{l=N_1+\cdots+N_{m-1}+1}^{N_1+\cdots+N_m}(a_{mr_l}^k)^2\right\} \\ \le& \mathrm{diag}\{N_1, \dots, N_m\} \le \bar{N}I_m,\end{aligned}$$

$$Q^TQ = \sum_{i=1}^{m}\sum_{j\in\mathcal{N}_i^{\mathcal{C}}}q_{ij}q_{ij}^T = \mathrm{diag}\{N_1, \dots, N_m\} \le \bar{N}I_m.$$

Then, we have

$$\|P_k\| = \sqrt{\lambda_{\max}(P_k^TP_k)} = \sqrt{\lambda_{\max}(P_kP_k^T)} \le \sqrt{\bar{N}},$$

$$\|Q\| = \sqrt{\lambda_{\max}(Q^TQ)} \le \sqrt{\bar{N}}.$$

This completes the proof. □

Lemma 5.8. *Under Assumptions 2, 3 and 5, U_k, V_k and ε_k satisfies:*
(i) $\|U_{k+l} - U_k\| = O(b_{k+l})$ for all $l \ge 0$;
(ii) $\|V_{k+l} - V_k\| = O(b_{k+l})$ for all $l \ge 0$;
(iii) $\|\varepsilon_{k+l} - \varepsilon_k\| = O(b_{k+l})$ for all $l \ge 0$.

Proof. From Assumptions 2–3, (5) and (6), we have

$$\|\theta_{k,i}\| \le \bar{\theta}, \quad \|\hat{\theta}_{k,ij}\| \le \bar{\theta}, \tag{10}$$

$$\|\phi_{k,i}\| \le \bar{\phi}, \quad \|\psi_k\| \le \bar{\psi}, \tag{11}$$

for all $i = 1, \dots, m$ and $j \in \mathcal{N}_i^{\mathcal{C}}$. We now prove

$$\left\|\tilde{\theta}_{k+l,i} - \tilde{\theta}_{k,i}\right\| \le \beta(\bar{\phi} + 2N_i\bar{\theta})\sum_{p=k+1}^{k+l}b_p, \tag{12}$$

$$\left\|\varepsilon_{k+l,ij} - \varepsilon_{k,ij}\right\| \le (\gamma\bar{\psi} + \beta\bar{\phi} + 2\beta N_i\bar{\theta})\sum_{p=k+1}^{k+l}b_p, \tag{13}$$

for all $l \ge 0$, $i = 1, \dots, m$ and $j \in \mathcal{N}_i^{\mathcal{C}}$. From Lemma 5.1, (6), (10) and (11), we have

$$\begin{aligned}&\left\|\tilde{\theta}_{k,i} - \tilde{\theta}_{k-1,i}\right\| \\ =& \Bigg\|\Pi_\Omega\Bigg\{\theta_{k-1,i} + \beta b_k\big(\phi_{k,i}(F_{k,i}(C_i - \phi_{k,i}^T\theta_{k-1,i}) - s_{k,i}) \\ &+ \sum_{j\in\mathcal{N}_i^k}a_{ij}^k(\hat{\theta}_{k-1,ij} - \theta_{k-1,i})\big)\Bigg\} - \theta_{k-1,i}\Bigg\| \\ \le& \Bigg\|\beta b_k\big(\phi_{k,i}(F_{k,i}(C_i - \phi_{k,i}^T\theta_{k-1,i}) - s_{k,i}) \\ &+ \sum_{j\in\mathcal{N}_i^k}a_{ij}^k(\hat{\theta}_{k-1,ij} - \theta_{k-1,i})\big)\Bigg\| \le \beta(\bar{\phi} + 2N_i\bar{\theta})b_k,\end{aligned}$$

and hence

$$\begin{aligned}\left\|\tilde{\theta}_{k+l,i} - \tilde{\theta}_{k,i}\right\| \le& \sum_{p=k+1}^{k+l}\|\tilde{\theta}_{p,i} - \tilde{\theta}_{p-1,i}\| \\ \le& \beta(\bar{\phi} + 2N_i\bar{\theta})\sum_{p=k+1}^{k+l}b_p.\end{aligned}$$

That is, (12) hold. By Lemma 5.1, (5) and (10)–(12), we have

$$\begin{aligned}&\left\|\varepsilon_{k,ij} - \varepsilon_{k-1,ij}\right\| \\ =& \left\|\left(\hat{\theta}_{k-1,ij} - \theta_{k,i}\right) - \left(\hat{\theta}_{k-1,ij} - \theta_{k-1,i}\right)\right\| \\ \le& \left\|\theta_{k,i} - \theta_{k-1,i}\right\| \\ \le& \beta(\bar{\phi} + 2N_i\bar{\theta})b_k, \text{ for } j \in \mathcal{N}_i^{\mathcal{C}} \setminus \mathcal{N}_i^k, \\ &\left\|\varepsilon_{k,ij} - \varepsilon_{k-1,ij}\right\| \\ =& \Bigg\|\left(\Pi_\Omega\left\{\hat{\theta}_{k-1,ij} + \gamma b_k\psi_k\left(G_{k,ij}(C_{ij} - \psi_k^T\hat{\theta}_{k-1,ij}) - z_{k,ij}\right)\right\}\right. \\ &\left. - \theta_{k,i}\right) - \left(\hat{\theta}_{k-1,ij} - \theta_{k-1,i}\right)\Bigg\|\end{aligned}$$

$$\leq \left\| \gamma b_k \psi_k \left(G_{k,ij}(C_{ij} - \psi_k^T \hat{\theta}_{k-1,ij}) - z_{k,ij} \right) \right\| + \left\| \theta_{k,i} - \theta_{k-1,i} \right\|$$
$$\leq (\gamma \bar{\psi} + \beta \bar{\phi} + 2\beta N_i \bar{\theta}) b_k, \text{ for } j \in \mathcal{N}_i^k,$$

and hence,

$$\left\| \varepsilon_{k+l,ij} - \varepsilon_{k,ij} \right\| \leq \sum_{p=k+1}^{k+l} \left\| \varepsilon_{p,ij} - \varepsilon_{p-1,ij} \right\|$$
$$\leq (\gamma \bar{\psi} + \beta \bar{\phi} + 2\beta N_i \bar{\theta}) \sum_{p=k+1}^{k+l} b_p.$$

That is, (13) holds. Then, from (9)–(13) and Assumption 5, we have

$$\|U_{k+l} - U_k\| = \left\| \mathbb{E}[(\varepsilon_{k+l} - \varepsilon_k)^T (\varepsilon_{k+l} + \varepsilon_k)] \right\| = O(b_{k+l}),$$
$$\|V_{k+l} - V_k\| = \left\| \mathbb{E}[(\tilde{\Theta}_{k+l} - \tilde{\Theta}_k)^T (\tilde{\Theta}_{k+l} + \tilde{\Theta}_k)] \right\| = O(b_{k+l}),$$
$$\|\varepsilon_{k+l} - \varepsilon_k\| = O(b_{k+l}),$$

for all $l \geq 0$. □

5.1. Proof of Lemma 4.1

At first, we consider $\varepsilon_{k,ij}$ for the case of $j \in \mathcal{N}_i^k$. From (5), (6) and Lemma 5.1, we have

$$\begin{aligned}
&\varepsilon_{k,ij}^T \varepsilon_{k,ij} \\
\leq & \varepsilon_{k-1,ij}^T \varepsilon_{k-1,ij} + 2\gamma b_k \varepsilon_{k-1,ij}^T \psi_k (\hat{G}_{k,ij} - z_{k,ij}) \\
& - 2\beta b_k \varepsilon_{k-1,ij}^T \phi_{k,j} (\hat{F}_{k,j} - s_{k,j}) + \gamma^2 b_k^2 \psi_k^T \psi_k (\hat{G}_{k,ij} - z_{k,ij})^2 \\
& - 2\beta b_k \varepsilon_{k-1,ij}^T \sum_{l \in \mathcal{N}_j^k} a_{jl}^k (\hat{\theta}_{k-1,jl} - \theta_{k-1,j}) \\
& - 2\gamma \beta b_k^2 \psi_k^T \phi_{k,j} (\hat{G}_{k,ij} - z_{k,ij}) \cdot (\hat{F}_{k,j} - s_{k,j}) \\
& - 2\gamma \beta b_k^2 (\hat{G}_{k,ij} - z_{k,ij}) \psi_k^T \sum_{l \in \mathcal{N}_j^k} a_{jl}^k (\hat{\theta}_{k-1,jl} - \theta_{k-1,j}) \\
& + 2\beta^2 b_k^2 (\hat{F}_{k,j} - s_{k,j}) \phi_{k,i}^T \sum_{l \in \mathcal{N}_j^k} a_{jl}^k (\hat{\theta}_{k-1,jl} - \theta_{k-1,j}) \\
& + \beta^2 b_k^2 \sum_{l \in \mathcal{N}_j^k} a_{jl}^k (\hat{\theta}_{k-1,jl} - \theta_{k-1,j})^T \sum_{l \in \mathcal{N}_j^k} a_{jl}^k \\
& \cdot (\hat{\theta}_{k-1,jl} - \theta_{k-1,j}) + \beta^2 b_k^2 \phi_{k,j}^T \phi_{k,j} (\hat{F}_{k,j} - s_{k,j})^2 \\
= & \varepsilon_{k-1,ij}^T \varepsilon_{k-1,ij} + 2\gamma b_k \varepsilon_{k-1,ij}^T \psi_k (\hat{G}_{k,ij} - z_{k,ij}) \\
& - 2\beta b_k \varepsilon_{k-1,ij}^T \sum_{l \in \mathcal{N}_j^k} a_{jl}^k (\hat{\theta}_{k-1,jl} - \theta_{k-1,j}) \\
& - 2\beta b_k \varepsilon_{k-1,ij}^T \phi_{k,j} (\hat{F}_{k,j} - s_{k,j}) + O\left(b_k^2\right),
\end{aligned} \tag{14}$$

where the last equality is got by (10)–(11) and the boundedness of $\hat{F}_{k,i}$ and $\hat{G}_{k,ij}$, which is got because $F_{k,i}(\cdot)$ and $G_{k,ij}(\cdot)$ are continuous functions.

By Assumption 4 and differential mean value theorem,

$$\begin{aligned}
&\mathbb{E}[2\gamma b_k \varepsilon_{k-1,ij}^T \psi_k (\hat{G}_{k,ij} - z_{k,ij}) | \mathcal{F}_{k-1}] \\
= & 2\gamma b_k \varepsilon_{k-1,ij}^T \psi_k \left(\hat{G}_{k,ij} - \mathbb{E}[z_{k,ij} | \mathcal{F}_{k-1}] \right) \\
= & 2\gamma b_k \varepsilon_{k-1,ij}^T \psi_k \big(G_{k,ij}(C_{ij} - \psi_k^T \hat{\theta}_{k-1,ij}) \\
& - G_{k,ij} \left(C_{ij} - \psi_k^T \theta_{k-1,j} \right) \big) \\
= & -2\gamma g_{k,ij} \left(\xi_{k,ij} \right) b_k \varepsilon_{k-1,ij}^T \psi_k \psi_k^T \varepsilon_{k-1,ij} \\
\leq & -2\gamma \underline{g} b_k \varepsilon_{k-1,ij}^T \psi_k \psi_k^T \varepsilon_{k-1,ij}.
\end{aligned} \tag{15}$$

where $\xi_{k,ij}$ is in the interval between $C_{ij} - \psi_k^T \hat{\theta}_{k-1,ij}$ and $C_{ij} - \psi_k^T \theta_{k-1,j}$ such that $G_{k,ij} \left(C_{ij} - \psi_k^T \hat{\theta}_{k-1,ij} \right) - G_{k,ij} \left(C_{ij} - \psi_k^T \theta_{k-1,j} \right) = -g_{k,ij} \left(\xi_{k,ij} \right) \psi_k^T \varepsilon_{k,ij}$.

By the differential mean value theorem, Cauchy–Schwarz inequality and Assumptions 2–4, we have

$$\begin{aligned}
& - \mathbb{E}[2\beta b_k \varepsilon_{k-1,ij}^T \phi_{k,j} (\hat{F}_{k,j} - s_{k,j}) | \mathcal{F}_{k-1}] \\
= & -2\beta b_k \varepsilon_{k-1,ij}^T \phi_{k,j} \left(\hat{F}_{k,j} - \mathbb{E}[s_{k,j} | \mathcal{F}_{k-1}] \right) \\
= & -2\beta b_k \varepsilon_{k-1,ij}^T \phi_{k,j} \left(F_{k,j}(C_j - \phi_{k,j}^T \theta_{k-1,j}) - F_{k,j}(C_j - \phi_{k,j}^T \theta) \right) \\
= & 2\beta b_k \varepsilon_{k-1,ij}^T \phi_{k,j} \phi_{k,j}^T \tilde{\theta}_{k-1,j} f_{k,j} \left(\zeta_{k,j} \right) \\
\leq & 2\beta \bar{f} \bar{\phi}^2 b_k \|\varepsilon_{k-1,ij}\| \|\tilde{\theta}_{k-1,j}\| \\
\leq & \beta \bar{f} \bar{\phi}^2 b_k \left(\frac{1}{\alpha_1} \|\varepsilon_{k-1,ij}\|^2 + \alpha_1 \|\tilde{\theta}_{k-1,j}\|^2 \right),
\end{aligned} \tag{16}$$

where $\zeta_{k,j}$ is in the interval between $C_j - \phi_{k,j}^T \theta_{k-1,j}$ and $C_j - \phi_{k,j}^T \theta$ such that $F_{k,j}(C_j - \phi_{k,j}^T \theta_{k-1,j}) - F_{k,j}(C_j - \phi_{k,j}^T \theta) = -f_{k,j} \left(\zeta_{k,j} \right) \phi_{k,j}^T \tilde{\theta}_{k-1,j}$, and $\alpha_1 \in (0, \infty)$.

Substituting (15) and (16) into (14) and calculating its mathematical expectation, we can obtain for $j \in \mathcal{N}_i^k$

$$\begin{aligned}
& \mathbb{E}[\varepsilon_{k,ij}^T \varepsilon_{k,ij}] \\
\leq & \mathbb{E}[\varepsilon_{k-1,ij}^T \varepsilon_{k-1,ij}] + 2\gamma b_k \mathbb{E}[\varepsilon_{k-1,ij}^T \psi_k (\hat{G}_{k,ij} - z_{k,ij})] \\
& - 2\beta b_k \mathbb{E}[\varepsilon_{k-1,ij}^T \sum_{l \in \mathcal{N}_j^k} a_{jl}^k (\hat{\theta}_{k-1,jl} - \theta_{k-1,j})] \\
& - 2\beta b_k \mathbb{E}[\varepsilon_{k-1,ij}^T \phi_{k,j} (\hat{F}_{k,j} - s_{k,j})] + O\left(b_k^2\right) \\
\leq & \mathbb{E}[\varepsilon_{k-1,ij}^T \varepsilon_{k-1,ij}] - 2\gamma \underline{g} b_k \mathbb{E}[\varepsilon_{k-1,ij}^T \psi_k \psi_k^T \varepsilon_{k-1,ij}] \\
& + \frac{\beta \bar{f} \bar{\phi}^2}{\alpha_1} b_k \mathbb{E}\|\varepsilon_{k-1,ij}\|^2 + \alpha_1 \beta \bar{f} \bar{\phi}^2 \mathbb{E}\|\tilde{\theta}_{k-1,j}\|^2 \\
& - 2\beta b_k \mathbb{E}[\varepsilon_{k-1,ij}^T \sum_{l \in \mathcal{N}_j^k} a_{jl}^k (\theta_{k-1,l} - \theta_{k-1,j})] \\
& - 2\beta b_k \mathbb{E}[\varepsilon_{k-1,ij}^T \sum_{l \in \mathcal{N}_j^k} a_{jl}^k \varepsilon_{k,jl}] + O\left(b_k^2\right).
\end{aligned} \tag{17}$$

Next, we consider $\varepsilon_{k,ij}$ for the case of $j \in \mathcal{N}_i^{\mathcal{C}} \setminus \mathcal{N}_i^k$. By (5), (6) and Lemma 5.1, we can obtain

$$\begin{aligned}
& \varepsilon_{k,ij}^T \varepsilon_{k,ij} \\
\leq & \big(\varepsilon_{k-1,ij} - \beta b_k \phi_{k,j} (\hat{F}_{k,j} - s_{k,j}) \\
& - \beta b_k \sum_{l \in \mathcal{N}_j^k} a_{jl}^k (\hat{\theta}_{k-1,jl} - \theta_{k-1,j}) \big)^T \\
& \cdot \big(\varepsilon_{k-1,ij} - \beta b_k \phi_{k,j} (\hat{F}_{k,j} - s_{k,j}) \\
& - \beta b_k \sum_{l \in \mathcal{N}_j^k} a_{jl}^k (\hat{\theta}_{k-1,jl} - \theta_{k-1,j}) \big) \\
= & \varepsilon_{k-1,ij}^T \varepsilon_{k-1,ij} - 2\beta b_k \varepsilon_{k-1,ij}^T \phi_{k,j} (\hat{F}_{k,j} - s_{k,j}) + O\left(b_k^2\right) \\
& - 2\beta b_k \varepsilon_{k-1,ij}^T \sum_{l \in \mathcal{N}_j^k} a_{jl}^k (\hat{\theta}_{k-1,jl} - \theta_{k-1,j}),
\end{aligned} \tag{18}$$

where the last equality is got similarly to (14) based on (10)–(11) and the boundedness of $\hat{F}_{k,i}$.

Taking (16) into (18) and calculating its mathematical expectation, we can get for $j \in \mathcal{N}_i^{\mathcal{C}} \setminus \mathcal{N}_i^k$

$$\begin{aligned}
& \mathbb{E}[\varepsilon_{k,ij}^T \varepsilon_{k,ij}] \\
\leq & \mathbb{E}[\varepsilon_{k-1,ij}^T \varepsilon_{k-1,ij}] + \beta \bar{f} \bar{\phi}^2 b_k \mathbb{E}\|\varepsilon_{k-1,ij}\|^2 / \alpha_1 + O\left(b_k^2\right)
\end{aligned}$$

$$+\alpha_1\beta\bar{f}\bar{\phi}^2\mathbb{E}\|\tilde{\theta}_{k-1,j}\|^2-2\beta b_k\mathbb{E}[\varepsilon_{k-1,ij}^T\sum_{l\in\mathcal{N}_j^k}a_{jl}^k\varepsilon_{k,jl}]$$
$$-2\beta b_k\mathbb{E}[\varepsilon_{k-1,ij}^T\sum_{l\in\mathcal{N}_j^k}a_{jl}^k(\theta_{k-1,l}-\theta_{k-1,j})]. \tag{19}$$

By (8), (9), (17), (19) and the notations in Section 4.1,

$$\begin{aligned}U_k\leq&U_{k-1}-2\gamma\underline{g}b_k\mathbb{E}[\varepsilon_{k-1}^T\Psi_kR_k\Psi_k^T\varepsilon_{k-1}]+\frac{\beta\bar{f}\bar{\phi}^2}{\alpha_1}b_k\mathbb{E}\|\varepsilon_{k-1}\|^2\\&+\alpha_1\beta\bar{f}\bar{\phi}^2b_k\mathbb{E}[\tilde{\Theta}_{k-1}^TQ^T\otimes I_nQ\otimes I_n\tilde{\Theta}_{k-1}]\\&-2\beta b_k\mathbb{E}[\varepsilon_{k-1}^TQ\otimes I_nP_k\otimes I_n\varepsilon_{k-1}]\\&-2\beta b_k\mathbb{E}[\varepsilon_{k-1}^TQ\otimes I_n\mathcal{L}_k\otimes I_n\tilde{\Theta}_{k-1}]+O\left(b_k^2\right)\\\leq&\left(1+\beta\bar{f}\bar{\phi}^2b_k/\alpha_1\right)U_{k-1}-2\gamma\underline{g}b_k\mathbb{E}[\varepsilon_{k-1}^T\Psi_kR_k\Psi_k^T\varepsilon_{k-1}]\\&-2\beta b_k\mathbb{E}[\varepsilon_{k-1}^TQP_k\otimes I_n\varepsilon_{k-1}]+\frac{\beta}{\alpha_2}b_k\mathbb{E}[\varepsilon_{k-1}^TQQ^T\otimes I_n\varepsilon_{k-1}]\\&+\alpha_1\beta\bar{f}\bar{\phi}^2b_k\mathbb{E}[\tilde{\Theta}_{k-1}^TQ^TQ\otimes I_n\tilde{\Theta}_{k-1}]\\&+\beta\alpha_2b_k\mathbb{E}[\tilde{\Theta}_{k-1}\mathcal{L}_k^T\mathcal{L}_k\otimes I_n\tilde{\Theta}_{k-1}]+O\left(b_k^2\right).\end{aligned}\tag{20}$$

Based on Lemma 5.7, the second term in the right side of (20) can be estimated as

$$\begin{aligned}&-2\gamma\underline{g}b_k\mathbb{E}[\varepsilon_{k-1}^T\Psi_kR_k\Psi_k^T\varepsilon_{k-1}]\\=&-2\gamma\underline{g}b_k\mathbb{E}[\mathbb{E}[\varepsilon_{k-1}^T\Psi_kR_k\Psi_k^T\varepsilon_{k-1}|\mathcal{F}_{k-1}]]\\=&-2\gamma\underline{g}b_k\mathbb{E}[\varepsilon_{k-1}^T\Psi_k\mathbb{E}[R_k]\Psi_k^T\varepsilon_{k-1}]\\\leq&-2\gamma\underline{g}\underline{\pi}b_k\mathbb{E}[\varepsilon_{k-1}^T\Psi_k\Psi_k^T\varepsilon_{k-1}]+O(\lambda_q^k),\end{aligned}\tag{21}$$

where the last inequality is got by the boundedness of Ψ_k and ε_{k-1}. From Assumption 1, we learn $\mathcal{L}_k\in\{\mathcal{L}^{(1)},\ldots,\mathcal{L}^{(s)}\}$ and

$$\|\mathcal{L}_k\|\leq\max_{1\leq u\leq s}\|\mathcal{L}^{(u)}\|\triangleq\lambda_m. \tag{22}$$

From Lemmas 5.7, taking (21) and (22) into (20) yields

$$\begin{aligned}U_k\leq&\left(1+\beta\left(\bar{f}\bar{\phi}^2/\alpha_1+2\bar{N}+\bar{N}/\alpha_2\right)b_k\right)U_{k-1}\\&-2\gamma\underline{g}\underline{\pi}b_k\mathbb{E}[\varepsilon_{k-1}^T\Psi_k\Psi_k^T\varepsilon_{k-1}]\\&+\beta(\alpha_1\bar{f}\bar{N}\bar{\phi}^2+\alpha_2\lambda_m^2)b_kV_{k-1}+O\left(b_k^2\right)\\\leq&U_{k-h}+\beta\left(\frac{\bar{f}\bar{\phi}^2}{\alpha_1}+2\bar{N}+\frac{\bar{N}}{\alpha_2}\right)\sum_{l=k-h}^{k-1}b_{l+1}U_l\\&-2\gamma\underline{g}\underline{\pi}\sum_{l=k-h}^{k-1}\mathbb{E}[\varepsilon_l^Tb_{l+1}\Psi_{l+1}\Psi_{l+1}^T\varepsilon_l]+O\left(b_k^2\right)\\&+\beta(\alpha_1\bar{f}\bar{N}\bar{\phi}^2+\alpha_2\lambda_m^2)\sum_{l=k-h}^{k-1}b_{l+1}V_l.\end{aligned}\tag{23}$$

By Remark 3.4, there exists c_h such that $c_h\triangleq\sup_{|p-q|<h}\frac{b_p}{b_q}<\infty$. So we have $b_{l+1}\geq\frac{1}{hc_h}\sum_{l=k-h}^{k-1}b_{l+1}$. Then, based on the encoding rule in Algorithm 1 and Assumption 5, we have

$$\begin{aligned}\sum_{l=k-h}^{k-1}b_{l+1}\Psi_{l+1}\Psi_{l+1}^T=&\frac{1}{hc_h}\sum_{l=k-h}^{k-1}b_{l+1}\sum_{l=k-h}^{k-1}\Psi_{l+1}\Psi_{l+1}^T\\\geq&\frac{\delta_\psi^2}{c_h}\sum_{l=k-h}^{k-1}b_{l+1}I_{nN}.\end{aligned}\tag{24}$$

From Lemma 5.8 and (23)–(24), we have

$$\begin{aligned}U_k\leq&\left(1+\beta\left(\frac{\bar{f}\bar{\phi}^2}{\alpha_1}+2\bar{N}+\frac{\bar{N}}{\alpha_2}\right)\sum_{l=k-h}^{k-1}b_{l+1}\right)U_{k-h}\\&-2\gamma\underline{g}\underline{\pi}\mathbb{E}[\varepsilon_{k-h}^T(\sum_{l=k-h}^{k-1}b_{l+1}\Psi_{l+1}\Psi_{l+1}^T)\varepsilon_{k-h}]\\&+\beta(\alpha_1\bar{f}\bar{N}\bar{\phi}^2+\alpha_2\lambda_m^2)\sum_{l=k-h}^{k-1}b_{l+1}V_{k-h}+O\left(b_k^2\right)\\\leq&\left(1-\left(\frac{2\gamma\underline{g}\underline{\pi}\delta_\psi^2}{c_h}-\beta\left(\frac{\bar{f}\bar{\phi}^2}{\alpha_1}+2\bar{N}+\frac{\bar{N}}{\alpha_2}\right)\right)\sum_{l=k-h}^{k-1}b_{l+1}\right)U_{k-h}\\&+\beta(\alpha_1\bar{f}\bar{N}\bar{\phi}^2+\alpha_2\lambda_m^2)\sum_{l=k-h}^{k-1}b_{l+1}V_{k-h}+O\left(b_k^2\right).\end{aligned}$$

This completes the proof. □

5.2. Proof of Lemma 4.2

From (5), (10)–(11), Lemma 5.1 and the boundedness of $\hat{F}_{k,i}$, we have

$$\begin{aligned}&\tilde{\theta}_{k,i}^T\tilde{\theta}_{k,i}\\\leq&\tilde{\theta}_{k-1,i}^T\tilde{\theta}_{k-1,i}+2\beta b_k\tilde{\theta}_{k-1,i}^T\phi_{k,i}(\hat{F}_{k,i}-s_{k,i})\\&+2\beta b_k\tilde{\theta}_{k-1,i}^T\sum_{j\in\mathcal{N}_i^k}a_{ij}^k(\hat{\theta}_{k-1,ij}-\theta_{k-1,i})\\&+2\beta^2b_k^2(\hat{F}_{k,i}-s_{k,i})\phi_{k,i}^T\sum_{j\in\mathcal{N}_i^k}a_{ij}^k(\hat{\theta}_{k-1,ij}-\theta_{k-1,i})\\&+\beta^2b_k^2\Big(\sum_{j\in\mathcal{N}_i^k}a_{ij}^k(\hat{\theta}_{k-1,ij}-\theta_{k-1,i})\Big)^T\sum_{j\in\mathcal{N}_i^k}a_{ij}^k\\&\cdot(\hat{\theta}_{k-1,ij}-\theta_{k-1,i})+\beta^2b_k^2\phi_{k,i}^T\phi_{k,i}(\hat{F}_{k,i}-s_{k,i})^2\\\leq&\tilde{\theta}_{k-1,i}^T\tilde{\theta}_{k-1,i}+2\beta b_k\tilde{\theta}_{k-1,i}^T\phi_{k,i}(\hat{F}_{k,i}-s_{k,i})+O(b_k^2)\\&+2\beta b_k\tilde{\theta}_{k-1,i}^T\sum_{j\in\mathcal{N}_i^k}a_{ij}^k(\hat{\theta}_{k-1,ij}-\theta_{k-1,i}).\end{aligned}\tag{25}$$

By Assumption 4 and differential mean value theorem,

$$\begin{aligned}&\mathbb{E}[2\beta b_k\tilde{\theta}_{k-1,i}^T\phi_{k,i}(\hat{F}_{k,i}-s_{k,i})|\mathcal{F}_{k-1}]\\=&2\beta b_k\tilde{\theta}_{k-1,i}^T\phi_{k,i}\left(\hat{F}_{k,i}-\mathbb{E}[s_{k,i}|\mathcal{F}_{k-1}]\right)\\=&2\beta b_k\tilde{\theta}_{k-1,i}^T\phi_{k,i}\left(F_{k,i}(C_i-\phi_{k,i}^T\theta_{k-1,i})-F_{k,i}(C_i-\phi_{k,i}^T\theta)\right)\\=&-2\beta f_{k,i}\left(\zeta_{k,i}\right)b_k\tilde{\theta}_{k-1,i}^T\phi_{k,i}\phi_{k,i}^T\tilde{\theta}_{k-1,i}\\\leq&-2\beta\underline{f}b_k\tilde{\theta}_{k-1,i}^T\phi_{k,i}\phi_{k,i}^T\tilde{\theta}_{k-1,i},\end{aligned}\tag{26}$$

where $\zeta_{k,i}$ is defined in (16).

Substituting (26) into (25) and calculating its mathematical expectation, we can obtain

$$\begin{aligned}&\mathbb{E}[\tilde{\theta}_{k,i}^T\tilde{\theta}_{k,i}]\\\leq&\mathbb{E}[\tilde{\theta}_{k-1,i}^T\tilde{\theta}_{k-1,i}]-2\beta b_k\underline{f}\mathbb{E}[\tilde{\theta}_{k-1,i}^T\phi_{k,i}\phi_{k,i}^T\tilde{\theta}_{k-1,i}]\\&+2\beta b_k\mathbb{E}[\tilde{\theta}_{k-1,i}^T\sum_{j\in\mathcal{N}_i^k}a_{ij}^k(\hat{\theta}_{k-1,ij}-\theta_{k-1,i})]+O(b_k^2)\\\leq&\mathbb{E}[\tilde{\theta}_{k-1,i}^T\tilde{\theta}_{k-1,i}]-2\beta b_k\underline{f}\mathbb{E}[\tilde{\theta}_{k-1,i}^T\phi_{k,i}\phi_{k,i}^T\tilde{\theta}_{k-1,i}]\\&+2\beta b_k\mathbb{E}[\tilde{\theta}_{k-1,i}^T\sum_{j\in\mathcal{N}_i^k}a_{ij}^k(\tilde{\theta}_{k-1,j}-\tilde{\theta}_{k-1,i})]\\&+2\beta b_k\mathbb{E}[\tilde{\theta}_{k-1,i}^T\sum_{j\in\mathcal{N}_i^k}a_{ij}^k\varepsilon_{k-1,ij}]+O(b_k^2).\end{aligned}\tag{27}$$

Based on (9), (27), Lemma 5.7 and the notations in Section 4.1, we have

$$\begin{aligned}V_k=&\mathbb{E}[\tilde{\Theta}_k^T\tilde{\Theta}_k]\leq V_{k-1}-2\beta b_k\underline{f}\mathbb{E}[\tilde{\Theta}_{k-1}^T\Phi_k\Phi_k^T\tilde{\Theta}_{k-1}]\\&+2\beta b_k\mathbb{E}[\tilde{\Theta}_{k-1}^TP_k\otimes I_n\varepsilon_{k-1}]\end{aligned}$$

$$
\begin{aligned}
&- 2\beta b_k \mathbb{E}[\tilde{\Theta}_{k-1}^T \mathcal{L}_k \otimes I_n \tilde{\Theta}_{k-1}] + O(b_k^2) \\
\leq & V_{k-1} - 2\beta b_k \underline{f} \mathbb{E}[\tilde{\Theta}_{k-1}^T \Phi_k \Phi_k^T \tilde{\Theta}_{k-1}] \\
&+ 2\beta b_k \mathbb{E}[\tilde{\Theta}_{k-1}^T \mathcal{L}_k \otimes I_n \tilde{\Theta}_{k-1}] + O(b_k^2) \\
&+ \beta/\alpha_3 b_k \mathbb{E}[\tilde{\Theta}_{k-1}^T \tilde{\Theta}_{k-1}] + \alpha_3 \beta \bar{N} b_k U_{k-1} \\
\leq & V_{k-h} + \frac{\beta}{\alpha_3} \sum_{l=k-h}^{k-1} b_{l+1} V_l + \alpha_3 \beta \bar{N} \sum_{l=k-h}^{k-1} b_{l+1} U_l \\
&- 2\beta \sum_{l=k-h}^{k-1} b_{l+1} \mathbb{E}[\tilde{\Theta}_l^T \mathcal{L}_{l+1} \otimes I_n \tilde{\Theta}_l] + O(b_k^2) \\
&- 2\beta \underline{f} \sum_{l=k-h}^{k-1} b_{l+1} \mathbb{E}[\tilde{\Theta}_l^T \Phi_{l+1} \Phi_{l+1}^T \tilde{\Theta}_l]. \qquad (28)
\end{aligned}
$$

By Remark 3.3, we have

$$
\mathbb{E}\mathcal{L}_k = \sum_{u=1}^{s} \pi^{(u)} \mathcal{L}^{(u)} + O(\lambda_a^k) = \mathcal{L}_{\mathcal{C}} + O(\lambda_a^k). \qquad (29)
$$

From Assumption 5, Lemma 5.8 and (28), we have

$$
\begin{aligned}
V_k \leq & V_{k-h} + \frac{\beta}{\alpha_3} \sum_{l=k-h}^{k-1} b_{l+1} V_{k-h} + \frac{\beta}{\alpha_3} \sum_{l=k-h}^{k-1} b_{l+1} (V_l - V_{k-h}) \\
&- 2\beta \underline{f} \mathbb{E}[\tilde{\Theta}_{k-h}^T \sum_{l=k-h}^{k-1} b_{l+1} \Phi_{l+1} \Phi_{l+1}^T \tilde{\Theta}_{k-h}] \\
&- 2\beta \underline{f} \sum_{l=k-h}^{k-1} b_{l+1} \mathbb{E}[\tilde{\Theta}_l^T \Phi_{l+1} \Phi_{l+1}^T (\tilde{\Theta}_l - \tilde{\Theta}_{k-h})] \\
&- 2\beta \mathbb{E}[\tilde{\Theta}_{k-h}^T \sum_{l=k-h}^{k-1} b_{l+1} \mathcal{L}_{l+1} \otimes I_n \tilde{\Theta}_{k-h}] + O(b_k^2) \\
&- 2\beta \sum_{l=k-h}^{k-1} b_{l+1} \mathbb{E}[\tilde{\Theta}_l^T \mathcal{L}_{l+1} \otimes I_n (\tilde{\Theta}_l - \tilde{\Theta}_{k-h})] \\
&+ \alpha_3 \beta \bar{N} \sum_{l=k-h}^{k-1} b_{l+1} U_{k-h} + \alpha_3 \beta \bar{N} \sum_{l=k-h}^{k-1} b_{l+1} (U_l - U_{k-h}) \\
\leq & V_{k-h} + \frac{\beta}{\alpha_3} \sum_{l=k-h}^{k-1} b_{l+1} V_{k-h} \\
&- \frac{2\beta \underline{f}}{h c_h} \sum_{l=k-h}^{k-1} b_{l+1} \mathbb{E}[\tilde{\Theta}_{k-h}^T \sum_{l=k-h}^{k-1} \Phi_{l+1} \Phi_{l+1}^T \tilde{\Theta}_{k-h}] \\
&- 2\beta \mathbb{E}[\tilde{\Theta}_{k-h}^T \sum_{l=k-h}^{k-1} b_{l+1} \mathcal{L}_{l+1} \otimes I_n \tilde{\Theta}_{k-h}] \\
&+ \alpha_3 \beta \bar{N} \sum_{l=k-h}^{k-1} b_{l+1} U_{k-h} + O(b_k^2). \qquad (30)
\end{aligned}
$$

Then from Assumption 5, Lemma 5.8, and (28)–(29),

$$
\begin{aligned}
& 2\beta \mathbb{E}[\tilde{\Theta}_{k-h}^T \sum_{l=k-h}^{k-1} b_{l+1} \mathcal{L}_{l+1} \otimes I_n \tilde{\Theta}_{k-h}] \\
= & 2\beta \mathbb{E}\left[\tilde{\Theta}_{k-h}^T \sum_{l=k-h}^{k-1} b_{l+1} \mathbb{E}[\mathcal{L}_{l+1} \otimes I_n | \mathcal{F}_{k-h}] \tilde{\Theta}_{k-h}\right] \\
= & 2\beta \mathbb{E}\left[\tilde{\Theta}_{k-h}^T \sum_{l=k-h}^{k-1} b_{l+1} \mathbb{E}[\mathcal{L}_{l+1} \otimes I_n] \tilde{\Theta}_{k-h}\right] \\
= & 2\beta \sum_{l=k-h}^{k-1} b_{l+1} \mathbb{E}[\tilde{\Theta}_{k-h}^T \mathcal{L}_{\mathcal{C}} \otimes I_n \tilde{\Theta}_{k-h}] + O(\lambda_a^k b_k) \\
= & \beta \sum_{l=k-h}^{k-1} b_{l+1} \mathbb{E}[\tilde{\Theta}_{k-h}^T (\mathcal{L}_{\mathcal{C}} + \mathcal{L}_{\mathcal{C}}^T) \otimes I_n \tilde{\Theta}_{k-h}] + O(\lambda_a^k b_k) \\
= & 2\beta \sum_{l=k-h}^{k-1} b_{l+1} \mathbb{E}[\tilde{\Theta}_{k-h}^T \hat{\mathcal{L}}_{\mathcal{C}} \otimes I_n \tilde{\Theta}_{k-h}] + O(\lambda_a^k b_k), \qquad (31)
\end{aligned}
$$

where $\hat{\mathcal{L}}_{\mathcal{C}}$ is the Laplacian matrix of the mirror graph of the union graph $\mathcal{G}_{\mathcal{C}}$. By Lemma 5.2 and Assumption 1, the smallest nonzero eigenvalue $\hat{\lambda}_2$ of $\hat{\mathcal{L}}_{\mathcal{C}}$ satisfies $\hat{\lambda}_2 > 0$. From Assumption 2, Lemma 5.3, and (31), we have

$$
\begin{aligned}
& - \frac{\underline{f}}{h} \sum_{l=k-h}^{k-1} b_{l+1} \mathbb{E}[\tilde{\Theta}_{k-h}^T \sum_{l=k-h}^{k-1} \Phi_{l+1} \Phi_{l+1}^T \tilde{\Theta}_{k-h}] \\
& - \mathbb{E}[\tilde{\Theta}_{k-h}^T \sum_{l=k-h}^{k-1} b_{l+1} \mathcal{L}_{l+1} \otimes I_n \tilde{\Theta}_{k-h}] \\
\leq & - \sum_{l=k-h}^{k-1} b_{l+1} \mathbb{E}[\tilde{\Theta}_{k-h}^T \left(\frac{\underline{f}}{h} \mathbb{E}\left[\sum_{l=k-h}^{k-1} \Phi_{l+1} \Phi_{l+1}^T \middle| \mathcal{F}_{k-h} \right] \right. \\
& \left. + \hat{\mathcal{L}}_{\mathcal{C}} \otimes I_n \right) \tilde{\Theta}_{k-h}] + O(\lambda_a^k b_k) \\
\leq & - \frac{\sigma}{m \underline{f} \delta_\phi^2} \mathbb{E}[\lambda_{\min} \left(\mathbb{E}\left[\frac{\underline{f}}{h} \sum_{i=1}^{m} \sum_{l=k-h}^{k-1} \phi_{l+1,i} \phi_{l+1,i}^T \middle| \mathcal{F}_{k-h} \right] \right) \\
& \cdot \sum_{l=k-h}^{k-1} b_{l+1} \tilde{\Theta}_{k-h}^T \tilde{\Theta}_{k-h}] + O(\lambda_a^k b_k) \\
\leq & - \sigma \sum_{l=k-h}^{k-1} b_{l+1} V_{k-h} + O(\lambda_a^k b_k), \qquad (32)
\end{aligned}
$$

where $\sigma = \frac{h \underline{f} \hat{\lambda}_2 \delta_\phi^2}{2 \underline{f} \bar{\phi}^2 + h \hat{\lambda}_2}$. Taking (32) into (30) gives

$$
\begin{aligned}
V_k \leq & \left(1 - \left(\frac{2\beta\sigma}{c_h} - \frac{\beta}{\alpha_3}\right) \sum_{l=k-h}^{k-1} b_{l+1}\right) V_{k-h} \\
& + \alpha_3 \beta \bar{N} \sum_{l=k-h}^{k-1} b_{l+1} U_{k-h} + O(b_k^2).
\end{aligned}
$$

This completes the proof. □

5.3. Proof of Theorem 4.1

Based on Lemmas 4.1 and 4.2, taking $\alpha_1 = c_h/(\sigma \bar{f} \bar{\phi}^2)$, $\alpha_2 = c_h \bar{N}/\sigma \lambda_m^2$ and $\alpha_3 = 2c_h/\sigma$ can give

$$
\begin{cases}
U_k \leq \left(1 - \left(\frac{2\gamma \underline{g} \pi \delta_\psi^2}{c_h} - \beta \left(\frac{\bar{f}^2 \bar{\phi}^4 \sigma}{c_h} + 2\bar{N} + \frac{\sigma \lambda_m^2}{c_h}\right)\right) \sum_{l=k-h}^{k-1} \right. \\
\qquad \left. \cdot b_{l+1}\right) U_{k-h} + \frac{2\beta c_h \bar{N}}{\sigma} \sum_{l=k-h}^{k-1} b_{l+1} V_{k-h} + O\left(b_k^2\right), \\
V_k \leq \left(1 - \frac{3\beta\sigma}{2c_h} \sum_{l=k-h}^{k-1} b_{l+1}\right) V_{k-h} \\
\qquad + \frac{2\beta c_h \bar{N}}{\sigma} \sum_{l=k-h}^{k-1} b_{l+1} U_{k-h} + O\left(b_k^2\right),
\end{cases} \qquad (33)
$$

Take $w_1 = \frac{2\gamma \underline{g} \pi \delta_\psi^2}{c_h} - \beta \left(\frac{\bar{f}^2 \bar{\phi}^4 \sigma}{c_h} + 2\bar{N} + \frac{\sigma \lambda_m^2}{c_h}\right)$, $w_2 = -\frac{2\beta c_h \bar{N}}{\sigma}$ and $w_3 = \frac{3\beta\sigma}{2c_h}$. Denote

$$
Z_k = \begin{bmatrix} U_k \\ V_k \end{bmatrix}, \quad W = \begin{bmatrix} w_1 & w_2 \\ w_2 & w_3 \end{bmatrix}.
$$

By Assumption 5, there exists $K > 0$ such that $\sum_{l=k-h}^{k-1} b_{l+1}\lambda_{\max}(W) > 1$ for $k > K$. From (33),

$$\begin{aligned}\|Z_k\| &\le \Big\|I_2 - W\sum_{l=k-h}^{k-1} b_{l+1}\Big\|\|Z_{k-h}\| + O\left(b_k^2\right)\\ &\le \big(1-\lambda_{\min}(W)\sum_{l=k-h}^{k-1} b_{l+1}\big)\|Z_{k-h}\| + O\left(b_k^2\right),\end{aligned} \tag{34}$$

for all $k > K$. Actually, in order to prove the convergence of U_k and V_k, we only need to prove $\lim_{k\to\infty}\|Z_k\| = 0$. Then, based on Lemma 5.4 and Assumption 5, we only need to prove that $\lambda_{\min}(W) > 0$ because of $\sum_{k=1}^{\infty} b_k = \infty$ and $\lim_{k\to\infty} b_k^2/(\sum_{l=k-h}^{k-1} b_{l+1}) = 0$.

Let $|\lambda I_2 - W| = (\lambda - w_1)(\lambda - w_3) - w_2^2 = 0$. Then,

$$\lambda_{\min}(W) = \frac{1}{2}\Big(w_1 + w_3 - \sqrt{(w_1+w_3)^2 - 4(w_1w_3 - w_2^2)}\Big).$$

If $\gamma > \frac{\beta}{2\underline{g}\pi\delta_\psi^2}\big(2c_h\bar{N} + \sigma\left(\bar{f}^2\bar{\phi}^4 + \lambda_m^2\right) + \frac{8c_h^4\bar{N}^2}{3\sigma^3}\big)$, then we can obtain $w_1w_3 > w_2^2$ and $w_1 > 0$, which implies $\lambda_{\min}(W) > 0$. Then, from (9), we obtain $\lim_{k\to\infty}\mathbb{E}\|\tilde{\Theta}_k\|^2 = 0$ and $\lim_{k\to\infty}\mathbb{E}\|\varepsilon_k\|^2 = 0$.

5.4. Proof of Theorem 4.2

Part I: The case of $b_k = \frac{1}{k^p}$, $p \in (1/2, 1)$.

Based on (34), we have

$$\begin{aligned}\|Z_k\| &\le \left(1 - \lambda_{\min}(W)\sum_{l=k-h}^{k-1}\frac{1}{(l+1)^p}\right)\|Z_{k-h}\| + O\left(\frac{1}{k^{2p}}\right),\\ &\le \left(1 - \frac{\lambda_{\min}(W)h}{k^p}\right)\|Z_{k-h}\| + O\left(\frac{1}{k^{2p}}\right),\\ &\le \prod_{l=0}^{\left\lfloor\frac{k-K}{h}\right\rfloor-1}\left(1 - \frac{\lambda_{\min}(W)h}{(k-lh)^p}\right)\left\|Z_{k-\left\lfloor\frac{k-K}{h}\right\rfloor h}\right\|\\ &\quad + \sum_{l=1}^{\left\lfloor\frac{k-K}{h}\right\rfloor}\prod_{q=0}^{l-1}\left(1 - \frac{\lambda_{\min}(W)h}{(k-qh)^p}\right)O\left(\frac{1}{(k-lh)^{2p}}\right)\\ &\le \prod_{l=\left\lceil\frac{K}{h}\right\rceil+\kappa+1}^{\left\lceil\frac{k}{h}\right\rceil}\left(1 - \frac{\lambda_{\min}(W)h}{(lh)^p}\right)\left\|Z_{k-\left\lfloor\frac{k-K}{h}\right\rfloor h}\right\|\\ &\quad + \sum_{l=\left\lceil\frac{K}{h}\right\rceil+1}^{\left\lfloor\frac{k}{h}\right\rfloor-1}\prod_{q=\left\lceil\frac{K}{h}\right\rceil+\kappa+l+1}^{\left\lceil\frac{k}{h}\right\rceil}\left(1 - \frac{\lambda_{\min}(W)h}{(qh)^p}\right)O\left(\frac{1}{(lh)^{2p}}\right)\\ &\le \prod_{l=\left\lceil\frac{K}{h}\right\rceil+\kappa+1}^{\left\lceil\frac{k}{h}\right\rceil}\left(1 - \frac{\lambda_{\min}(W)h^{1-p}}{l^p}\right)\left\|Z_{k-\left\lfloor\frac{k-K}{h}\right\rfloor h}\right\|\\ &\quad + \sum_{l=\left\lceil\frac{K}{h}\right\rceil+1}^{\left\lfloor\frac{k}{h}\right\rfloor-1}\prod_{q=\left\lceil\frac{K}{h}\right\rceil+\kappa+l+1}^{\left\lceil\frac{k}{h}\right\rceil}\left(1 - \frac{\lambda_{\min}(W)h^{1-p}}{q^p}\right)O\left(\frac{1}{l^{2p}}\right).\end{aligned}$$

Then by it and Lemma 5.5, we learn $\|Z_k\| = O\left(\frac{1}{k^p}\right)$, which implies $\mathbb{E}\|\tilde{\Theta}_k\|^2 = O\left(\frac{1}{k^p}\right)$ and $\mathbb{E}\|\varepsilon_k\|^2 = O\left(\frac{1}{k^p}\right)$.

Part II: The case of $b_k = \frac{1}{k}$.

Based on (34), we have

$$\begin{aligned}\|Z_k\| &\le \left(1 - \lambda_{\min}(W)\sum_{l=k-h}^{k-1}\frac{1}{l+1}\right)\|Z_{k-h}\| + O\left(\frac{1}{k^2}\right),\\ &\le \left(1 - \frac{\lambda_{\min}(W)h}{k}\right)\|Z_{k-h}\| + O\left(\frac{1}{k^2}\right),\\ &\le \prod_{l=0}^{\left\lfloor\frac{k-K}{h}\right\rfloor-1}\left(1 - \frac{\lambda_{\min}(W)h}{k-lh}\right)\left\|Z_{k-\left\lfloor\frac{k-K}{h}\right\rfloor h}\right\|\\ &\quad + \sum_{l=1}^{\left\lfloor\frac{k-K}{h}\right\rfloor}\prod_{q=0}^{l-1}\left(1 - \frac{\lambda_{\min}(W)h}{k-qh}\right)O\left(\frac{1}{(k-lh)^2}\right)\\ &\le \prod_{l=\left\lceil\frac{K}{h}\right\rceil+\kappa+1}^{\left\lceil\frac{k}{h}\right\rceil}\left(1 - \frac{\lambda_{\min}(W)h}{lh}\right)\left\|Z_{k-\left\lfloor\frac{k-K}{h}\right\rfloor h}\right\|\\ &\quad + \sum_{l=\left\lceil\frac{K}{h}\right\rceil+1}^{\left\lfloor\frac{k}{h}\right\rfloor-1}\prod_{q=\left\lceil\frac{K}{h}\right\rceil+\kappa+l+1}^{\left\lceil\frac{k}{h}\right\rceil}\left(1 - \frac{\lambda_{\min}(W)h}{qh}\right)O\left(\frac{1}{(lh)^2}\right)\\ &\le \prod_{l=\left\lceil\frac{K}{h}\right\rceil+\kappa+1}^{\left\lceil\frac{k}{h}\right\rceil}\left(1 - \frac{\lambda_{\min}(W)}{l}\right)\left\|Z_{k-\left\lfloor\frac{k-K}{h}\right\rfloor h}\right\|\\ &\quad + \sum_{l=\left\lceil\frac{K}{h}\right\rceil+1}^{\left\lfloor\frac{k}{h}\right\rfloor-1}\prod_{q=\left\lceil\frac{K}{h}\right\rceil+\kappa+l+1}^{\left\lceil\frac{k}{h}\right\rceil}\left(1 - \frac{\lambda_{\min}(W)}{q}\right)O\left(\frac{1}{l^2}\right),\end{aligned}$$

where $\kappa = \left\lceil\frac{k-K}{h}\right\rceil - \left\lfloor\frac{k-K}{h}\right\rfloor$. Then by Lemma 5.6, we just need to prove $\lambda_{\min}(W) > 1$ to ensure $\|Z_k\| = O\left(\frac{1}{k}\right)$, implying $\mathbb{E}\|\tilde{\Theta}_k\|^2 = O\left(\frac{1}{k}\right)$ and $\mathbb{E}\|\varepsilon_k\|^2 = O\left(\frac{1}{k}\right)$.

Similar to the proof of Theorem 4.1, if $\gamma > \frac{\beta}{2\underline{g}\pi\delta_\psi^2}\big(2c_h\bar{N} + \sigma\left(\bar{f}^2\bar{\phi}^4 + \lambda_m^2\right) + \frac{8\beta c_h^4\bar{N}^2}{\sigma^2(3\beta\sigma - 2c_h)} + \frac{c_h}{\beta}\big)$, we have $w_1 > \frac{w_2^2}{w_3 - 1} + 1$. If $\beta > \frac{2c_h}{3\sigma}$, then, $w_3 - 1 > 0$, $w_1(w_3 - 1) > w_2^2 + w_3 - 1$. Therefore, $(w_1 + w_3)^2 - 4(w_1w_3 - w_2^2) < (w_1 + w_3 - 2)^2$.

Hence, $\lambda_{\min}(W) > 1$. This completes this part's proof.

6. Numerical example

This section demonstrates the effectiveness of the proposed EFTQDI algorithm through a numerical example. Additionally, we illustrate the joint effect of the sensors: with only one-bit information exchange, the network sensors collectively achieve an estimation task that no individual sensor could accomplish alone.

Example 1. Consider a WSN with $m = 6$ sensors, whose dynamics obey (1)–(2) with $n = 3$. The communication graph sequence $\{\mathcal{G}_k\}$ switches among $\mathcal{G}^{(1)}$, $\mathcal{G}^{(2)}$, $\mathcal{G}^{(3)}$ and $\mathcal{G}^{(4)}$ as shown in Fig. 1. For all $u = 1, 2, 3, 4$, $a_{ij}^{(u)} = \frac{2}{5}$ if $(j, i) \in \mathcal{E}^{(u)}$; and 0, otherwise. The graph sequence $\{\mathcal{G}_k\}$ switches according to a Markovian chain $\{m_k\}$, with initial probability $p_{u,1} = \mathbb{P}\left\{\mathcal{G}_1 = \mathcal{G}^{(u)}\right\} = \frac{1}{4}$ and the following transition probability matrix:

$$\mathbf{P} = \{p_{uv}\}_{4\times4} = \begin{bmatrix} 1/2 & 1/2 & 0 & 0\\ 0 & 1/2 & 1/2 & 0\\ 0 & 0 & 1/2 & 1/2\\ 1/2 & 0 & 0 & 1/2\end{bmatrix}$$

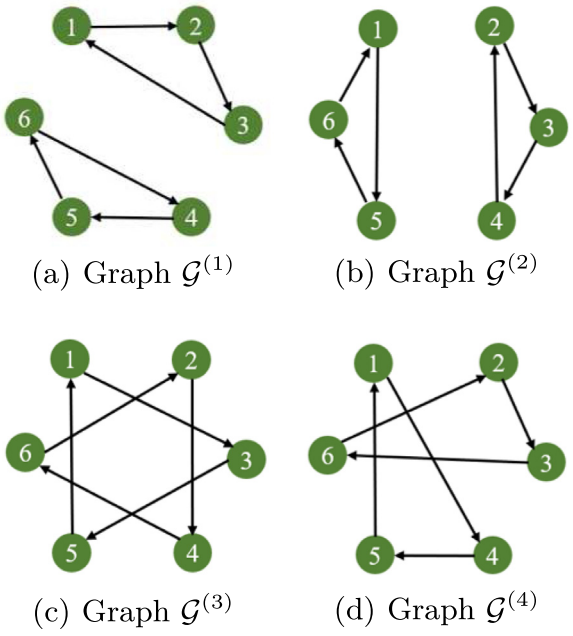

Fig. 1. Switching communication topology.

where $p_{uv} = \mathbb{P}\{m_k = v \mid m_{k-1} = u\}$. Therefore, the stationary distribution $\pi_u = \frac{1}{4}$ for all $u = 1, 2, 3, 4$.

In the dynamic model (1), the unknown parameter $\theta = [1, -1, 1]^T$ with prior information $\theta \in \Omega = [0, 2] \times [-2, 0] \times [0, 2]$. The observation noise $d_{k,i}$ is independent and identically distributed Gaussian with zero mean and variance 8^2. Let $\phi_{k,i} \in \mathbb{R}^3$ be generated by the following state space model

$$\begin{cases} x_{k,i} = A_i x_{k-1,i} + B_i \eta_{k,i}, \\ \phi_{k,i} = C_i x_{k,i}, \end{cases} \quad i = 1., \ldots, 6,$$

where $x_{k,i} \in \mathbb{R}^3$ with $x_{0,i} = [1.3, 1.3, 1.3]^T$, $\eta_{k,i} \in \mathbb{R}$ is independent and identically distributed with $\eta_{k,i}$ following uniform distribution on the interval $[-0.1, 0.1]$ and

$A_1 = \text{diag}\{1, 1/2, 1/2\}, A_4 = \text{diag}\{1, 5/6, 5/6\}$

$A_2 = \text{diag}\{1/2, 1, 1/2\}, A_5 = \text{diag}\{5/6, 1, 5/6\},$

$A_3 = \text{diag}\{1/2, 1/2, 1\}, A_6 = \text{diag}\{5/6, 5/6, 1\}$

$B_1 = B_4 = [1, 0, 0]^T, B_2 = B_5 = [0, 1, 0]^T,$

$B_3 = B_6 = [0, 0, 1]^T,$

$C_1 = \text{diag}\{1, 0, 0\}, C_4 = \text{diag}\{-1, 0, 0\},$

$C_2 = \text{diag}\{0, 1, 0\}, C_5 = \text{diag}\{0, -1, 0\},$

$C_3 = \text{diag}\{0, 0, 1\}, C_6 = \text{diag}\{0, 0, -1\},$

It can be verified that Assumption 2 is satisfied with $h = 2$. Besides, the thresholds of quantized measurements are $C_i = 1$ for all $i = 1, \ldots, 6$. In the quantized communication mechanism (2), the communication noise $\omega_{k,ij}$ is independent and identically distributed Gaussian with zero mean and variance 1, and the threshold of quantized communications is $C_{ij} = 0$ for all the channel $(j, i) \in \mathcal{E}_{\mathcal{C}}$.

Then, we apply the proposed EFTQDI algorithm with two types of step sizes to produce the estimates, where the linear encoding rule ψ_k in (4) switches sequentially among $\{[1, 0, 0]^T, [0, 1, 0]^T, [0, 0, 1]^T\}$. The simulation is repeated 100 times using the same initial values $\hat{\theta}_{0,ij} = [1/2, -1/2, 1/2]$ and $\theta_{0,i} = [1/2, -1/2, 1/2]$ to calculate the empirical variance of estimation errors, representing the mean square errors.

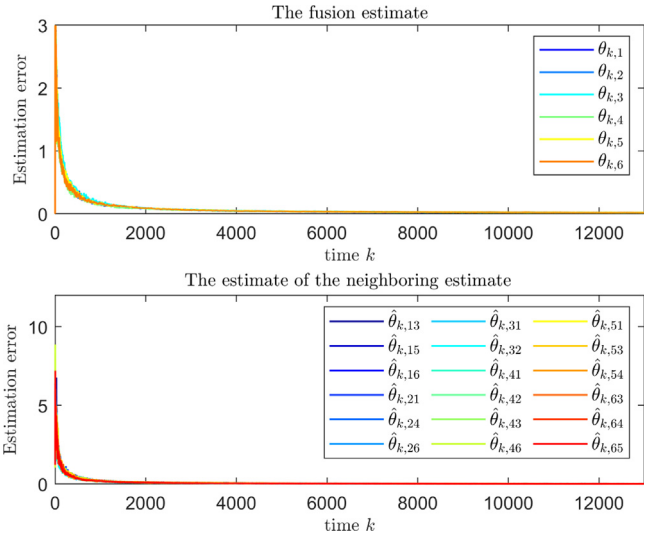

Fig. 2. Convergence of the two estimates with $b_k = \frac{1}{k}$.

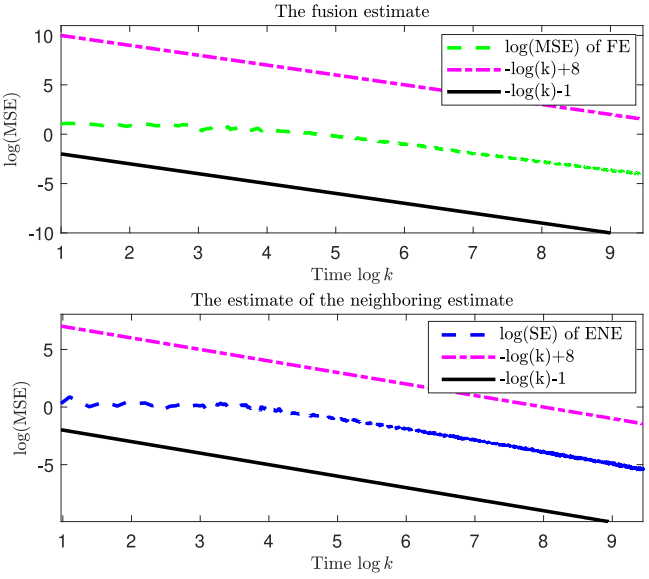

Fig. 3. Convergence rate of the two estimates with $b_k = \frac{1}{k}$.

In the first case, the step size for the EFTQDI algorithm is set as $b_k = \frac{1}{k}$ with the step coefficients $\beta = 39$ and $\gamma = 74$. The mean square errors (MSEs) of both the fusion estimate and the estimate of the neighboring estimate are shown in Fig. 2. These results indicate that the fusion estimate and the estimate of the neighboring estimate converge to the true parameter and fusion estimate, respectively. Additionally, Fig. 3 shows a linear relationship between the logarithm of the MSEs and the logarithm of the index k, which indicates the mean square convergence rates of two estimates are $O(\frac{1}{k})$.

Fig. 4 shows the MSEs trajectories of the proposed EFTQDI algorithm compared to its non-cooperative counterpart (i.e., $\theta_{k,i} = \Pi_{\Omega}\{\theta_{k-1,i} + \beta b_k \phi_{k,i}(\hat{F}_{k,i} - s_{k,i})\}$). The results indicate that, unlike the non-cooperative algorithm, which does not converge to zero, the EFTQDI algorithm's estimate converges to the true parameter. This highlights the joint effect of sensors under quantized communication, as one-bit information exchange among sensors enables the estimation task that individual sensors cannot achieve alone.

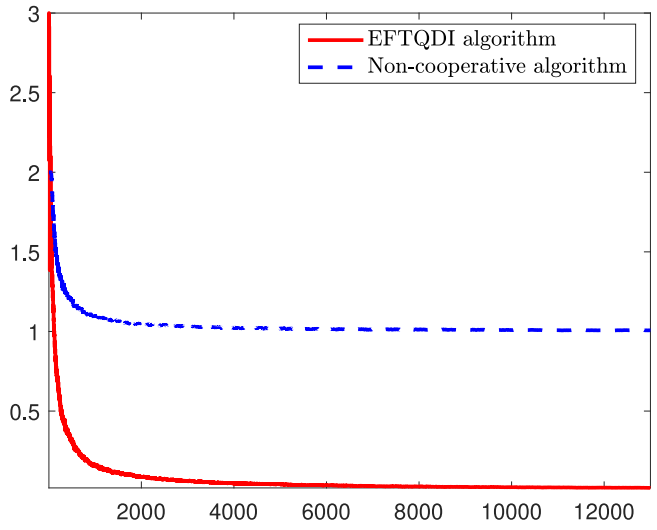

Fig. 4. The comparison between the EFTQDI algorithm and non-cooperative algorithm under $b_k = \frac{1}{k}$.

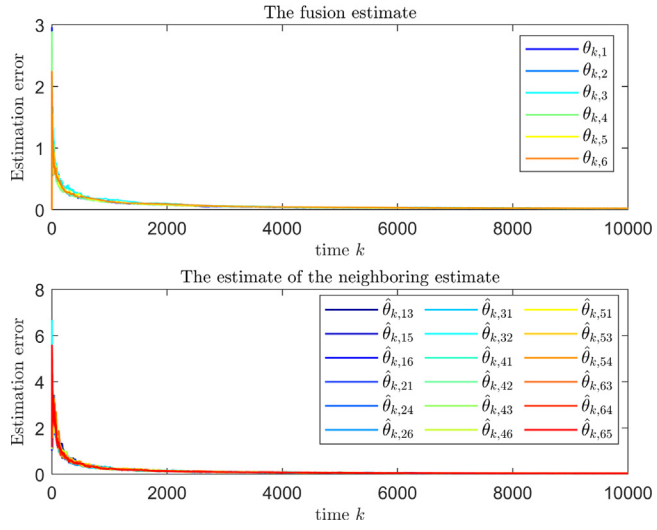

Fig. 5. Convergence of the two estimates with $b_k = \frac{1}{k^{4/5}}$.

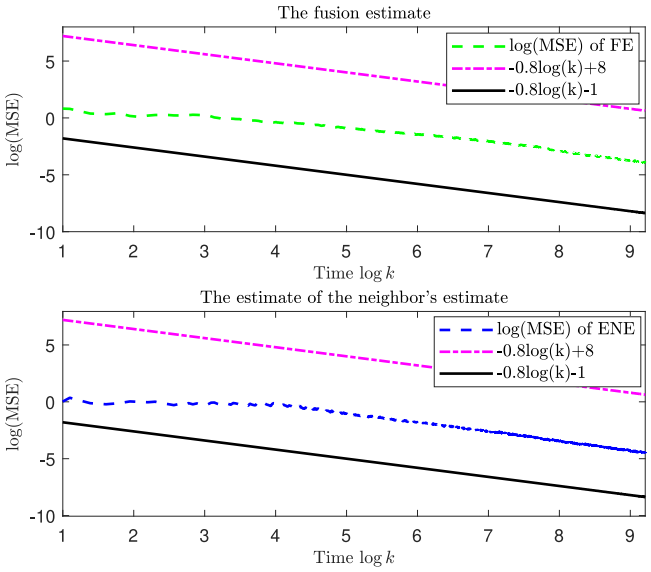

Fig. 6. Convergence rate of the two estimates with $b_k = \frac{1}{k^{4/5}}$.

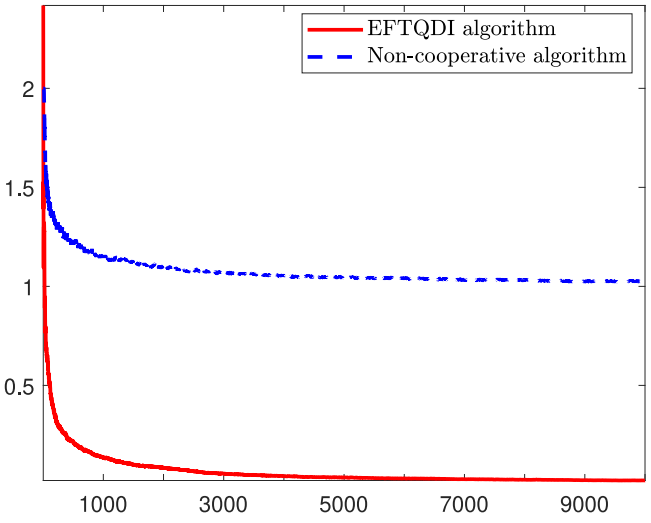

Fig. 7. The comparison between the EFTQDI algorithm and non-cooperative algorithm under $b_k = \frac{1}{k^{4/5}}$.

In the second case, the step size in the EFTQDI algorithm is set as $b_k = \frac{1}{k^{4/5}}$ with step coefficients $\beta = 16$ and $\gamma = 65$. Fig. 5 shows that both the fusion estimate and the estimate of the neighboring estimate converge, respectively. Besides, Fig. 6 shows that the logarithms of MSEs of the two estimates scale linearly with $\frac{4}{5}\log k$, indicating a mean square convergence rate of $O(\frac{1}{k^{4/5}})$. Fig. 7 shows the MSE trajectories of the EFTQDI algorithm and its corresponding non-cooperative algorithm, which demonstrates again the joint effect of the sensors under quantized communication.

7. Conclusion

This paper addresses distributed parameter estimation in stochastic dynamic systems with quantized measurements, focusing on quantized communication and Markov-switching directed topologies. A compression encoding method satisfying persistent excitation is introduced, ensuring original signal recovery from quantized transmitted data. Based on this, an EFTQDI algorithm is proposed, leveraging a stochastic approximation idea. This approach first estimates the neighboring estimates through quantized communication, followed by a consensus-based fusion strategy to integrate these estimates. By constructing and analyzing two coupled Lyapunov functions for estimation errors, we establish the mean-square convergence of these two estimates under a cooperative excitation condition and the assumption that the union topology includes a spanning tree. Furthermore, we demonstrate that the mean-square convergence rate aligns with the step size order given appropriate coefficient conditions.

Future work could explore the combination of event-driven mechanisms and quantized protocol, which both can reduce communication load. Developing effective event-driven quantized transmission protocols to ensure distributed estimation with minimal data transfer remains a compelling challenge.

References

Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensor networks: a survey. *Computer Networks*, *38*, 393–422.

Bianchi, P., Fort, G., & Hachem, W. (2013). Performance of a distributed stochastic approximation algorithm. *Institute of Electrical and Electronics Engineers. Transactions on Information Theory*, *59*, 7405–7418.
Calamai, P. H., & Moré, J. J. (1987). Projected gradient methods for linearly constrained problems. *Mathematical Programming*, *39*, 93–116.
Dimakis, A., Kar, S., Moura, J., Rabbat, M., & Scaglione, A. (2010). Gossip algorithms for distributed signal processing. *Proceedings of the IEEE*, *98*, 1847–1864.
Ding, D., Wang, Z., Ho, D., & Wei, G. (2017). Distributed recursive filtering for stochastic systems under uniform quantizations and deception attacks through sensor networks. *Automatica*, *78*, 231–240.
Fu, K., Chen, H. F., & Zhao, W. (2022). Distributed system identification for linear stochastic systems with binary sensors. *Automatica*, *141*, Article 110298.
Gan, D., & Liu, Z. (2024). Convergence of the distributed SG algorithm under cooperative excitation condition. *IEEE Transactions on Neural Networks and Learning Systems*, *35*, 7087–7101.
Ge, X., Han, Q. L., Zhang, X., Ding, L., & Yang, F. (2020). Distributed event-triggered estimation over sensor networks: A survey. *IEEE Transactions on Cybernetics*, *50*, 1306–1320.
Guo, J., & Zhao, Y. L. (2013). Recursive projection algorithm on FIR system identification with binary-valued observations. *Automatica*, *49*, 3396–3401.
Han, F., Wang, Z., Dong, H., Yi, X., & Alsaadi, F. (2023a). Distributed H_∞-consensus estimation for random parameter systems over binary sensor networks: A local performance analysis method. *IEEE Transactions on Network Science and Engineering*, *10*, 2334–2346.
Han, F., Wang, Z., Dong, H., Yi, X., & Alsaadi, F. (2023b). Recursive distributed filtering for time-varying systems over sensor network via rayleigh fading channels: Tackling binary measurements. *IEEE Transactions on Signal and Information Processing over Networks*, *9*, 110–122.
Kar, S., Moura, J., & Ramanan, K. (2012). Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication. *Institute of Electrical and Electronics Engineers. Transactions on Information Theory*, *58*, 3575–3605.
Lei, J., & Chen, H. F. (2020). Distributed stochastic approximation algorithm with expanding truncations. *IEEE Transactions on Automatic Control*, *65*, 664–679.
Li, J., Wang, Z., Lu, R., & Xu, Y. (2023). Distributed filtering under constrained bit rate over wireless sensor networks: Dealing with bit rate allocation protocol. *IEEE Transactions on Automatic Control*, *68*, 1642–1654.
Li, T., & Zhang, J. F. (2010). Consensus conditions of multi-agent systems with time-varying topologies and stochastic communication noises. *IEEE Transactions on Automatic Control*, *55*, 2043–2057.
Mateos, G., & Giannakis, G. (2012). Distributed recursive least-squares: Stability and performance analysis. *IEEE Transactions on Signal Processing*, *60*, 3740–3754.
Polyak, B. (1987). *Introduction to optimization*. New York: Optimization Soft-ware Inc..
Pottie, G., & Kaiser, W. (2000). Wireless integrated network sensors. *Communications of the ACM*, *43*, 51–58.
Rawat, P., Singh, K., Chaouchi, H., & Bonnin, J. (2014). Wireless sensor networks: a survey on recent developments and potential synergies. *Journal of Supercomputing*, *68*, 1–48.
Ren, W., & Beard, R. (2005). Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Transactions on Automatic Control*, *50*, 655–661.
Seneta, E. (2006). *Non-negative matrices and Markov chains*. NY, USA: Springer.
Sinha, N., & Griscik, M. (1971). A stochastic approximation method. *IEEE Transactions on Systems, Man and Cybernetics*, *SMC-1*, 338–344.
Swenson, B., Sridhar, A., & Poor, H. (2020). On distributed stochastic gradient algorithms for global optimization. In *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing* (pp. 8594–8598).
Taj, M., & Cavallaro, A. (2011). Distributed and decentralized multicamera tracking. *IEEE Signal Processing Magazine*, *28*, 46–58.
Takahashi, N., Yamada, I., & Sayed, A. (2010). Diffusion least-mean squares with adaptive combiners: Formulation and performance analysis. *IEEE Transactions on Signal Processing*, *58*, 4795–4810.
Wang, J., Ke, J., & Zhang, J. F. (2024). Differentially private bipartite consensus over signed networks with time-varying noises. *IEEE Transactions on Automatic Control*, *69*, 5788–5803.
Wang, T., Tan, J., & Zhao, Y. L. (2018). Asymptotically efficient non-truncated identification for FIR systems with binary-valued outputs. *Science China. Information Sciences*, *61*, Article 129208.
Wang, Y., & Zhang, J. F. (2019). Distributed parameter identification of quantized ARMAX systems. In *Proceedings of the 38th Chinese control conference* (pp. 1701–1706).
Wang, L. Y., Zhang, J. F., & Yin, G. (2003). System identification using binary sensors. *IEEE Transactions on Automatic Control*, *48*, 1892–1907.
Wang, Y., Zhao, Y. L., & Zhang, J. F. (2021). Distributed recursive projection identification with binary-valued observations. *Journal of Systems Science and Complexity*, *34*, 2048–2068.
Wang, Y., Zhao, Y. L., Zhang, J. F., & Guo, J. (2022). A unified identification algorithm of FIR systems based on binary observations with time-varying thresholds. *Automatica*, *135*, Article 109990.
Xie, S., & Guo, L. (2018). Analysis of normalized least mean squares-based consensus adaptive filters under a general information condition. *SIAM Journal on Control and Optimization*, *56*, 3404–3431.
Xie, S., Zhang, Y., & Guo, L. (2021). Convergence of a distributed least squares. *IEEE Transactions on Automatic Control*, *66*, 4952–4959.
You, K. (2015). Recursive algorithms for parameter estimation with adaptive quantizer. *Automatica*, *52*, 192–201.
Zhang, H., Wang, T., & Zhao, Y. L. (2021). Asymptotically efficient recursive identification of FIR systems with binary-valued observations. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *51*, 2687–2700.
Zhang, Q., & Zhang, J. F. (2012). Distributed parameter estimation over unreliable networks with markovian switching topologies. *IEEE Transactions on Automatic Control*, *57*, 2545–2560.
Zhang, L., Zhao, Y. L., & Guo, L. (2022). Identification and adaptation with binary-valued observations under non-persistent excitation condition. *Automatica*, *138*, Article 110158.
Zhao, Y. L., Zhang, H., & Kang, G. (2023). System identification under saturated precise or set-valued measurements. *Science China. Information Sciences*, *66*, Article 112204.
Zhu, S., Chen, C., Xu, J., Guan, X., Xie, L., & Johansson, K. H. (2018). Mitigating quantization effects on distributed sensor fusion: A least squares approach. *IEEE Transactions on Signal Processing*, *66*, 3459–3474.
Zhu, S., Liu, S., Soh, Y., & Xie, L. (2017). Performance analysis of averaging based distributed estimation algorithm with additive quantization model. *Automatica*, *80*, 95–101.
Zhu, S., Soh, Y., & Xie, L. (2015). Distributed parameter estimation with quantized communication via running average. *IEEE Transactions on Signal Processing*, *63*, 4634–4646.

Ying Wang received the B.S. degree in mathematics from Wuhan University, Wuhan, China, in 2017, and the Ph.D. degree in system theory from the Academy of Mathematics and Systems Science (AMSS), Chinese Academy of Sciences (CAS), Beijing, China, in 2022.

She is currently a postdoctoral fellow at the AMSS, CAS, and Division of Decision and Control Systems, KTH Royal Institute of Technology, Stockholm, Sweden. Her research interests include parameter identification and adaptive control of quantized systems. She was a recipient of the Special Prize of the Presidential Scholarship of Chinese Academy of Sciences in 2022, and the Outstanding Doctoral Dissertation Award of the Chinese Association of Automation in 2024.

Jian Guo received the B.S. degree in Mathematics from Xi'an Jiaotong University, Xi'an, China, in 2019, and the Ph.D. degree in system analysis and integration from the Academy of Mathematics and Systems Science (AMSS), Chinese Academy of Sciences (CAS), Beijing, China, in 2024. He is currently a postdoc researcher at the Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China.

His research interests include sparse identification, control of multi-agent systems, and dynamic stochastic variational inequalities.

Yanlong Zhao received the B.S. degree in mathematics from Shandong University, Jinan, China, in 2002, and the Ph.D. degree in systems theory from the Academy of Mathematics and Systems Science (AMSS), Chinese Academy of Sciences (CAS), Beijing, China, in 2007.

Since 2007, he has been with the AMSS, CAS, where he is currently a full Professor and a Vice Director of State Key Laboratory of Mathematical Sciences. His research interests include the identification and control of quantized systems, information theory and modeling of financial systems.

Yanlong Zhao was awarded the Second Prize of the State Natural Science Award of China in 2015. He has been a Deputy Editor-in-Chief of Journal of Systems Science and Complexity, an Associate Editor for Automatica, SIAM Journal on Control and Optimization, and IEEE Transactions on Systems, Man and Cybernetics: Systems. He was a Vice-President of Asian Control Association and a Vice-President of IEEE CSS Beijing Chapter, and is now Vice-President and Fellow of the Chinese Association of Automation (CAA) and the Chair of Technical Committee on Control Theory, CAA.

Ji-Feng Zhang received the B.S. degree in mathematics from Shandong University, China, in 1985, and the Ph.D. degree from the Institute of Systems Science, Chinese Academy of Sciences (CAS), China, in 1991. He is now with the Zhongyuan University of Technology and the Academy of Mathematics and Systems Science, CAS.

His current research interests include system modeling, adaptive control, stochastic systems, and multi-agent systems.

He is an IEEE Fellow, IFAC Fellow, CAA Fellow, SIAM Fellow, member of the European Academy of Sciences and Arts, and Academician of the International Academy for Systems and Cybernetic Sciences. He received the Second Prize of the State Natural Science Award of China in 2010 and 2015, respectively. He was a Vice-Chair of the IFAC Technical Board, member of the Board of Governors, IEEE Control Systems Society; Convenor of Systems Science Discipline, Academic Degree Committee of the State Council of China; Vice-President of the Chinese Association of Automation, the Systems Engineering Society of China, and the Chinese Mathematical Society. He served as Editor-in-Chief, Deputy Editor-in-Chief or Associate Editor for more than 10 journals, including Science China Information Sciences, IEEE Transactions on Automatic Control and SIAM Journal on Control and Optimization, etc.