SIGNAL-COMPARISON-BASED DISTRIBUTED ESTIMATION UNDER DECAYING AVERAGE DATA RATE COMMUNICATIONS*

JIEMING KE†, XIAODONG LU‡, YANLONG ZHAO§, AND JI-FENG ZHANG¶

Abstract. The paper investigates the distributed estimation problem under low data rate communications. Based on the signal-comparison (SC) consensus protocol under binary-valued communications, a new consensus+innovations type distributed estimation algorithm is proposed. Firstly, the high-dimensional estimates are compressed into binary-valued messages by using a periodic compressive strategy, dithering noises and a sign function. Next, based on the dithering noises and expanding triggering thresholds, a new stochastic event-triggered mechanism is proposed to reduce the communication frequency. Then, a modified SC consensus protocol is applied to fuse the neighborhood information. Finally, a stochastic approximation estimation algorithm is used to process innovations. The proposed SC-based algorithm has the advantages of high effectiveness and low communication cost. For the effectiveness, the estimates of the SC-based algorithm converge to the true value in the almost sure and mean square sense, and a polynomial almost sure convergence rate is also obtained. For the communication cost, the local and global average data rates decay to zero at a polynomial rate. The trade-off between the convergence rate and the communication cost is established through event-triggered coefficients. A better convergence rate can be achieved by decreasing event-triggered coefficients, while lower communication cost can be achieved by increasing event-triggered coefficients. A simulation example is given to demonstrate the theoretical results.

Key words. distributed estimation, data rate, event-triggered mechanism, stochastic approximation

MSC codes. 68W15, 93B30, 68P30, 62L20

1. Introduction.

Distributed estimation is of great practical significance in many practical fields, such as electric power grid [11] and cognitive radio systems [24], and therefore has been being an attractive topic [7, 12, 23, 30]. In the distributed estimation problem, the subsystem of each sensor is not necessarily observable. Therefore, communications between sensors are required to fuse the observations of the distributed sensors, which brings communication cost problems. Firstly, due to the bandwidth limitations in the real digital networks, high data rate communications may cause network congestion. Secondly, the transmission energy cost is positively correlated with the bit numbers of communication messages [16]. Therefore, it is

*Submitted to the editors DATE.

Funding: The work was supported by National Key R&D Program of China under Grant 2018YFA0703800, National Natural Science Foundation of China under Grants T2293772 and 62025306, and CAS Project for Young Scientists in Basic Research under Grant YSBR-008.

†Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, People's Republic of China, and School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China (kejieming@amss.ac.cn)

‡School of Automation and Electrical Engineering University of Science and Technology Beijing, Beijing 100083, People's Republic of China, and Key Laboratory of Knowledge Automation for Industrial Processes, Ministry of Education, Beijing 100083, People's Republic of China (lxd1025744209@163.com)

§Corresponding author. Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, People's Republic of China, and School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100149, People's Republic of China (ylzhao@amss.ac.cn)

¶School of Automation and Electrical Engineering, Zhongyuan University of Technology, Zhengzhou 450007, Henan Province, People's Republic of China, and Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, People's Republic of China. (jif@iss.ac.cn)

important to propose a distributed estimation algorithm under low data rate communications.

There have been many works in quantization methods to reduce the communication cost for distributed algorithms [2, 3, 4, 13, 39, 40], many of which are based on infinity level quantizers. For example, Aysal et al. adopt infinite level probabilistic quantizers to construct a quantized consensus algorithm [2]. Furthermore, Carli et al. [3, 4] propose an important technique based on infinite level logarithm quantizers to give quantized coordination algorithms and a quantized average consensus algorithm. Kar and Moura [13] appear to be the first to consider distributed estimation under quantized communications. They improve the probabilistic quantizer-based consensus algorithm in [2] by using the stochastic approximation method. Based on the technique, the estimates of corresponding consensus+innovations distributed estimation algorithm converge to the true value. Besides, when there is only one observation for each sensor, Zhu et al. [39, 40] propose running average distributed estimation algorithms based on probabilistic quantizers.

Due to the data rate limitations in real digital networks, distributed algorithms under finite data rate communications are developed. This is a challenging task because information contained in the interactive messages is limited. To solve the difficulty, Li et al. [17], Liu et al. [18], and Meng et al. [20] design zooming-in methods for the consensus problems under finite data rate communications. The methods are effective to deal with the quantization error. When communication noises exist, Zhao et al. [38] and Wang et al. [32] propose an empirical measurement-based consensus algorithm and a recursive projection consensus algorithm under binary-valued communications, respectively.

Distributed estimation under finite data rate communications has also been extensively investigated [5, 15, 21, 22, 25, 35]. Xie and Li [35] design finite level dynamical quantization method for distributed least mean square estimation under finite data rate communications. Sayin and Kozat [25] propose a single bit diffusion algorithm, which requires least data rate among existing works. Assuming that the Euclidean norm of messages can be transmitted with high precision, Carpentiero et al. [5] and Lao et al. [15] apply the quantizer in [1] and propose adapt-compress-then-combine diffusion algorithm and quantized adapt-then-combine diffusion algorithm, respectively. The estimates of these algorithms are all mean square bounded, but the almost sure and mean square convergence is not achieved. Additionally, the offline distributed estimation problem under finite data rate can be modelled as a distributed learning problem, which is solved by Michelusi et al. [21] and Nassif et al. [22]. However, under finite data rate communications, how to design an online distributed estimation algorithm with estimation errors converging to zero is still an open problem.

Despite the remarkable progress in distributed estimation under finite data rate communications [5, 15, 25, 35], we propose a novel distributed estimation with better effectiveness and lower communication cost. For the effectiveness, the estimates of the algorithm converge to the true value. For the communication cost, the average data rates decay to zero.

Both of the two issues are challenging. For the effectiveness, the main difficulty lies in the selection of consensus protocols to fuse the neighborhood information. Note that consensus protocol is an important part for both the consensus+innovation type distributed estimation algorithms and the diffusion type distributed estimation algorithms. A proper selection of consensus protocols can solve many communication problems in distributed estimation, including the communication cost problem. Under finite data rate communications, there have been many consensus protocols [17, 18, 20, 32, 38], but many of them have limitations when applied to distributed estimation. For example, the consensus protocol in [38] requires the states to keep constant in most of the times, which results in a relatively poor effectiveness. Besides, the consensus protocols in [17, 18, 20, 32] are proved to achieve consensus only when all the states are located in known compact sets. This limits their application in the distributed estimation problem due to the randomness of measurements and the lack of *a priori* information on the location of unknown parameter.

The limitations can be overcome by using the signal-comparison (SC) consensus protocol that we [14] propose recently. Firstly, the convergence analysis of the SC protocol does not require that all the states are located in known compact sets. Secondly, the SC protocol updates the states at every moment, and therefore achieves a better convergence rate compared with [38]. Hence, the SC protocol is suitable to be applied in the distributed estimation.

For the communication cost, if information is transmitted at every moment, the minimum data rate is 1. Therefore, the communication frequency should be reduced to achieve a average data rate that decay to zero. The event-triggered strategy is an important method to reduce communication frequency, and is widely applied in consensus control [27, 34], distributed Nash equilibrium [28] and impulsive synchronization [33]. For the distributed estimation problem, He et al. [10] propose an event-triggered algorithm where the communication rate can decay to zero at a polynomial rate. However, the mechanism requires accurate transmission of local estimates, making it difficult to extend to the quantized communication case. Therefore, it is important to propose a new event-triggered mechanism for the distributed estimation under quantized communications.

For the distributed estimation problem under quantized communications, we propose a new stochastic event-triggered mechanism, which consists of dithering noises and expanding triggering thresholds. The mechanism is suitable for the quantized communication case, because it regards whether the information is transmitted as part of quantized information.

Based on the SC consensus protocol and the stochastic event-triggered mechanism, we construct the SC-based distributed estimation algorithm. The main contributions are summarized as follows.

1. For the effectiveness, the estimates of the SC-based algorithm converge to the true value in the almost sure and mean square sense. A polynomial almost sure convergence rate is obtained for the SC-based algorithm. Under finite data rate communications, the SC-based distributed estimation algorithm is the first to achieve convergence. Moreover, it is the first to characterize the almost sure properties of a distributed estimation algorithm under finite data rate communications.
2. For the communication cost, the average data rates of the SC-based algorithm decay to zero almost surely. The upper bounds of local average data rates are estimated, and both the local and global average data rates converge to zero at a polynomial rate. The SC-based algorithm requires the least average data rates among existing works for distributed estimation [13, 21, 22, 25, 35].
3. The trade-off between the convergence rate and the communication cost is established via event-triggered coefficients. A better convergence rate can be achieved by decreasing event-triggered coefficients, while a lower communication cost can be achieved by increasing event-triggered coefficients. The operator of each sensor can decide its own preference on the trade-off by

selecting the event-triggered coefficients of adjacent communication channels.

The remainder of the paper is organized as follows. Section 2 formulates the problem. Section 3 introduces the SC consensus protocol and proposes the SC-based distributed estimation algorithm. Section 4 analyzes the convergence properties of the algorithm. Section 5 calculates the average data rates of the SC-based algorithm to measure the communication cost. Section 6 discusses the trade-off between the convergence rate and the communication cost for the algorithm. Section 7 gives a simulation example to demonstrate the theoretical results. Section 8 concludes the paper.

Notation. In the rest of the paper, $\mathbb{N}$, $\mathbb{R}$, $\mathbb{R}^n$, and $\mathbb{R}^{n\times m}$ are the sets of natural numbers, real numbers, n-dimensional real vectors, and $n\times m$-dimensional real matrices, respectively. $\|x\|$ is the Euclidean norm for vector x, and $\|A\|$ is the induced matrix norm for matrix A. Besides, $\|x\|_1$ is the L_1 norm. I_n is an $n\times n$ identity matrix. $\mathbf{1}_n$ is the n-dimensional vector whose elements are all ones. $\operatorname{diag}\{\cdot\}$ denotes the block matrix formed in a diagonal manner of the corresponding numbers or matrices. $\operatorname{col}\{\cdot\}$ denotes the column vector stacked by the corresponding numbers or vectors. $\otimes$ denotes the Kronecker product. Given two series $\{a_k\}$ and $\{b_k\}$, $a_k = O(b_k)$ means that $a_k = c_k b_k$ for a bounded c_k, and $a_k = o(b_k)$ means that $a_k = c_k b_k$ for a c_k that converges to 0.

2. Problem formulation.

This section introduces the graph preliminaries and formulates the distributed estimation problem under decaying average data rate communications.

2.1. Graph preliminaries.

In this paper, the communications between sensors can be described by an undirected weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$. $\mathcal{V} = \{1, \dots, N\}$ is the set of the sensors. $\mathcal{E} = \{(i,j) : i,j \in \mathcal{V}\}$ is the edge set. $(i,j)\in\mathcal{E}$ if and only if the sensor i and the sensor j can communicate with each other. $\mathcal{A} = (a_{ij})_{N\times N}$ represents the symmetric weighted adjacency matrix of the graph whose elements are all non-negative. $a_{ij} > 0$ if and only if $(i,j)\in\mathcal{E}$. Besides, $\mathcal{N}_i = \{j : (i,j)\in\mathcal{E}\}$ is used to denote the sensor i's the neighbor set. Define Laplacian matrix as $\mathcal{L} = \mathcal{D} - \mathcal{A}$, where $\mathcal{D} = \operatorname{diag}\left(\sum_{i\in\mathcal{N}_1} a_{i1}, \dots, \sum_{i\in\mathcal{N}_N} a_{iN}\right)$. The graph $\mathcal{G}$ is said to be connected if $\operatorname{rank}(\mathcal{L}) = N-1$.

2.2. Problem statement.

Consider a network $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ with N sensors. The sensor i observes the unknown parameter $\theta\in\mathbb{R}^n$ from the observation model

$$\mathtt{y}_{i,k} = H_{i,k}\theta + \mathtt{w}_{i,k},$$

where k is the time index, $H_{i,k}\in\mathbb{R}^{m_i\times n}$ is the measurement matrix, $\mathtt{w}_{i,k}\in\mathbb{R}^{m_i}$ is the observation noise, and $\mathtt{y}_{i,k}\in\mathbb{R}^{m_i}$ is the observation. Define σ-algebra $\mathcal{F}_k^w = \sigma(\{\mathtt{w}_{i,t} : i\in\mathcal{V},\ 1\le t\le k\})$.

The assumptions of the observation model are given as below.

Assumption 2.1. There exists $\bar{H} > 0$ such that $\|H_{i,k}\| \le \bar{H}$ for all $k\ge 1$ and $i = 1,\dots,N$. There exists a positive integer p and a positive real number δ such that

$$\frac{1}{p}\sum_{t=k}^{k+p-1}\sum_{i=1}^{N} H_{i,t}^{\top} H_{i,t} \ge \delta I_n,\ \ k\ge 1. \tag{2.1}$$

Remark 2.2. The condition (2.1) is the cooperative persistent excitation condition, and is common in existing literature for distributed estimation. For example, [12, 23] assumes that $H_{i,k}$ is constant for all k and $\frac{1}{N}\sum_{i=1}^{N} H_{i,k}^{\top}\Sigma_w^{-1}H_{i,k}$ is invertible, where Σ_w is the nonsingular covariance of $\mathtt{w}_{i,k}$. This condition is a special case for Assumption 2.1.

Assumption 2.3. $\{\mathtt{w}_{i,k}, \mathcal{F}_k\}$ is a martingale difference sequence such that

$$\sup_{i\in\mathcal{V},\ k\in\mathbb{N}} \mathbb{E}\left[\|\mathtt{w}_{i,k}\|^{\rho}\middle|\mathcal{F}_{k-1}^{w}\right] < \infty, \text{ a.s.} \tag{2.2}$$

for some $\rho > 2$.

Remark 2.4. $\mathtt{w}_{i,k}$ and $\mathtt{w}_{j,k}$ is allowed to be correlated for $i \neq j$, which makes our model applicable to more practical scenarios, such as the distributed target localization [13].

Assumption 2.5. The communication graph $\mathcal{G}$ is connected.

The goal of this paper is to cooperatively estimate the unknown parameter θ. Cooperative estimation requires information exchange between sensors, which brings communication cost. We use the average data rates to describe the communication cost of the distributed estimation.

DEFINITION 2.6. *Given time interval $[1,k]\cap\mathbb{N}$, the local average data rate for the communication channel where the sensor i sends messages to the neighbor j*

$$\mathtt{B}_{ij}(k) = \frac{\sum_{t=1}^{k}\zeta_{ij}(t)}{k}, \tag{2.3}$$

where $\zeta_{ij}(t)$ is the bit number of the message that the sensor i sends to the sensor j at time t. The global average data rate of communication is

$$\mathtt{B}(k) = \frac{\sum_{(i,j)\in\mathcal{E}}\sum_{t=1}^{k}\zeta_{ij}(t)}{2kM},$$

where M is the edge number of the communication graph.

Remark 2.7. From Definition 2.6, one can get $\mathtt{B}(k) = \frac{\sum_{(i,j)\in\mathcal{E}}\mathtt{B}_{ij}(k)}{2M}$.

Remark 2.8. The average data rates are used to describe the communication cost because they can represent the consumption of bandwidth, and are also related to transmission energy cost [16].

There have been distributed estimation algorithms with $\mathtt{B}(k) < \infty$. For example, $\mathtt{B}(k)$ of the distributed least mean square algorithm with $2K+1$ level dynamical quantizer in [35] is $n\lceil\log_2(2K+1)\rceil$, where $\lceil\cdot\rceil$ is the minimum integer that is no smaller than the given number. $\mathtt{B}(k)$ of the single-bit diffusion algorithm in [25] is 1. For effectiveness, these algorithms are shown to be mean square stable [25, 35].

Here, we propose a new distributed estimation algorithm with better effectiveness and lower communication cost. For the effectiveness, the estimation errors converge to zero at a polynomial rate. For the communication cost, $\mathtt{B}_{ij}(k)$ for all communication channels $(i,j)\in\mathcal{E}$ and $\mathtt{B}(k)$ also converge to zero.

3. Algorithm construction.

The section constructs the distributed estimation algorithm under the consensus+innovations framework [13], where a consensus protocol is necessary to fuse the messages transmitted in the network. Therefore, the SC consensus algorithm [14] is firstly introduced as the foundation of our distributed estimation algorithm.

3.1. The SC consensus protocol [14]. In [14], we consider the first order multi-agent system

$$\mathtt{x}_{i,k} = \mathtt{x}_{i,k-1} + \mathtt{u}_{i,k},\ \forall i = 1, \ldots, N, \tag{3.1}$$

where $\mathtt{x}_{i,k} \in \mathbb{R}$ is the agent i's state, and $\mathtt{u}_{i,k} \in \mathbb{R}$ is the input to be designed. The SC consensus protocol for the system (3.1) is given as in Algorithm 3.1.

Algorithm 3.1 The SC consensus protocol

Input: initial state sequence $\{x_{i,0}\}$, threshold C, step-size sequence $\{\alpha_k\}$.
Output: state sequence $\{\mathtt{x}_{i,k}\}$.
for $k = 1, 2, \ldots,$ **do**
Encoding: The agent i generates the binary-valued message as

$$\mathtt{s}_{i,k} = \begin{cases} 1, & \text{if } \mathtt{x}_{i,k} + \mathtt{d}_{i,k} < C; \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathtt{d}_{i,k}$ is the noise.
Consensus: The agent i receives the binary-valued messages $\mathtt{s}_{j,k}$ for all $j \in \mathcal{N}_i$, and updates its states by

$$\mathtt{x}_{i,k} = \mathtt{x}_{i,k-1} + \alpha_k \sum_{j \in \mathcal{N}_i} a_{ij} \left(\mathtt{s}_{i,k-1} - \mathtt{s}_{j,k-1} \right). \tag{3.2}$$

end for

The effectiveness of Algorithm 3.1 is analyzed in [14]. One of the main results is shown below.

THEOREM 3.1 (Theorem 1 of [14]). *Assume that the communication graph is connected, $\sum_{k=1}^{\infty} \alpha_k = \infty$, $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$, and the noise sequence $\{\mathtt{d}_{i,k}\}$ is independent and identically distributed (i.i.d.) with a strictly increasing distribution function $F(\cdot)$. Then, for Algorithm 3.1, we have $\lim_{k\to\infty} \mathtt{x}_{i,k} = \frac{1}{N}\sum_{j=1}^{N} x_{j,0}$ almost surely.*

Remark 3.2. Theorem 3.1 shows that Algorithm 3.1 can achieve the almost sure consensus. Therefore, Algorithm 3.1 can be used to solve the information transmission problem of distributed identification under binary-valued communications.

Remark 3.3. The design idea of Algorithm 3.1 is based on the comparison of the binary-valued messages $\mathtt{s}_{i,k}$ and $\mathtt{s}_{j,k}$. If $\mathtt{s}_{i,k} - \mathtt{s}_{j,k} = 1$, then $\mathtt{s}_{i,k} = 1$ and $\mathtt{s}_{j,k} = 0$. From the distributions of $\mathtt{s}_{i,k}$ and $\mathtt{s}_{j,k}$, one can get that $\mathtt{x}_{i,k}$ is more likely to be less than $\mathtt{x}_{j,k}$. Therefore, in Algorithm 3.1, $\mathtt{x}_{i,k}$ increases, and $\mathtt{x}_{j,k}$ decreases. Conversely, if $\mathtt{s}_{i,k} - \mathtt{s}_{j,k} = -1$, then $\mathtt{x}_{i,k}$ decreases, and $\mathtt{x}_{j,k}$ increases.

Remark 3.4. The noise $\mathtt{d}_{i,k}$ with strictly increasing distribution function is necessary for Algorithm 3.1. Without such a noise, the states $\mathtt{x}_{i,k}$ will keep constant if all the states are greater (or smaller) than the threshold C, and hence, the consensus may not be achieved. With the noise $\mathtt{d}_{i,k}$, $\mathbb{E}\left[\mathtt{s}_{i,k}|\mathtt{x}_{i,k}\right]$ is strictly decreasing with $\mathtt{x}_{i,k}$. Therefore, when $\mathtt{x}_{i,k} \neq \mathtt{x}_{j,k}$, the stochastic properties of $\mathtt{s}_{i,k}$ and $\mathtt{s}_{j,k}$ are different even if $\mathtt{x}_{i,k}$ and $\mathtt{x}_{j,k}$ are all greater (or smaller) than the threshold C. The consensus can be thereby achieved.

3.2. The SC-based distributed estimation algorithm. The subsection propose the SC-based distributed estimation algorithm in Algorithm 3.2.

Algorithm 3.2 The SC-based distributed estimation algorithm.

Input: initial estimate sequence $\{\hat{\theta}_{i,0}\}$, event-triggered coefficient sequence $\{\nu_{ij}\}$ with $\nu_{ij} = \nu_{ji} \geq 0$, noise coefficient sequence $\{b_{ij}\}$ with $b_{ij} = b_{ji} > 0$, step-size sequences $\{\alpha_{ij,k}\}$ with $\alpha_{ij,k} = \alpha_{ji,k} > 0$ and $\{\beta_{i,k}\}$ with $\beta_{i,k} > 0$.

Output: estimate sequence $\{\hat{\theta}_{i,k}\}$.

for $k = 1, 2, \ldots,$ **do**

Compressing: If $k = nq + l$ for some $q \in \mathbb{N}$ and $l \in \{1, \ldots, n\}$, then the sensor i generates φ_k as the n-dimensional vector whose l-th element is 1 and the others are 0. The sensor i uses φ_k to compress the previous local estimate $\hat{\theta}_{i,k-1}$ into the scalar $\mathtt{x}_{i,k} = \varphi_k^\top \hat{\theta}_{i,k-1}$.

Encoding: The sensor i generates the dithering noise $\mathtt{d}_{i,k}$ with Laplacian distribution $Lap(0, 1)$. Then, the sensor i generates the binary-valued message for the neighbor j

$$\mathtt{s}_{ij,k} = \begin{cases} 1, & \text{if } \mathtt{x}_{i,k} + b_{ij}\mathtt{d}_{i,k} > 0; \\ -1, & \text{otherwise.} \end{cases}$$

Data Transmission: Set $C_{ij,k} = \nu_{ij} b_{ij} \ln k$. If $|\mathtt{x}_{i,k} + b_{ij}\mathtt{d}_{i,k}| > C_{ij,k}$, then the sensor i sends the 1 bit message $\mathtt{s}_{ij,k}$ to the neighbor j. Otherwise, the sensor i does not send any message to the neighbor j.

Data Receiving: If the sensor i receives 1 bit message $\mathtt{s}_{ji,k}$ from its neighbor j, then set $\hat{\mathtt{s}}_{ji,k} = \mathtt{s}_{ji,k}$. Otherwise, set $\hat{\mathtt{s}}_{ji,k} = 0$.

Information fusion: Apply the modified Algorithm 3.1 to fuse the neighborhood information.

$$\check{\theta}_{i,k} = \hat{\theta}_{i,k-1} + \varphi_k \sum_{j \in \mathcal{N}_i} \alpha_{ij,k} a_{ij} \left(\hat{\mathtt{s}}_{ji,k} - G_{ij,k}(\mathtt{x}_{i,k}) \right) \tag{3.3}$$

where $G_{ij,k}(x) = F((x - C_{ij,k})/b_{ij}) - F((-x - C_{ij,k})/b_{ij})$, and $F(\cdot)$ is the distribution function of $Lap(0, 1)$.

Estimate update: Use the observation $\mathtt{y}_{i,k}$ to update the local estimate.

$$\hat{\theta}_{i,k} = \check{\theta}_{i,k} + \beta_{i,k} H_{i,k}^\top \left(\mathtt{y}_{i,k} - H_{i,k} \hat{\theta}_{i,k-1} \right). \tag{3.4}$$

end for

In Algorithm 3.2, dithering noise $\mathtt{d}_{i,k}$ is used for the encoding step and the event-triggered condition. The independence assumption for $\mathtt{d}_{i,k}$ is required.

Assumption 3.5. $\mathtt{d}_{i,k}$ and $\mathtt{d}_{j,t}$ are independent when $k \neq t$ or $i \neq j$. And, $\mathtt{d}_{i,k}$ and $\mathtt{w}_{j,t}$ are independent for all $i, j \in \mathcal{V}$ and $k, t \in \mathbb{N}$.

Following remarks are given for Algorithm 3.2.

Remark 3.6. The requirement that $\alpha_{ij,k} = \alpha_{ji,k}$ in Algorithm 3.2 is weak among existing literature. In the distributed estimation algorithms in [12, 13, 19, 30], it is required that $\alpha_{ij,k} = \alpha_{i'j',k}$ for all $(i, j), (i', j') \in \mathcal{E}$. He et al. [10] and Zhang and Zhang [37] relax this condition, but still require that $\lim_{k\to\infty} \frac{\alpha_{ij,k}}{\alpha_{i'j',k}} = 1$ for all $(i,j),(i',j') \in \mathcal{E}$, and hence the step-sizes $\alpha_{ij,k}$ converge to 0 with the same order. For comparison, in Algorithm 3.2, $\alpha_{ij,k} = \alpha_{i'j',k}$ is required only when $i = j'$ and $j = i'$, which is more easily implemented since it only requires the communication between adjacent sensors i and j, and the step-sizes $\alpha_{ij,k}$ in Algorithm 3.2 are allowed to converge to 0 with different orders. Here, we give one of the techniques to achieve $\alpha_{ij,k} = \alpha_{ji,k}$, which is a two-step protocol before running Algorithm 3.2. Firstly, the operators of the sensors i and j select positive numbers $\bar{\alpha}_{ij,1}$, $\bar{\gamma}_{ij}$ and $\bar{\alpha}_{ji,1}$, $\bar{\gamma}_{ji}$, respectively, and then transmit the selected numbers to each other. Secondly, set $\alpha_{ij,k} = \alpha_{ji,k} = \frac{\alpha_{ij,1}}{k^{\gamma_{ij}}}$, where $\alpha_{ij,1} = \frac{\bar{\alpha}_{ij,1}+\bar{\alpha}_{ji,1}}{2}$ and $\gamma_{ij} = \frac{\bar{\gamma}_{ij}+\bar{\gamma}_{ji}}{2}$. By using this technique, it requires only finite bits of communications to achieve $\alpha_{ij,k} = \alpha_{ji,k}$ if $\bar{m}\bar{\alpha}_{ij,1}$, $\bar{m}\bar{\gamma}_{ij}$, $\bar{m}\bar{\alpha}_{ji,1}$, $\bar{m}\bar{\gamma}_{ji}$ are all integers for some positive $\bar{m}$. Similar techniques can be applied to achieve $\nu_{ij} = \nu_{ji}$ and $b_{ij} = b_{ji}$ in Algorithm 3.2.

Remark 3.7. A new stochastic event-triggered mechanism is applied to Algorithm 3.2. The main idea is to use the dithering noises and the expanding triggering thresholds. When $\nu_{ij} > 0$, the threshold $C_{ij,k}$ goes to infinity. Hence, the probability that $|\mathtt{x}_{i,k} + b_{ij}\mathtt{d}_{i,k}| > C_{ij,k}$ decays to zero, which implies that the communication frequency is reduced.

Remark 3.8. The stochastic event-triggered mechanism used in Algorithm 3.2 is significantly different from existing ones. When the information is not transmitted at a certain moment, the traditional event-triggered mechanisms [10] use the recently received message as an approximation of the untransmitted message. Note that in the binary-valued communication case, 1 and -1 represent opposite information. Then, in this case, approximation technique of [10] can only be used when the recently received message is the same as the untransmitted message. This constraint makes it difficult to reduce communication frequency to zero through event-triggered mechanisms. To overcome the difficulty, a new approximation method is used in Algorithm 3.2. When the information is not transmitted at a certain moment, our stochastic event-triggered mechanism uses 0 as an approximation of the untransmitted information. The approximation technique expands the binary-valued message $\mathtt{s}_{ji,k}$ to triple-valued message $\hat{\mathtt{s}}_{ji,k}$. The message $\hat{\mathtt{s}}_{ji,k}$ contains information on whether $\mathtt{s}_{ji,k}$ is transmitted or not. Hence, the statistical properties of whether $\mathtt{s}_{ji,k}$ is transmitted can be better utilized.

Remark 3.9. In Algorithm 3.2, the dithering noise $\mathtt{d}_{i,k}$ is artificial, and generated under a given distribution function. The necessity of introducing $\mathtt{d}_{i,k}$ is similar to that in Algorithm 3.1, which has been explained in Remark 3.4. For similar reasons, dithering noises are often used to avoid the influence of quantization error [2, 9, 31]. Besides, in Algorithm 3.2, the dithering noise $\mathtt{d}_{i,k}$ is not necessarily Laplacian distributed. $\mathtt{d}_{i,k}$ can be any other types with continuous and strictly increasing distribution $F(\cdot)$, including Gaussian noises and the heavy-tailed noises [19]. For the polynomial decaying rate of $\mathtt{B}(k)$, the triggering threshold $C_{ij,k}$ can be changed accordingly.

Remark 3.10. In (3.3), we use $G_{ij,k}(\mathtt{x}_{i,k})$ to replace $\hat{\mathtt{s}}_{ij,k}$ in order to reduce the variances of the estimates, because $\mathbb{E}\left[\hat{\mathtt{s}}_{ij,k}|\mathcal{F}_{k-1}\right] = G_{ij,k}(\mathtt{x}_{i,k})$, where $\mathcal{F}_k = \sigma(\{\mathtt{w}_{i,t}, \mathtt{d}_{i,t} : i = 1,\dots,N,\ 1 \le t \le k\})$.

4. Convergence analysis. The convergence properties of Algorithm 3.2 is analyzed in this section. The almost sure convergence and mean square convergence are obtained in Subsection 4.1. Then, the almost sure convergence rate is calculated in Subsection 4.2.

4.1. Convergence. This subsection focuses on the almost sure and mean square convergence of Algorithm 3.2. The following theorem gives a new step-size condition, where the step-sizes are allowed to converge to zero with different orders, and the estimates of Algorithm 3.2 are proved to converge to the true value almost surely.

THEOREM 4.1. *Suppose the step-size sequences* $\{\alpha_{ij,k}\}$ *and* $\{\beta_{i,k}\}$ *satisfy*

i) $\sum_{k=1}^{\infty}\alpha_{ij,k}^2<\infty$ *and* $\alpha_{ij,k+1}=O\left(\alpha_{ij,k}\right)$ *for all* $(i,j)\in\mathcal{E}$*;*

ii) $\sum_{k=1}^{\infty}\beta_{i,k}^2<\infty$ *and* $\beta_{i,k+1}=O\left(\beta_{i,k}\right)$ *for all* $\forall i\in\mathcal{V}$*;*

iii) $\sum_{k=1}^{\infty}z_k=\infty$ *for* $z_k=\min\left\{\frac{\alpha_{ij,k}}{k^{\nu_{ij}}},(i,j)\in\mathcal{E};\beta_{i,k},i\in\mathcal{V}\right\}$*.*

Then, under Assumptions 2.1*,* 2.3*,* 2.5*, and* 3.5*, the estimate* $\hat{\mathtt{\theta}}_{i,k}$ *in Algorithm* 3.2 *converges to the true value* θ *almost surely.*

Proof. By $\mathbb{E}\left[\hat{\mathtt{s}}_{ji,k}|\mathcal{F}_{k-1}\right]=G_{ji,k}(\mathtt{x}_{j,k})$, one can get

$$
\begin{aligned}
(4.1)\qquad &\mathbb{E}\left[\left(\hat{\mathtt{s}}_{ji,k}-G_{ji,k}(\mathtt{x}_{j,k-1})\right)^2\Big|\mathcal{F}_{k-1}\right]\\
=&\mathbb{E}\left[\hat{\mathtt{s}}_{ji,k}^2\Big|\mathcal{F}_{k-1}\right]-G_{ij,k}^2(\mathtt{x}_{j,k})\\
=&F((\mathtt{x}_{j,k}-C_{ji,k})/b_{ji})+F((-\mathtt{x}_{j,k}-C_{ji,k})/b_{ji})-G_{ji,k}^2(\mathtt{x}_{j,k}),
\end{aligned}
$$

where the σ-algebra $\mathcal{F}_{k-1}$ is defined in Remark 3.10. Besides by the Lagrange mean value theorem [41], given $(i,j)\in\mathcal{E}$, there exists $\xi_{ij,k}$ between $\mathtt{x}_{i,k}$ and $\mathtt{x}_{j,k}$ such that

$$
G_{ji,k}(\mathtt{x}_{j,k})-G_{ij,k}(\mathtt{x}_{i,k})=g_{ij,k}(\xi_{ij,k})\left(\mathtt{x}_{j,k}-\mathtt{x}_{i,k}\right),
$$

where

$$
g_{ij,k}(x)=g_{ji,k}(x)=\left(f\left(\frac{x-C_{ij,k}}{b_{ij}}\right)+f\left(\frac{-x-C_{ij,k}}{b_{ij}}\right)\right)\Big/b_{ij}\ ,
$$

and $f(\cdot)$ is the density function of $Lap(0,1)$. Denote $\tilde{\mathtt{\theta}}_{i,k}=\hat{\mathtt{\theta}}_{i,k}-\theta$. Then, it holds that

$$
\begin{aligned}
\mathbb{E}\left[\|\tilde{\mathtt{\theta}}_{i,k}\|^2\Big|\mathcal{F}_{k-1}\right]=&\|\tilde{\mathtt{\theta}}_{i,k-1}\|^2-2\beta_{i,k}\left(H_{i,k}\tilde{\mathtt{\theta}}_{i,k-1}\right)^2\\
&+2\varphi_k^\top\tilde{\mathtt{\theta}}_{i,k-1}\sum_{j\in\mathcal{N}_i}\alpha_{ij,k}a_{ij}g_{ij,k}(\xi_{ij,k})\left(\mathtt{x}_{j,k}-\mathtt{x}_{i,k}\right)\\
&+O\left(\beta_{i,k}^2\left(\|\tilde{\mathtt{\theta}}_{i,k-1}\|^2+1\right)+\sum_{j\in\mathcal{N}_i}\alpha_{ij,k}^2\right).
\end{aligned}
$$

Denote $\tilde{\mathtt{x}}_{i,k}=\varphi_k^\top\tilde{\mathtt{\theta}}_{i,k-1}=\mathtt{x}_{i,k}-\varphi_k^\top\theta$ and $\tilde{\mathtt{X}}_k=[\tilde{\mathtt{x}}_{1,k},\ldots,\tilde{\mathtt{x}}_{N,k}]^\top$. Then, one can get

$$
\begin{aligned}
(4.2)\qquad &\sum_{i=1}^N 2\varphi_k^\top\tilde{\mathtt{\theta}}_{i,k-1}\sum_{j\in\mathcal{N}_i}\alpha_{ij,k}a_{ij}g_{ij,k}(\xi_{ij,k})\left(\mathtt{x}_{j,k}-\mathtt{x}_{i,k}\right)\\
=&\sum_{i=1}^N 2\tilde{\mathtt{x}}_{i,k}\sum_{j\in\mathcal{N}_i}\alpha_{ij,k}a_{ij}g_{ij,k}(\xi_{ij,k})\left(\tilde{\mathtt{x}}_{j,k}-\tilde{\mathtt{x}}_{i,k}\right)=-2\tilde{\mathtt{X}}_k^\top\mathtt{L}_{G,k}\tilde{\mathtt{X}}_k,
\end{aligned}
$$

where $\mathtt{L}_{G,k}=(\mathtt{l}_{ij,k}^G)_{N\times N}$ is a Laplacian matrix with $\mathtt{l}_{ii,k}^G=\sum_{j\in\mathcal{N}_i}\alpha_{ij,k}a_{ij}g_{ij,k}(\xi_{ij,k})$

and $\mathtt{l}^G_{ij,k} = -\alpha_{ij,k} a_{ij} g_{ij,k}(\xi_{ij,k})$ for $i \neq j$. Therefore, we have

$$\mathbb{E}\left[\sum_{i=1}^{N} \|\tilde{\theta}_{i,k}\|^2 \middle| \mathcal{F}_{k-1}\right] = \sum_{i=1}^{N} \|\tilde{\theta}_{i,k-1}\|^2 - 2\sum_{i=1}^{N} \beta_{i,k}\left(H_{i,k}\tilde{\theta}_{i,k-1}\right)^2 - 2\tilde{\mathtt{X}}_k^\top \mathtt{L}_{G,k}\tilde{\mathtt{X}}_k + O\left(\sum_{i=1}^{N} \beta_{i,k}^2\left(\|\tilde{\theta}_{i,k-1}\|^2 + 1\right) + \sum_{(i,j)\in\mathcal{E}} \alpha_{ij,k}^2\right). \tag{4.3}$$

Then, by Theorem 1.3.2 of [8], $\sum_{i=1}^N \|\tilde{\theta}_{i,k}\|^2$ converges to a finite value almost surely, and

$$\sum_{k=1}^{\infty}\left(\sum_{i=1}^{N} \beta_{i,k}\left(H_{i,k}\tilde{\theta}_{i,k-1}\right)^2 + \tilde{\mathtt{X}}_k^\top \mathtt{L}_{G,k}\tilde{\mathtt{X}}_k\right) < \infty, \text{ a.s.}, \tag{4.4}$$

By the convergence of $\sum_{i=1}^N \|\tilde{\theta}_{i,k}\|^2$, $\tilde{\mathtt{x}}_{i,k} = \varphi_k^\top \tilde{\theta}_{i,k}$ is uniformly bounded almost surely. Then, by Lemma A.1 in Appendix A, it holds that

$$\underline{\mathtt{g}} := \inf_{(i,j)\in\mathcal{E}, k\in\mathbb{N}} k^{\nu_{ij}} g_{ij,k}(\xi_{ij,k}) > 0, \text{ a.s.} \tag{4.5}$$

Hence, one can get

$$\mathtt{L}_{G,k} \geq \left(\min_{(i,j)\in\mathcal{E}} \frac{\alpha_{ij,k}}{k^{\nu_{ij}}}\right) \underline{\mathtt{g}}\lambda_2(\mathcal{L})\left(I_N - J_N\right), \tag{4.6}$$

where $\lambda_2(\mathcal{L})$ is the second smallest eigenvalue of $\mathcal{L}$, and $J_N = \frac{1}{N}\mathbf{1}_N^\top \mathbf{1}_N$.

Denote

$$\begin{aligned}
\tilde{\Theta}_k &= \mathrm{col}\{\tilde{\theta}_{1,k}, \ldots, \tilde{\theta}_{N,k}\}, \ \mathbb{H}_k = \mathrm{diag}\{H_{1,k}^\top H_{1,k}, \ldots, H_{N,k}^\top H_{N,k}\}, \\
\mathbb{H}_{\beta,k} &= \mathrm{diag}\{\beta_{1,k} H_{1,k}^\top H_{1,k}, \ldots, \beta_{N,k} H_{N,k}^\top H_{N,k}\}, \\
\Phi_k &= \mathbb{H}_k + \underline{\mathtt{g}}\lambda_2(\mathcal{L})\left(I_N - J_N\right) \otimes \varphi_k\varphi_k^\top, \\
\mathtt{W}_k &= \mathrm{col}\{\beta_{1,k} H_{1,k}^\top \mathtt{w}_{1,k}, \ldots, \beta_{N,k} H_{N,k}^\top \mathtt{w}_{N,k}\}, \\
&\quad + \left[\left(\varphi_k \sum_{j\in\mathcal{N}_1} \alpha_{1j,k} a_{1j}(\hat{\mathtt{s}}_{j1,k} - G_{j1,k}(\mathtt{x}_{j,k}))\right)^\top, \ldots,\right. \\
&\quad \left.\left(\varphi_k \sum_{j\in\mathcal{N}_N} \alpha_{Nj,k} a_{Nj} a_{Nj}(\hat{\mathtt{s}}_{jN,k} - G_{jN,k}(\mathtt{x}_{j,k}))\right)^\top\right]^\top.
\end{aligned}$$

Then, $\mathtt{W}_k$ is $\mathcal{F}_k$-measurable, and

$$\begin{aligned}
\tilde{\Theta}_k &= \left(I_{N\times n} - \mathbb{H}_{\beta,k} - \mathtt{L}_{G,k} \otimes \varphi_k\varphi_k^\top\right)\tilde{\Theta}_{k-1} + \mathtt{W}_k, \\
\mathbb{E}\left[\mathtt{W}_k|\mathcal{F}_{k-1}\right] &= 0, \ \mathbb{E}\left[\|\mathtt{W}_k\|^2\middle|\mathcal{F}_{k-1}\right] = O\left(\sum_{i=1}^{N} \beta_{i,k}^2 + \sum_{(i,j)\in\mathcal{E}} \alpha_{ij,k}^2\right).
\end{aligned} \tag{4.7}$$

By the almost sure uniform boundedness of $\tilde{\Theta}_k$ and (4.7), one can get

$$\mathtt{P}_k := \frac{\tilde{\Theta}_k - \tilde{\Theta}_{k-1} - \mathtt{W}_k}{\sum_{i=1}^N \beta_{i,k} + \sum_{(i,j)\in\mathcal{E}} \alpha_{ij,k}} \tag{4.8}$$

is $\mathcal{F}_{k-1}$-measurable and almost surely uniformly bounded. By (4.6), it holds that

$$\sum_{i=1}^{N}\beta_{i,k}\left(H_{i,k}\tilde{\theta}_{i,k-1}\right)^2+\tilde{\mathtt{X}}_k^\top \mathtt{L}_{G,k}\tilde{\mathtt{X}}_k\geq z_k\tilde{\Theta}_k^\top\Phi_k\tilde{\Theta}_k. \tag{4.9}$$

Besides by Lemma 5.4 of [36], there exists $\underline{\mathtt{H}}>0$ almost surely such that

$$\sum_{t=k-np+1}^{k}\Phi_t=\sum_{t=k-np+1}^{k}\mathbb{H}_t+\underline{\mathtt{g}}\lambda_2(\mathcal{L})p\left(I_N-J_N\right)\otimes I_n\geq\underline{\mathtt{H}}. \tag{4.10}$$

By (4.8), one can get

$$\begin{aligned}
&\sum_{t=npr+1}^{npr+np}z_t\tilde{\Theta}_{pr+p}^\top\Phi_t\tilde{\Theta}_{npr+np}-\sum_{t=npr+1}^{npr+np}z_t\tilde{\Theta}_t^\top\Phi_t\tilde{\Theta}_t \\
&=\sum_{t=npr+1}^{npr+np}z_t\sum_{l=t+1}^{npr+np}\left(\tilde{\Theta}_l^\top\Phi_t\tilde{\Theta}_l-\tilde{\Theta}_{l-1}^\top\Phi_t\tilde{\Theta}_{l-1}\right)\\
&=\sum_{t=npr+1}^{npr+np}z_t\sum_{l=t+1}^{npr+np}\left(2\mathtt{W}_l^\top\Phi_t\tilde{\Theta}_{l-1}+\mathtt{W}_l^\top\Phi_t\mathtt{W}_l\right)\\
&\quad+O\left(\sum_{t=npr+1}^{npr+np}z_t\left(\sum_{l=t+1}^{npr+np}\sum_{i=1}^{N}\beta_{i,l}+\sum_{l=t+1}^{npr+np}\sum_{(i,j)\in\mathcal{E}}\alpha_{ij,l}\right)\right)\\
&\quad+\sum_{t=npr+1}^{npr+np}2z_t\left(\sum_{l=t+1}^{npr+np}\sum_{i=1}^{N}\beta_{i,l}\mathtt{W}_l^\top\Phi_t\mathtt{P}_l+\sum_{l=t+1}^{npr+np}\sum_{(i,j)\in\mathcal{E}}\alpha_{ij,l}\mathtt{W}_l^\top\Phi_t\mathtt{P}_l\right),\ \text{a.s.}
\end{aligned} \tag{4.11}$$

By $\sum_{k=1}^{\infty}\alpha_{ij,k}^2<\infty$ and $\sum_{k=1}^{\infty}\beta_{i,k}^2<\infty$, we have

$$\sum_{r=1}^{\infty}\sum_{t=npr+1}^{npr+np}z_t\left(\sum_{l=t+1}^{npr+np}\sum_{i=1}^{N}\beta_{i,l}+\sum_{l=t+1}^{npr+np}\sum_{(i,j)\in\mathcal{E}}\alpha_{ij,l}\right)<\infty.$$

By Theorem 1.3.10 of [8], one can get

$$\sum_{r=1}^{\infty}\sum_{t=npr+1}^{npr+np}\sum_{l=t+1}^{npr+np}2z_t\mathtt{W}_l^\top\Phi_t\tilde{\Theta}_{l-1}<\infty,\ \text{a.s.},$$

$$\sum_{r=1}^{\infty}\sum_{t=npr+1}^{npr+np}\sum_{l=t+1}^{npr+np}2z_t\left(\sum_{i=1}^{N}\beta_{i,l}\mathtt{W}_l^\top\Phi_t\mathtt{P}_l+\sum_{(i,j)\in\mathcal{E}}\alpha_{ij,l}\mathtt{W}_l^\top\Phi_t\mathtt{P}_l\right)<\infty,\ \text{a.s.}$$

By Theorem 1.3.9 of [8] with $\alpha=1$, we have

$$\sum_{r=1}^{\infty}\sum_{t=npr+1}^{npr+np}\sum_{l=t+1}^{npr+np}z_t\mathbb{E}\|\mathtt{W}_l\|^2\cdot\frac{1}{\mathbb{E}\|\mathtt{W}_l\|^2}\left(\mathtt{W}_l^\top\Phi_t\mathtt{W}_l-\mathbb{E}\left[\mathtt{W}_l^\top\Phi_t\mathtt{W}_l\middle|\mathcal{F}_{l-1}\right]\right)<\infty,\ \text{a.s.}$$

Besides, $\mathbb{E}\left[\mathtt{W}_l^\top\Phi_t\mathtt{W}_l\middle|\mathcal{F}_{l-1}\right]=O\left(\left(\sum_{i=1}^{N}\beta_{i,l}+\sum_{(i,j)\in\mathcal{E}}\alpha_{ij,l}\right)^2\right)$ almost surely. Then,

$$\sum_{r=1}^{\infty}\sum_{t=npr+1}^{npr+np}z_t\sum_{l=t+1}^{npr+np}\mathbb{E}\left[\mathtt{W}_l^\top\Phi_t\mathtt{W}_l\middle|\mathcal{F}_{l-1}\right]<\infty,\ \text{a.s.}$$

Therefore by (4.4), (4.9)-(4.11), we have

$$
\begin{aligned}
&\underline{\mathtt{H}}\sum_{r=1}^{\infty}\left(\min_{npr+1\le t\le npr+np} z_t\right)\|\tilde{\Theta}_{npr+np}\|^2\\
\le&\sum_{r=1}^{\infty}\left(\min_{npr+1\le t\le npr+np} z_t\right)\tilde{\Theta}_{npr+np}^{\top}\left(\sum_{t=npr+1}^{npr+np}\Phi_t\right)\tilde{\Theta}_{npr+np}\\
\le&\sum_{r=1}^{\infty}\sum_{t=npr+1}^{npr+np} z_t\tilde{\Theta}_{npr+np}^{\top}\Phi_t\tilde{\Theta}_{npr+np}=\sum_{k=1}^{\infty} z_k\tilde{\Theta}_k^{\top}\Phi_k\tilde{\Theta}_k+O(1)\\
\le&\sum_{k=1}^{\infty}\left(\sum_{i=1}^{N}\beta_{i,k}\left(H_{i,k}\tilde{\theta}_{i,k-1}\right)^2+\tilde{\mathtt{X}}_k^{\top}\mathtt{L}_{G,k}\tilde{\mathtt{X}}_k\right)+O(1)<\infty,\ \text{a.s.}
\end{aligned}
$$

Then, by Lemma A.2 in Appendix A, there exist $\mathtt{k}_1<\mathtt{k}_2<\cdots$ such that $\lim_{t\to\infty}\|\tilde{\Theta}_{\mathtt{k}_t}\|^2 = 0$ almost surely. Note that $\sum_{i=1}^{N}\|\tilde{\theta}_{i,k}\|^2=\|\tilde{\Theta}_k\|^2$ converges to a finite value. Then, the value is 0, which proves the theorem. ∎

Remark 4.2. The estimates of Algorithm 3.2 can converge to the true value because the algorithm is designed by using the idea of stochastic approximation [6]. In Algorithm 3.2, $\hat{\mathtt{s}}_{ji,k}-G_{ij,k}(\mathtt{x}_{i,k})=G_{ij,k}(\mathtt{x}_{j,k})-G_{ij,k}(\mathtt{x}_{i,k})+\hat{\mathtt{s}}_{ji,k}-G_{ij,k}(\mathtt{x}_{j,k})$ and $\mathtt{y}_{i,k}-H_{i,k}\hat{\theta}_{i,k-1}=-H_{i,k}\tilde{\theta}_{i,k-1}+\mathtt{w}_{i,k}$, where $\hat{\mathtt{s}}_{ji,k}-G_{ij,k}(\mathtt{x}_{j,k})$ and $\mathtt{w}_{i,k}$ are martingale difference with bounded variance, and

$$G_{ij,k}(\varphi_k^{\top}\hat{\theta}_j)-G_{ij,k}(\varphi_k^{\top}\hat{\theta}_i)=0,\ \forall(i,j)\in\mathcal{E},k\in\mathbb{N};\ H_{i,k}(\hat{\theta}_i-\theta)=0,\forall i\in\mathcal{V},k\in\mathbb{N}$$

holds if and only if $\hat{\theta}_i=\theta$ for all i. Besides, under i) and ii) of Theorem 4.1, the step-sizes converge to 0. These algorithm characteristics based on stochastic approximation enable the estimates to converge to the true value [6].

Remark 4.3. If $\alpha_{ij,k}$ and $\beta_{i,k}$ are all polynomial, iii) of Theorem 4.1 is equivalent to $\sum_{k=1}^{\infty}\frac{\alpha_{ij,k}}{k^{\nu_{ij}}}=\infty$ for all $(i,j)\in\mathcal{E}$ and $\sum_{k=1}^{\infty}\beta_{i,k}=\infty$ for all $i\in\mathcal{V}$. Under this case, the step-sizes can be designed in a distributed manner.

Remark 4.4. Note that $2\sum_{t=1}^{k}\frac{\alpha_{ij,k}}{k^{\nu_{ij}}}\le\sum_{t=1}^{k}\alpha_{ij,k}^2+\sum_{t=1}^{k}\frac{1}{k^{2\nu_{ij}}}$. Then, the conditions i) and iii) imply $\nu_{ij}\le\frac{1}{2}$. Especially, if $\alpha_{ij,k}$ is polynomial, then $\nu_{ij}<\frac{1}{2}$.

The following theorem proves the mean square convergence of Algorithm 3.2.

THEOREM 4.5. *Under the condition of Theorem* 4.1*, the estimate* $\hat{\theta}_{i,k}$ *in Algorithm* 3.2 *converges to the true value* θ *in the mean square sense.*

Proof. Since we have proved the almost sure convergence of Algorithm 3.2, by Theorem 2.6.4 of [26], it suffices to prove the uniform integrability of the algorithm. Here, we continue to use the notations of $\mathtt{L}_{G,k}$, $\tilde{\Theta}_k$, $\mathbb{H}_{\beta,k}$, and $\mathtt{W}_k$ in the proof of Theorem 4.1.

Denote $\mathtt{A}_k=I_{N\times n}-\mathbb{H}_{\beta,k}-\mathtt{L}_{G,k}\otimes\varphi_k\varphi_k^{\top}$. When k is sufficiently large, $\|\mathtt{A}_k\|\le 1$. Then, by (4.7),

$$
\begin{aligned}
(4.12)\quad &\mathbb{E}\|\tilde{\Theta}_k\|^2\ln\left(1+\|\tilde{\Theta}_k\|^2\right)\\
\le\ &\mathbb{E}\left(\|\tilde{\Theta}_{k-1}\|^2+2\mathtt{W}_k^{\top}\mathtt{A}_k\tilde{\Theta}_{k-1}+\|\mathtt{W}_k\|^2\right)\ln\left(1+\|\tilde{\Theta}_{k-1}\|^2+2\mathtt{W}_k^{\top}\mathtt{A}_k\tilde{\Theta}_{k-1}+\|\mathtt{W}_k\|^2\right).
\end{aligned}
$$

By b) of Lemma A.3 in Appendix A,

$$
\begin{aligned}
(4.13)\qquad &\mathbb{E}\|\tilde{\Theta}_{k-1}\|^2 \ln\left(1+\|\tilde{\Theta}_{k-1}\|^2+2\mathtt{W}_k^\top \mathtt{A}_k\tilde{\Theta}_{k-1}+\|\mathtt{W}_k\|^2\right)\\
\le&\mathbb{E}\|\tilde{\Theta}_{k-1}\|^2 \ln\left(1+\|\tilde{\Theta}_{k-1}\|^2\right)+\mathbb{E}\frac{\|\tilde{\Theta}_{k-1}\|^2}{1+\|\tilde{\Theta}_{k-1}\|^2}\left(2\mathtt{W}_k^\top \mathtt{A}_k\tilde{\Theta}_{k-1}+\|\mathtt{W}_k\|^2\right)\\
\le&\mathbb{E}\|\tilde{\Theta}_{k-1}\|^2 \ln\left(1+\|\tilde{\Theta}_{k-1}\|^2\right)+\mathbb{E}\|\tilde{\Theta}_{k-1}\|^2\mathbb{E}\|\mathtt{W}_k\|^2.
\end{aligned}
$$

By a), c) and d) of Lemma A.3 in Appendix A,

$$
\begin{aligned}
(4.14)\qquad &\mathbb{E}2\mathtt{W}_k^\top \mathtt{A}_k\tilde{\Theta}_{k-1} \ln\left(1+\|\tilde{\Theta}_{k-1}\|^2+2\mathtt{W}_k^\top \mathtt{A}_k\tilde{\Theta}_{k-1}+\|\mathtt{W}_k\|^2\right)\\
\le&\mathbb{E}2\mathtt{W}_k^\top \mathtt{A}_k\tilde{\Theta}_{k-1} \ln\left(1+\|\tilde{\Theta}_{k-1}\|^2+\|\mathtt{W}_k\|^2\right)+\mathbb{E}\left(2\mathtt{W}_k^\top \mathtt{A}_k\tilde{\Theta}_{k-1}\right)^2\\
\le&\mathbb{E}2\mathtt{W}_k^\top \mathtt{A}_k\tilde{\Theta}_{k-1} \ln\left(1+\|\tilde{\Theta}_{k-1}\|^2\right)+4\mathbb{E}\|\mathtt{W}_k\|^2\mathbb{E}\|\tilde{\Theta}_{k-1}\|^2\\
&+\mathbb{E}2|\mathtt{W}_k^\top \mathtt{A}_k\tilde{\Theta}_{k-1}|\left(\ln\left(1+\|\tilde{\Theta}_{k-1}\|^2+\|\mathtt{W}_k\|^2\right)-\ln\left(1+\|\tilde{\Theta}_{k-1}\|^2\right)\right)\\
\le&\mathbb{E}2\|\mathtt{W}_k\|\|\tilde{\Theta}_{k-1}\|\ln\left(1+\|\mathtt{W}_k\|^2\right)+4\mathbb{E}\|\tilde{\Theta}_{k-1}\|^2\mathbb{E}\|\mathtt{W}_k\|^2\\
\le&O\left(\mathbb{E}\|\tilde{\Theta}_{k-1}\|\mathbb{E}\|\mathtt{W}_k\|^2\right)+4\mathbb{E}\|\tilde{\Theta}_{k-1}\|^2\mathbb{E}\|\mathtt{W}_k\|^2.
\end{aligned}
$$

By a) and d) of Lemma A.3 in Appendix A,

$$
\begin{aligned}
(4.15)\qquad &\mathbb{E}\|\mathtt{W}_k\|^2 \ln\left(1+\|\tilde{\Theta}_{k-1}\|^2+2\mathtt{W}_k^\top \mathtt{A}_k\tilde{\Theta}_{k-1}+\|\mathtt{W}_k\|^2\right)\\
\le&\mathbb{E}\|\mathtt{W}_k\|^2 \ln\left(1+2\|\tilde{\Theta}_{k-1}\|^2+2\|\mathtt{W}_k\|^2\right)\\
\le&\mathbb{E}\|\mathtt{W}_k\|^2 \ln\left(1+2\|\tilde{\Theta}_{k-1}\|^2\right)+\mathbb{E}\|\mathtt{W}_k\|^2\ln\left(1+2\|\mathtt{W}_k\|^2\right)\\
\le&2\mathbb{E}\|\tilde{\Theta}_{k-1}\|^2\mathbb{E}\|\mathtt{W}_k\|^2+O\left(\mathbb{E}\|\mathtt{W}_k\|^{\min\{\rho,4\}}\right),
\end{aligned}
$$

where ρ is given in Assumption 2.3. Taken the expectation over (4.3), we have $\mathbb{E}\|\tilde{\Theta}_k\|^2$ is uniformly bounded. By Lyapunov inequality [26], one can get $\mathbb{E}\|\tilde{\Theta}_k\|$ is also uniformly bounded. Besides, $\mathbb{E}\|\mathtt{W}_k\|^2=O\left(\left(\sum_{i=1}^N \beta_{i,k}^2+\sum_{(i,j)\in\mathcal{E}}\alpha_{ij,k}^2\right)\right)$, and $\mathbb{E}\|\mathtt{W}_k\|^{\min\{\rho,4\}}=O\left(\left(\sum_{i=1}^N \beta_{i,k}^{\min\{\rho,4\}}+\sum_{(i,j)\in\mathcal{E}}\alpha_{ij,k}^{\min\{\rho,4\}}\right)\right)$. Hence, (4.12)-(4.15) imply that $\mathbb{E}\|\tilde{\Theta}_k\|^2\ln\left(1+\|\tilde{\Theta}_k\|^2\right)$ is uniformly bounded. Note that

$$
\begin{aligned}
&\lim_{x\to\infty}\sup_{k\in\mathbb{N}}\int_{\{\|\tilde{\Theta}_k\|^2>x\}}\|\tilde{\Theta}_k\|^2\mathrm{d}\mathbb{P}\\
\le&\lim_{x\to\infty}\sup_{k\in\mathbb{N}}\frac{1}{\ln(1+x)}\int_{\{\|\tilde{\Theta}_k\|^2>x\}}\|\tilde{\Theta}_k\|^2\ln\left(1+\|\tilde{\Theta}_k\|^2\right)\mathrm{d}\mathbb{P}\\
\le&\lim_{x\to\infty}\sup_{k\in\mathbb{N}}\frac{1}{\ln(1+x)}\mathbb{E}\|\tilde{\Theta}_k\|^2\ln\left(1+\|\tilde{\Theta}_k\|^2\right)=0.
\end{aligned}
$$

Then, $\|\tilde{\Theta}_k\|^2$ is uniformly integrable. Hence, the theorem can be proved by Theorem 2.6.4 of [26] and Theorem 4.1. ∎

Remark 4.6. If (2.2) holds for any $\rho>0$, then similar to Theorem 4.5, we can prove the L^r convergence of Algorithm 3.2 for any positive integer r.

Remark 4.7. Under finite data rate, existing literature [25, 35] focuses on the mean square stability in terms of effectiveness, and gives the upper bounds of the mean square estimation errors for corresponding algorithms. There are two important breakthroughs in Theorems 4.1 and 4.5. Firstly, Theorem 4.5 shows that our algorithm can not only achieve mean square stability, but also can achieve mean square convergence. The mean square estimation errors of our algorithm can converge to zero. Secondly, Theorem 4.1 shows that the estimates of our algorithm can converge not only in the mean square sense, but also in the almost sure sense. The almost sure convergence property can better describe the characteristics of a single trajectory. When using our algorithm, there is no need to worry about the small probability event that the estimation errors do not converge to zero, as it will not occur almost surely.

4.2. Convergence rate. To quantitatively demonstrate the effectiveness, the following theorem calculates the almost sure convergence rate of Algorithm 3.2.

THEOREM 4.8. *In Algorithm* 3.2, *set* $\alpha_{ij,k}=\frac{\alpha_{ij,1}}{k^{\gamma_{ij}}}$ *and* $\beta_{i,k}=\frac{\beta_{i,1}}{k}$ *with*

i) $\alpha_{ij,1}=\alpha_{ji,1}>0$ *for all* $(i,j)\in\mathcal{E}$, *and* $\beta_{i,1}>0$ *for all* $i\in\mathcal{V}$;

ii) $1/2<\gamma_{ij}\le 1$ *and* $\nu_{ij}+\gamma_{ij}\le 1$ *for all* $(i,j)\in\mathcal{E}$.

Then, under Assumptions 2.1, 2.3, 2.5, *and* 3.5, *the almost sure convergence rate of the estimation error for the sensor* i *is*

$$\tilde{\theta}_{i,k}=\begin{cases} O\left(\frac{1}{k^{a}}\right), & \text{if } 2h-2a>1;\\ O\left(\frac{\ln k}{k^{h-1/2}}\right), & \text{if } 2h-2a=1;\\ O\left(\frac{\sqrt{\ln k}}{k^{h-1/2}}\right), & \text{if } 2h-2a<1,\end{cases}\quad a.s.,$$

where $h=\min_{(i,j)\in\mathcal{E}}\left(\frac{\nu_{ij}}{2}+\gamma_{ij}\right)$, $\lambda_2(\mathcal{L})$ *is defined in* (4.6), $\mathcal{E}'=\{(i,j)\in\mathcal{E}:\nu_{ij}+\gamma_{ij}=1\}$, *and*

$$a=\begin{cases}\frac{\delta(\min_{i\in\mathcal{V}}\beta_{i,1})}{N}, & \text{if } \mathcal{E}'=\varnothing;\\ \frac{\delta\lambda_2(\mathcal{L})(\min_{i\in\mathcal{V}}\beta_{i,1})\left(\min_{(i,j)\in\mathcal{E}'}\alpha_{ij,1}\frac{\exp(-\|\theta\|_1/b_{ij})}{b_{ij}}\right)}{2Nn\bar{H}^2(\min_{i\in\mathcal{V}}\beta_{i,1})+N\lambda_2(\mathcal{L})\left(\min_{(i,j)\in\mathcal{E}'}\alpha_{ij,1}\frac{\exp(-\|\theta\|_1/b_{ij})}{b_{ij}}\right)}, & \text{if } \mathcal{E}'\neq\varnothing.\end{cases}$$

Proof. The key of the proof is to use Lemma A.4 in Appendix A. Here, we continue to use the notations of $\mathtt{L}_{G,k}$, $\tilde{\Theta}_k$, $\mathbb{H}_k$, $\mathbb{H}_{\beta,k}$, Φ_k, and $\mathtt{W}_k$ in the proof of Theorem 4.1. Under the step-sizes in this theorem, by (4.7), one can get

$$\tilde{\Theta}_k=\left(I_{N\times n}-\frac{1}{k}\left(k\mathbb{H}_{\beta,k}+k\mathtt{L}_{G,k}\otimes\varphi_k\varphi_k^{\top}\right)\right)\tilde{\Theta}_{k-1}+\mathtt{W}_k. \tag{4.16}$$

Since $\mathbb{E}\left[(\hat{\mathtt{s}}_{ji,k}-G_{ji,k}(\mathtt{x}_{j,k}))^2\middle|\mathcal{F}_{k-1}\right]=O\left(\frac{1}{k^{\nu_{ij}}}\right)$ almost surely, we have

$$\mathbb{E}\left[\|\mathtt{W}\|_k^2\middle|\mathcal{F}_{k-1}\right]=O\left(\frac{1}{k^2}+\frac{1}{k^{\min_{(i,j)\in\mathcal{E}}(\nu_{ij}+2\gamma_{ij})}}\right)=O\left(\frac{1}{k^{2h}}\right),\ \text{a.s.}$$

Besides by (4.5), one can get

$$\mathtt{L}_{G,k}=O\left(\frac{1}{k^{\min_{(i,j)\in\mathcal{E}}(\nu_{ij}+\gamma_{ij})}}\right),\ \text{a.s.}$$

Therefore, we have $k\mathbb{H}_{\beta,k}+k\mathtt{L}_{G,k}\otimes\varphi_k\varphi_k^{\top}=O\left(k^{1-\min_{(i,j)\in\mathcal{E}}(\nu_{ij}+\gamma_{ij})}\right)$ almost surely.

Firstly, we show that $h \leq \min\left\{1, \frac{3+2h-2(1-\min_{(i,j)\in\mathcal{E}}(\nu_{ij}+\gamma_{ij}))}{3}\right\}$. Note that $h = \min_{(i,j)\in\mathcal{E}}\left(\frac{\nu_{ij}}{2}+\gamma_{ij}\right) \leq \min_{(i,j)\in\mathcal{E}}(\nu_{ij}+\gamma_{ij})$. Then, one can get $h \leq 1$ and

$$h < \frac{1+4h}{4} \leq \frac{3+2h-2\left(1-\min_{(i,j)\in\mathcal{E}}(\nu_{ij}+\gamma_{ij})\right)}{4}.$$

Secondly, we estimate the lower bound of $\frac{1}{np}\sum_{t=k-np+1}^{k}\left(t\mathbb{H}_{\beta,t}+t\mathtt{L}_{G,t}\otimes\varphi_t\varphi_t^\top\right)$. By (4.6) and (4.10), one can get

$$\sum_{t=k-np+1}^{k}\left(t\mathbb{H}_{\beta,t}+t\mathtt{L}_{G,t}\otimes\varphi_t\varphi_t^\top\right) \geq z_1\sum_{t=k-np+1}^{k}\Phi_t \geq \underline{\mathtt{H}} > 0, \text{ a.s.,}$$

where $z_1 = \min\{\alpha_{ij,1}, (i,j)\in\mathcal{E}; \beta_{i,1}, i\in\mathcal{V}\}$. Then, by Lemma A.4, $\tilde{\Theta}_k = O\left(\frac{1}{k^\psi}\right)$ for some $\psi > 0$ almost surely. Hence, by the Lagrange mean value theorem [41] and Lemma A.1, we have $g_{ij}(\xi_{ij,k}) - g_{ij}(\varphi_k^\top\theta) = O\left(\frac{1}{k^{\nu_{ij}+\psi}}\right)$ almost surely, which implies

$$g_{ij}(\xi_{ij,k}) \geq \frac{\exp\left(\frac{-|\varphi_k^\top\theta|-C_{ij,k}}{b_{ij}}\right)}{b_{ij}} + O\left(\frac{1}{k^{\nu_{ij}+\psi}}\right) \geq \frac{e^{-\|\theta\|_1/b_{ij}}}{b_{ij}k^{\nu_{ij}}} + O\left(\frac{1}{k^{\nu_{ij}+\psi}}\right), \text{a.s.} \tag{4.17}$$

By Assumption 2.1, (4.17), and Lemma 5.4 of [36], it holds that

$$\begin{aligned}
&\sum_{t=k-np+1}^{k}\left(t\mathbb{H}_{\beta,t}+t\mathtt{L}_{G,t}\otimes\varphi_t\varphi_t^\top\right) \\
\geq & \sum_{t=k-np+1}^{k}\left(t\mathbb{H}_{\beta,t}+R_t\left(I_N-J_N\right)\otimes\varphi_t\varphi_t^\top\right) \\
\geq & \sum_{t=k-np+1}^{k}\left(\min_{i\in\mathcal{V}}\beta_{i,1}\right)\mathbb{H}_t + \left(\min_{k-np+1\leq t\leq k}R_t\right)\left(I_N-J_N\right)\otimes\left(\sum_{t=k-np+1}^{k}\varphi_t\varphi_t^\top\right) \\
= & \sum_{t=k-np+1}^{k}\left(\left(\min_{i\in\mathcal{V}}\beta_{i,1}\right)\mathbb{H}_t + \frac{1}{n}\left(\min_{k-np+1\leq t\leq k}R_t\right)\left(I_N-J_N\right)\otimes I_n\right) \\
\geq & \frac{np\delta\left(\min_{i\in\mathcal{V}}\beta_{i,1}\right)\left(\min_{k-np+1\leq t\leq k}R_t\right)}{2Nn\bar{H}^2\left(\min_{i\in\mathcal{V}}\beta_{i,1}\right)+N\left(\min_{k-np+1\leq t\leq k}R_t\right)}I_{Nn}, \\
= & npaI_{Nn} + O\left(\frac{1}{\psi'}\right), \text{a.s.,}
\end{aligned} \tag{4.18}$$

for some $\psi' > 0$, where $R_k = \left(\min_{(i,j)\in\mathcal{E}}\alpha_{ij,1}\frac{e^{-\|\theta\|_1/b_{ij}}}{b_{ij}}k^{1-\nu_{ij}-\gamma_{ij}}\left(1+O\left(\frac{1}{k^\psi}\right)\right)\right)\lambda_2(\mathcal{L})$ and J_N is defined in (4.6).

Then, by (4.16) and Lemma A.4, we have

$$\tilde{\Theta}_k = \begin{cases} O\left(\frac{1}{k^a}\right), & \text{if } 2h-2a>1; \\ O\left(\frac{\ln k}{k^{h-1/2}}\right), & \text{if } 2h-2a=1; \\ O\left(\frac{\sqrt{\ln k}}{k^{h-1/2}}\right), & \text{if } 2h-2a<1, \end{cases} \text{ a.s.} \qquad \square$$

Remark 4.9. Given ν_{ij} and γ_{ij}, an almost sure convergence rate of $O\left(\frac{\sqrt{\ln k}}{k^{h-1/2}}\right)$ can be achieved by properly selecting $\alpha_{ij,1}$, $\beta_{i,1}$ and b_{ij}. Especially, when $\nu_{ij}=0$, $\gamma_{ij}=1$, and a is sufficiently large, Algorithm 3.2 can achieve an almost sure convergence rate of $O(\sqrt{\ln k/k})$, which is the best one among existing literature [10, 12, 37] even without data rate constraints. For comparison, He et al. [10] and Kar et al. [12] show that their distributed estimation algorithm achieve a almost sure convergence rate of $o\left(k^{-\tau}\right)$ for some $\tau\in[0,\frac{1}{2})$. Zhang and Zhang [37] prove that $\frac{1}{k}\sum_{t=1}^{k}\|\tilde{\Theta}_t\|=o\left((b(k)k)^{-1/2}\right)$ almost surely for their algorithm, where $b(k)$ is the step-size satisfying the stochastic approximation condition $\sum_{k=0}^{\infty}b(k)=\infty,\sum_{k=0}^{\infty}b^2(k)<\infty$. The theoretical result of Theorem 4.8 is better than these ones. Our technique can be applied in the almost sure convergence rate analysis of other distributed estimation algorithms. For example, if the step-size $b(t)$ in the distributed estimation algorithm (3) of [37] is selected as $\frac{\beta}{k}$ with sufficiently large β, then by Lemma A.4, an almost sure convergence rate of $O(\sqrt{\ln k/k})$ can also be achieved.

Remark 4.10. When $\nu_{ij}<1$ for some $(i,j)\in\mathcal{E}$, we have $h=\min_{(i,j)\in\mathcal{E}}\left(\frac{\nu_{ij}}{2}+\gamma_{ij}\right)$ <1. Therefore, the almost sure convergence rate of $O(\sqrt{\ln k/k})$ cannot be obtained. This is because the communication frequency is reduced. Similar results can be seen in [10]. The trade-off between the convergence rate and the communication cost is discussed in Section 6.

5. Communication cost.

This section analyzes the communication cost of Algorithm 3.2 by calculating the average data rates defined in Definition 2.6.

Firstly, the local average data rates of Algorithm 3.2 are calculated.

THEOREM 5.1. *Under the condition of Theorem* 4.1*, the local average data rate* $\mathtt{B}_{ij}(k)=O\left(\frac{1}{k^{\nu_{ij}}}\right)$ *almost surely. Furthermore, if* $\nu_{ij}=0$*, then* $\mathtt{B}_{ij}(k)=1$*. And, if* $\nu_{ij}>0$ *and the step-sizes are set as Theorem* 4.8 *and* $a>h-1/2$*, then*

$$\mathtt{B}_{ij}(k)\leq\frac{\exp\left(\|\theta\|_1/b_{ij}\right)}{(1-\nu_{ij})k^{\nu_{ij}}}+O\left(\frac{\sqrt{\ln k}}{k^{h-1/2+\nu_{ij}}}\right),\ a.s.$$

Proof. If $\nu_{ij}=0$, then $C_{ij,k}=0$. In this case, the sensor i transmits 1 bit of message to the sensor j at every moment almost surely, which implies $\mathtt{B}_{ij}(k)=1$ almost surely. Therefore, it suffices to discuss the case of $\nu_{ij}>0$.

By the definition of $\zeta_{ij}(k)$, we have $\zeta_{ij}(k)$ is $\mathcal{F}_k$-measurable, and

$$\begin{aligned}\mathbb{P}\{\zeta_{ij}(k)=1\}=&F\left(\frac{\mathtt{x}_{i,k}-C_{ij,k}}{b_{ij}}\right)+F\left(\frac{-\mathtt{x}_{i,k}-C_{ij,k}}{b_{ij}}\right),\\ \mathbb{P}\{\zeta_{ij}(k)=0\}=&1-F\left(\frac{\mathtt{x}_{i,k}-C_{ij,k}}{b_{ij}}\right)-F\left(\frac{-\mathtt{x}_{i,k}-C_{ij,k}}{b_{ij}}\right).\end{aligned}$$

Firstly, we estimate $\sum_{t=1}^{k}\mathbb{E}\left[\zeta_{ij}(t)|\mathcal{F}_{t-1}\right]$. By Theorem 4.1, $\mathtt{x}_{i,k}=\varphi_k^{\top}\hat{\theta}_{i,k}$ is uniformly bounded almost surely. Therefore, when k is sufficiently large,

$$\begin{aligned}\mathbb{E}\left[\zeta_{ij}(k)|\mathcal{F}_{k-1}\right]=&F\left(\frac{\mathtt{x}_{i,k}-C_{ij,k}}{b_{ij}}\right)+F\left(\frac{-\mathtt{x}_{i,k}-C_{ij,k}}{b_{ij}}\right)\\ =&\frac{\exp\left((\mathtt{x}_{i,k}-C_{ij,k})/b_{ij}\right)+\exp\left((-\mathtt{x}_{i,k}-C_{ij,k})/b_{ij}\right)}{2}\\ =&\frac{\exp\left(\mathtt{x}_{i,k}/b_{ij}\right)+\exp\left(-\mathtt{x}_{i,k}/b_{ij}\right)}{2k^{\nu_{ij}}}=O\left(\frac{1}{k^{\nu_{ij}}}\right),\ \text{a.s.}\end{aligned}\tag{5.1}$$

Hence, $\mathbb{E}\left[\zeta_{ij}(k)|\mathcal{F}_{k-1}\right] = O\left(\frac{1}{k^{\nu_{ij}}}\right)$ for $\nu_{ij} \geq 0$ almost surely, which implies

$$\sum_{t=1}^{k} \mathbb{E}\left[\zeta_{ij}(t)|\mathcal{F}_{t-1}\right] = O\left(k^{1-\nu_{ij}}\right), \text{ a.s.} \tag{5.2}$$

Secondly, we estimate $\sum_{t=1}^{k} \zeta_{ij}(t) - \mathbb{E}\left[\zeta_{ij}(t)|\mathcal{F}_{t-1}\right]$. Since $\nu_{ij} \leq \frac{1}{2}$ under the condition of Theorem 4.1, $1 - \nu_{ij} > \frac{1}{2} - \frac{\nu_{ij}}{4}$. By $\mathbb{E}\left[\zeta_{ij}(k)|\mathcal{F}_{k-1}\right] = O\left(\frac{1}{k^{\nu_{ij}}}\right)$ almost surely and $\zeta_{ij}(k) = 0$ or 1, we have

$$\begin{aligned}
&\mathbb{E}\left[\left\|\zeta_{ij}(k) - \mathbb{E}\left[\zeta_{ij}(k)|\mathcal{F}_{k-1}\right]\right\|^4 \middle| \mathcal{F}_{k-1}\right] \\
\leq &\mathbb{E}\left[\left(\zeta_{ij}(k) - \mathbb{E}\left[\zeta_{ij}(k)|\mathcal{F}_{k-1}\right]\right)^2 \middle| \mathcal{F}_{k-1}\right] \\
= &\mathbb{E}\left[\zeta_{ij}|\mathcal{F}_{k-1}\right] - \left(\mathbb{E}\left[\zeta_{ij}(k)|\mathcal{F}_{k-1}\right]\right)^2 = O\left(\frac{1}{k^{\nu_{ij}}}\right), \text{ a.s.}
\end{aligned}$$

Then, by Theorem 1.3.10 of [8], it holds that

$$\begin{aligned}
&\sum_{t=1}^{k} \left(\zeta_{ij}(t) - \mathbb{E}\left[\zeta_{ij}(t)|\mathcal{F}_{t-1}\right]\right) \\
= &\sum_{t=1}^{k} \frac{1}{t^{\nu_{ij}/4}} \cdot t^{\nu_{ij}/4} \left(\zeta_{ij}(t) - \mathbb{E}\left[\zeta_{ij}(t)|\mathcal{F}_{t-1}\right]\right) = O\left(k^{\frac{1}{2} - \frac{\nu_{ij}}{4}} \sqrt{\ln \ln k}\right), \text{ a.s.}
\end{aligned} \tag{5.3}$$

(5.2) and (5.3) imply $\sum_{t=1}^{k} \zeta_{ij}(t) = O(k^{1-\nu_{ij}})$ almost surely. Therefore, $\mathtt{B}_{ij}(k) = O\left(\frac{1}{k^{\nu_{ij}}}\right)$ almost surely.

If the step-sizes are set as Theorem 4.8 and $a > h - 1/2$, then by Theorem 4.8, $\tilde{\theta}_{i,k} = O\left(\frac{\sqrt{\ln k}}{k^{h-1/2}}\right)$ almost surely for all $i \in \mathcal{V}$. Then, by (5.1), we have

$$\mathbb{E}\left[\zeta_{ij}(k)|\mathcal{F}_{k-1}\right] \leq \frac{\exp\left(\|\theta\|_1 / b_{ij}\right)}{k^{\nu_{ij}}} + O\left(\frac{\sqrt{\ln k}}{k^{h-1/2+\nu_{ij}}}\right), \text{ a.s.}$$

Therefore, one can get

$$\mathtt{B}_{ij}(k) \leq \frac{\exp\left(\|\theta\|_1 / b_{ij}\right)}{(1-\nu_{ij})k^{\nu_{ij}}} + O\left(\frac{\sqrt{\ln k}}{k^{h-1/2+\nu_{ij}}}\right), \text{ a.s.} \qquad \square$$

Remark 5.2. By Theorem 5.1, the decaying rate of $\mathtt{B}_{ij}(k)$ only depends on ν_{ij}. Therefore, the operators of sensors i and j can directly set and easily know the decaying rate of $\mathtt{B}_{ij}(k)$ before running the algorithm.

Remark 5.3. The noise coefficient b_{ij} influences the almost sure convergence rate and the average data rate. By Theorem 4.8, an almost sure convergence rate of $O\left(\frac{\sqrt{\ln k}}{k^{h-1/2}}\right)$ can be achieved when $2h - 2a < 1$, where a is a function of b_{ij}. By Theorem 5.1, the upper bound of $\mathtt{B}_{ij}(k)$ is monotonically non-increasing with b_{ij}. Therefore, increasing b_{ij} while maintaining $2h - 2a < 1$ can reduce the communication cost without losing the almost sure convergence rate.

Then, we can estimate the global average data rate.

THEOREM 5.4. *Under the condition of Theorem* 4.1*, the global average data rate* $\mathtt{B}(k) = O\left(\frac{1}{k^{\underline{\nu}}}\right)$ *almost surely, where* $\underline{\nu} = \min_{(i,j)\in\mathcal{E}} \nu_{ij}$.

Proof. The theorem can be proved by Theorem 5.1 and $\mathtt{B}(k) = \frac{\sum_{(i,j)\in\mathcal{E}} \mathtt{B}_{ij}(k)}{2M}$. □

Remark 5.5. If the step-sizes are set as Theorem 4.8 and $a > h - 1/2$, the upper bound of global average data rate $\mathtt{B}(k)$ can also be obtained by Theorem 5.1 and $\mathtt{B}(k) = \frac{\sum_{(i,j)\in\mathcal{E}} \mathtt{B}_{ij}(k)}{2M}$.

6. Trade-off between convergence rate and communication cost. In Sections 4 and 5, we quantitatively demonstrate the effectiveness of Algorithm 3.2 by the almost sure convergence rate and the communication cost by the average data rates. This section establishes the trade-off between the convergence rate and the communication cost.

By Theorem 4.8, the convergence rate of Algorithm 3.2 is influenced by the selection of step-sizes $\alpha_{ij,k}$ and $\beta_{i,k}$. The following theorem optimizes almost sure convergence rate by properly selecting the step-sizes.

THEOREM 6.1. *In Algorithm* 3.2*, set* $\nu_{ij} \in [0, \frac{1}{2})$*. Then, under the condition of Theorem* 4.1*, there exist step-sizes* $\alpha_{ij,k}$ *and* $\beta_{i,k}$ *such that* $\tilde{\theta}_{i,k} = O\left(\frac{\sqrt{\ln k}}{k^{1/2-\bar{\nu}/2}}\right)$ *almost surely, where* $\bar{\nu} = \max_{(i,j)\in\mathcal{E}} \nu_{ij}$*.*

Proof. Set $\gamma_{ij} = 1 - \nu_{ij}$. Then, h in Theorem 4.8 equals to $1 - \bar{\nu}/2$. Besides, when $\alpha_{ij,1}$ and $\beta_{i,1}$ are sufficiently large, a in Theorem 4.8 is larger than $2h - 1$. Then, the theorem can be proved by Theorem 4.8. □

Remark 6.2. The proof of Theorem 6.1 provides a selection method to optimize the convergence rate of the algorithm.

Theorem 6.1 shows that when properly selecting the step-sizes, the key factor to determine the almost sure convergence rate of Algorithm 3.2 is the event-triggered coefficient ν_{ij}. The optimal almost sure convergence rate of Algorithm 3.2 gets faster under smaller ν_{ij}.

On the other hand, Theorem 5.1 shows that ν_{ij} is the decaying rate of the local average data rate for the communication channel $(i, j) \in \mathcal{E}$. Theorem 5.4 shows that $\underline{\nu} = \min_{(i,j)\in\mathcal{E}} \nu_{ij}$ is the decaying rate of the global average data rate. Therefore, the average data rates of Algorithm 3.2 get smaller under large ν_{ij}.

Therefore, there is a trade-off between the convergence rate and the communication cost. The operator of each sensor i can decrease ν_{ij} of the adjacent communication channel $(i, j) \in \mathcal{E}$ for a better convergence rate, or increase ν_{ij} for a lower communication cost.

7. Simulation. This section gives a numerical example to illustrate the effectiveness and the average data rates of Algorithm 3.2.

Consider a network with 8 sensors. The communication topology is shown in Figure 1. $a_{ij} = 1$ if $(i, j) \in \mathcal{E}$, and 0, otherwise. For the sensor i, the measurement matrix $H_{i,k} = \begin{bmatrix} 1 & 0 \end{bmatrix}$ if i is odd, and $\begin{bmatrix} 0 & 1 \end{bmatrix}$ if i is even. The observation noise $w_{i,k}$ is i.i.d. Gaussian with zero mean and standard deviation 0.1. The true value $\theta = \begin{bmatrix} 1 & -1 \end{bmatrix}^\top$.

In Algorithm 3.2, set $b_{ij} = \frac{1}{2}$ and $\nu_{ij} = \frac{1}{4}$. The step-sizes $\alpha_{ij,k} = \frac{5}{k^{3/4}}$ and $\beta_{i,k} = \frac{5}{k}$. Figure 2 shows the trajectory of $\frac{1}{N}\sum_{i=1}^{N} \|\tilde{\theta}_{i,k}\|^2$, which demonstrates the convergence of Algorithm 3.2.

To show the balance between the convergence rate and the communication cost, set $b_{ij} = \frac{1}{2}$, $\nu_{ij} = \nu = 0, \frac{1}{9}, \frac{2}{9}, \frac{3}{9}, \frac{4}{9}$, and the step-sizes $\alpha_{ij,k} = \frac{5}{k^{1-\nu}}$ and $\beta_{i,k} = \frac{5}{k}$. The simulation is repeated 50 times. Denote $\tilde{\theta}^t_{i,k}$ as the estimation error of the sensor i

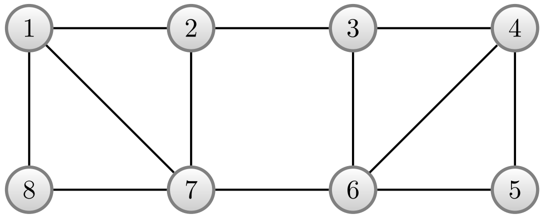

FIG. 1. *Communication topology.*

FIG. 2. *The trajectory of* $\frac{1}{N}\sum_{i=1}^{N}\|\tilde{\theta}_{i,k}\|^2$

at time k in the t-th run. Figure 3 depicts the log-log plot of $\frac{1}{N}\sum_{i=1}^{N}\|\tilde{\theta}_{i,k}^{t}\|^2$, which demonstrates that the convergence rate is faster under a smaller ν. Figure 4 shows the log-log plot of $\mathtt{B}(k)$, which illustrates that the global average data rate is smaller under a larger ν. Figures 3 and 4 reveal the trade-off between the convergence rate and the data rate.

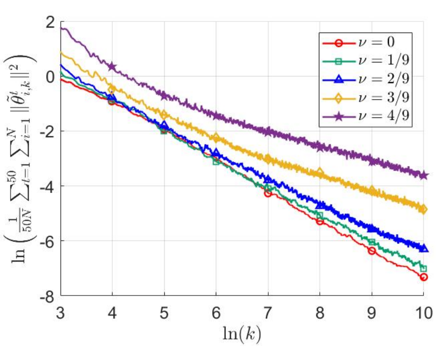

FIG. 3. *Convergence rates with different* ν

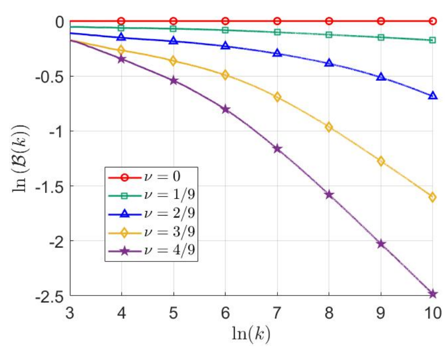

FIG. 4. *Average data rates with different* ν

Figures 5 and 6 compare Algorithm 3.2 with the single bit diffusion algorithm [25] and the distributed least mean square (LMS) algorithm [35], which demonstrates that Algorithm 3.2 can achieve higher estimation accuracy at a lower communication data rate compared to the algorithms in [25, 35].

8. Conclusion. This paper considers the distributed estimation under low communication cost, which is described by the average data rates. We propose a novel distributed estimation algorithm, where the SC consensus protocol [14] is used to fuse neighborhood information, and a new stochastic event-triggered mechanism is

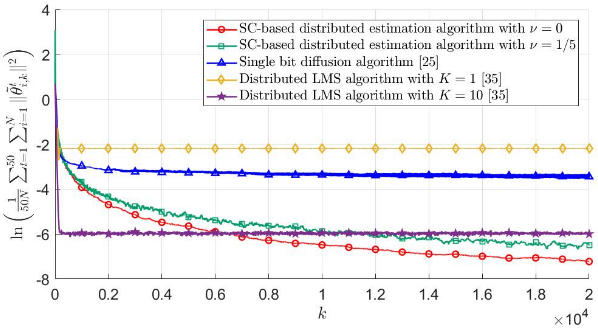

FIG. 5. *The trajectories of* $\ln\left(\frac{1}{50N}\sum_{t=1}^{50}\sum_{i=1}^{N}\|\tilde{\theta}_{i,k}^{t}\|^2\right)$ *or different algorithms*

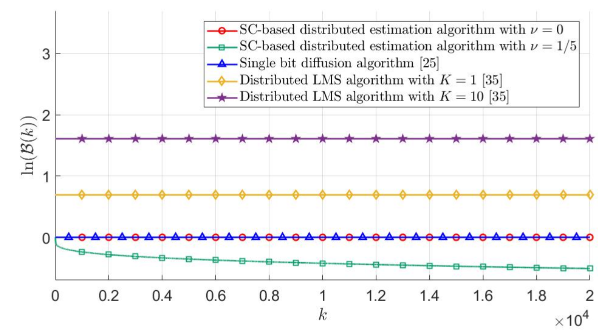

FIG. 6. *Average data rates for different algorithms*

designed to reduce the communication frequency. The algorithm has advantages both in the effectiveness and communication cost. For the effectiveness, the estimates of the algorithm are proved to converge to the true value in the almost sure and mean square sense, and polynomial almost sure convergence rate is also obtained. For the communication cost, the local and global average data rates are proved to decay to zero at polynomial rates. Besides, the trade-off between convergence rate and communication cost is established through event-triggered coefficients. A better convergence rate can be achieved by decreasing event-triggered coefficients, while lower communication cost can be achieved by increasing event-triggered coefficients.

There are interesting issues for future works. For example, how to extend the results to the cases with more complex communication graphs, such as directed graphs and switching graphs? Besides, Gan and Liu [7] consider the distributed order estimation, and Xie and Guo [36] investigate distributed adaptive filtering. These issues also suffer the communication cost problems. Then, how to apply our technique to these works to save the communication cost?

Appendix A. Lemmas.

LEMMA A.1. *Let* $f(\cdot)$ *be the density function of* $Lap(0,1)$. *Given* $C_k = \nu b \ln k$ *with* $\nu \geq 0$ *and* $b > 0$, *and a compact set* $\mathcal{X}$, *we have* $\inf_{x\in\mathcal{X},k\in\mathbb{N}} \frac{k^{\nu}}{b} f((x - C_k)/b) > 0$.

Proof. If $\nu = 0$, then $C_k = 0$ for all k. Therefore, $\inf_{x\in\mathcal{X},k\in\mathbb{N}} \frac{1}{b} f(x/b) > 0$ by the compactness of $\mathcal{X}$.

If $\nu > 0$, then $\lim_{k\to\infty} C_k = \infty$, which together with the compactness of $\mathcal{X}$ implies that there exists k_0 such that $x - C_k < 0$ for all $x \in \mathcal{X}$ and $k \geq k_0$. Hence,

$$\inf_{x\in\mathcal{X},k\geq k_0} \frac{k^\nu}{b} f\left(\frac{x - C_k}{b}\right) = \inf_{x\in\mathcal{X},k\geq k_0} \frac{k^\nu}{2b} e^{(x-\nu b\ln k)/b} = \frac{1}{2b} e^{\min \mathcal{X}/b} > 0.$$

Besides by the compactness of $\mathcal{X}$, one can get $\inf_{x\in\mathcal{X}} \frac{k^\nu}{b} f((x - C_k)/b) > 0$ for all $k < k_0$. The lemma is proved. □

Lemma A.2. *If positive sequence $\{z_k\}$ satisfies $\sum_{k=1}^\infty z_k = \infty$ and $z_{k+1} = O(z_k)$, then for any $l \in \{1, \ldots, n\}$, $\sum_{q=1}^\infty \min_{n(q-1)+l<t\leq nq+l} z_t = \infty$.*

Proof. Set $\bar{z} = \sup\left\{1, \frac{z_k+1}{z_k}, k \in \mathbb{N}\right\} < \infty$. Then, $z_k \geq \frac{z_{k+1}}{\bar{z}}$. Therefore,

$$\sum_{q=1}^\infty \min_{n(q-1)+l<t\leq nq+l} z_t \geq \sum_{q=1}^\infty \max_{nq+l<t\leq n(q+1)+l} \frac{z_t}{\bar{z}^{2n}} \geq \frac{1}{n\bar{z}^{2n}} \sum_{k=l+n+1}^\infty z_k = \infty. \qquad \square$$

Lemma A.3. *a) $\ln(1 + x + y) \leq \ln(1 + x) + \ln(1 + y)$ for all $x, y \geq 0$;*
b) $\ln(1 + x) - \ln(1 + y) \leq \frac{x-y}{1+y}$ for all $x, y \geq 0$;
c) $\frac{\ln(1+x)-\ln(1+y)}{x-y} \leq 1$ for all $x, y \geq 0$;
d) $\sup_{x>0} \frac{\ln(1+x)}{x^p} < \infty$ for all $p \in (0, 1]$.

Proof. a), b) and c) can be proved by $\ln(1 + x + y) \leq \ln((1 + x)(1 + y)) = \ln(1 + x) + \ln(1 + y)$, Proposition 5.4.6 of [41] and the Lagrange mean value theorem [41], respectively. For d), if $p = 1$, then we have $\sup_{x\geq 0} \frac{\ln(1+x)}{x} \leq 1$. If $p \in (0, 1)$, then $x_1 > x_2 > 0$ such that $\ln(1 + x) < x^p$ for all $x \in (0, x_2) \cup (x_1, \infty)$. Therefore, $\sup_{x\geq 0} \frac{\ln(1+x)}{x^p} \leq \max\left\{\sup_{x\in[x_2,x_1]} \frac{\ln(1+x)}{x^p}, 1\right\} < \infty$. □

Lemma A.4. *Assume that*

i) $\{\mathcal{F}_k\}$ is a σ-algebra sequence satisfying $\mathcal{F}_{k-1} \subseteq \mathcal{F}_k$ for all k;
ii) $\{\mathtt{U}_k\}$ is a matrix sequence satisfying that $\mathtt{U}_k$ is $\mathcal{F}_{k-1}$-measurable, $\mathtt{U}_k = O(k^\mu)$ for some $0 \leq \mu < \frac{1}{2}$ almost surely, $\mathtt{U}_k + \mathtt{U}_k^\top$ is positive semi-definite for all k, and

$$\frac{1}{2p} \sum_{t=k-p+1}^{k} \mathtt{U}_t + \mathtt{U}_t^\top \geq a I_n \tag{A.1}$$

for some $p \in \mathbb{N}$, $a > 0$ and all $k \in \mathbb{N}$ almost surely;
iii) $\{\mathtt{W}_k, \mathcal{F}_k\}$ is a martingale difference sequence such that $\mathbb{E}\left[\|\mathtt{W}_k\|^\rho|\mathcal{F}_{k-1}\right] = O\left(\frac{1}{k^{\rho h}}\right)$ almost surely for some $\rho > 2$ and $\frac{1}{2} < h \leq \min\{1, \frac{3+2h-2\mu}{4}\}$;
iv) $\{\mathtt{X}_k, \mathcal{F}_k\}$ is a sequence of adaptive random variables;
v) There exists $c > 1$ almost surely such that

$$\mathtt{X}_k = \left(I_n - \frac{\mathtt{U}_k}{k} + O\left(\frac{1}{k^c}\right)\right) \mathtt{X}_{k-1} + \mathtt{W}_k. \tag{A.2}$$

Then,

$$\mathtt{X}_k = \begin{cases} O\left(\frac{1}{k^a}\right), & \text{if } 2h - 2a > 1; \\ O\left(\frac{\ln k}{k^{h-1/2}}\right), & \text{if } 2h - 2a = 1; \\ O\left(\frac{\sqrt{\ln k}}{k^{h-1/2}}\right), & \text{if } 2h - 2a < 1, \end{cases} \quad a.s.$$

Proof. Denote $\bar{\mathtt{U}}_t = \frac{\mathtt{U}_t+\mathtt{U}_t^\top}{2}$. Then, by (A.2),

$$\begin{aligned}
&\mathbb{E}\left[\left\|\mathtt{X}_k\right\|^2 \middle| \mathcal{F}_{k-1}\right] \\
&=\left(1+O\left(\frac{1}{k^{\min\{c,2-2\mu\}}}\right)\right)\left\|\mathtt{X}_{k-1}\right\|^2-\frac{2}{k}\mathtt{X}_{k-1}^\top\bar{\mathtt{U}}_k\mathtt{X}_{k-1}+O\left(\frac{1}{k^{2b}}\right), \text{ a.s.}
\end{aligned}$$

Hence, by Theorem 1.3.2 of [8], we have $\left\|\mathtt{X}_k\right\|^2$ converges to a finite value almost surely, which implies the almost sure boundedness of $\mathtt{X}_k$.

We estimate the almost sure convergence rate of $\mathtt{X}_k$ in the following two cases.

Case 1: $2h-2a>1$. In this case, we have

$$\begin{aligned}
\text{(A.3)}\quad &\mathbb{E}\left[(k+1)^{2a}\left\|\mathtt{X}_k\right\|^2 \middle| \mathcal{F}_{k-1}\right] \\
\leq&\left(1+\frac{2a}{k}+O\left(\frac{1}{k^{\min\{c,2-2\mu\}}}\right)\right)k^{2a}\left\|\mathtt{X}_{k-1}\right\|^2-\frac{2}{k^{1-2a}}\mathtt{X}_{k-1}^\top\bar{\mathtt{U}}_k\mathtt{X}_{k-1}+O\left(\frac{1}{k^{2h-2a}}\right) \\
=&\left(1+O\left(\frac{1}{k^{\min\{c,2-2\mu\}}}\right)\right)k^{2a}\left\|\mathtt{X}_{k-1}\right\|^2+\frac{2a}{k^{1-2a}}\left\|\mathtt{X}_{k-1}\right\|^2-\frac{2}{k^{1-2a}}\mathtt{X}_{k-1}^\top\bar{\mathtt{U}}_k\mathtt{X}_{k-1} \\
&+O\left(\frac{1}{k^{2h-2a}}\right), \text{ a.s.}
\end{aligned}$$

Next, we will prove that $\sup_{k\in\mathbb{N}}\sum_{t=1}^k\left(\frac{2a}{t^{1-2a}}\left\|\mathtt{X}_{t-1}\right\|^2-\frac{2}{t^{1-2a}}\mathtt{X}_{t-1}^\top\bar{\mathtt{U}}_t\mathtt{X}_{t-1}\right)<\infty$ almost surely. Note that $1-2a>2-2h\geq 0$. Then, by (A.1), one can get

$$\begin{aligned}
\text{(A.4)}\quad &\sum_{t=1}^k\left(\frac{2a}{t^{1-2a}}\left\|\mathtt{X}_{t-1}\right\|^2-\frac{2}{t^{1-2a}}\mathtt{X}_{t-1}^\top\bar{\mathtt{U}}_t\mathtt{X}_{t-1}\right) \\
\leq&\sum_{r=0}^{\lfloor\frac{k}{p}\rfloor-1}\sum_{t=pr+1}^{pr+p}\left(\frac{2a}{t^{1-2a}}\left\|\mathtt{X}_{t-1}\right\|^2-\frac{2}{t^{1-2a}}\mathtt{X}_{t-1}^\top\bar{\mathtt{U}}_t\mathtt{X}_{t-1}\right)+O\left(\frac{1}{k^{1-2a}}\right) \\
\leq&\sum_{r=0}^{\lfloor\frac{k}{p}\rfloor-1}\frac{2}{(pr+p)^{1-2a}}\sum_{t=pr+1}^{pr+p}\left(a\left\|\mathtt{X}_{t-1}\right\|^2-\mathtt{X}_{t-1}^\top\bar{\mathtt{U}}_t\mathtt{X}_{t-1}\right)+\sum_{r=0}^{\lfloor\frac{k}{p}\rfloor-1}O\left(\frac{1}{r^{2-2a}}\right)+O(1) \\
\leq&\sum_{r=0}^{\lfloor\frac{k}{p}\rfloor-1}\frac{2}{(pr+p)^{1-2a}}\sum_{t=pr+1}^{pr+p}\left(\mathtt{X}_{pr+p-1}^\top\bar{\mathtt{U}}_t\mathtt{X}_{pr+p-1}-\mathtt{X}_{t-1}^\top\bar{\mathtt{U}}_t\mathtt{X}_{t-1}\right) \\
&+\sum_{r=0}^{\lfloor\frac{k}{p}\rfloor-1}\frac{2a}{(pr+p)^{1-2a}}\sum_{t=pr+1}^{pr+p}\left(\left\|\mathtt{X}_{t-1}\right\|^2-\left\|\mathtt{X}_{pr+p-1}\right\|^2\right)+O(1).
\end{aligned}$$

Besides,

$$\begin{aligned}
\text{(A.5)}\quad &\sum_{t=pr+1}^{pr+p}\left(\mathtt{X}_{pr+p-1}^\top\bar{\mathtt{U}}_t\mathtt{X}_{pr+p-1}-\mathtt{X}_{t-1}^\top\bar{\mathtt{U}}_t\mathtt{X}_{t-1}\right) \\
=&\sum_{t=pr+1}^{pr+p}\sum_{l=t}^{pr+p-1}\left(\mathtt{X}_l^\top\bar{\mathtt{U}}_t\mathtt{X}_l-\mathtt{X}_{l-1}^\top\bar{\mathtt{U}}_t\mathtt{X}_{l-1}\right) \\
=&\sum_{t=pr+1}^{pr+p}\sum_{l=t}^{pr+p-1}\left(2\mathtt{W}_l^\top\bar{\mathtt{U}}_t\left(I_n-\frac{\bar{\mathtt{U}}_l}{l}+O\left(\frac{1}{l^c}\right)\right)\mathtt{X}_{l-1}+\mathtt{W}_l^\top\bar{\mathtt{U}}_t\mathtt{W}_l\right)+O\left(r^{2\mu-1}\right), \text{ a.s.}
\end{aligned}$$

When $t \in \{pq+1, \dots, pq+l\}$ and $l = \{t, \dots, pr+p-1\}$, it holds that

$$\frac{4}{(pr+p)^{1-2a} l^b} \bar{\mathtt{U}}_t \left(I_n - \frac{\bar{\mathtt{U}}_l}{l} + O\left(\frac{1}{l^c}\right)\right) \mathtt{X}_{l-1} = O\left(\frac{1}{r^{1+b-2a-\mu}}\right).$$

Note that $1+h-2a-\mu \geq 2h-2a-\mu > \frac{1}{2}$. Then, by Theorem 1.3.10 of [8], we have

(A.6)
$$\sum_{r=0}^{\lfloor \frac{k}{p} \rfloor -1} \sum_{t=pr+1}^{pr+p} \sum_{l=t}^{pr+p-1} (l^b \mathtt{W}_l)^\top \left(\frac{4}{(pr+p)^{1-2a} l^b} \bar{\mathtt{U}}_t \left(I_n - \frac{\bar{\mathtt{U}}_l}{l} + O\left(\frac{1}{l^c}\right)\right) \mathtt{X}_{l-1} \right) = O(1), \text{ a.s.}$$

Additionally, by $1+2b-2a-\mu > 2-\mu > 1$ and Theorem 1.3.9 of [8] with $\alpha = 1$,

$$\begin{aligned}
&\sum_{r=0}^{\lfloor \frac{k}{p} \rfloor -1} \sum_{t=pr+1}^{pr+p} \frac{2}{(pr+p)^{1-2a}} \sum_{l=t}^{pr+p-1} \mathtt{W}_l^\top \bar{\mathtt{U}}_t \mathtt{W}_l \\
=& \sum_{r=0}^{\lfloor \frac{k}{p} \rfloor -1} \sum_{t=pr+1}^{pr+p} \sum_{l=t}^{pr+p-1} \frac{2}{(pr+p)^{1-2a} t^{2b-\mu}} \cdot t^{2b-\mu} \left(\mathtt{W}_l^\top \bar{\mathtt{U}}_t \mathtt{W}_l - \mathbb{E}\left[\mathtt{W}_l^\top \bar{\mathtt{U}}_t \mathtt{W}_l \middle| \mathcal{F}_{l-1} \right] \right) \\
&+ \sum_{r=0}^{\lfloor \frac{k}{p} \rfloor -1} \sum_{t=pr+1}^{pr+p} \frac{2}{(pr+p)^{1-2a}} \sum_{l=t}^{pr+p-1} \mathbb{E}\left[\mathtt{W}_l^\top \bar{\mathtt{U}}_t \mathtt{W}_l \middle| \mathcal{F}_{l-1} \right] = O(1), \text{ a.s.,}
\end{aligned}$$

which together with (A.5) and (A.6) implies that

$$\sum_{r=0}^{\lfloor \frac{k}{p} \rfloor -1} \frac{2}{(pr+p)^{1-2a}} \sum_{t=pr+1}^{pr+p} \left(\mathtt{X}_{pr+p-1}^\top \bar{\mathtt{U}}_t \mathtt{X}_{pr+p-1} - \mathtt{X}_{t-1}^\top \bar{\mathtt{U}}_t \mathtt{X}_{t-1} \right) = O(1), \text{ a.s.}$$

Similarly, one can get

$$\sum_{r=0}^{\lfloor \frac{k}{p} \rfloor -1} \frac{2a}{(pr+p)^{1-2a}} \sum_{t=pr+1}^{pr+p} \left(\|\mathtt{X}_{t-1}\|^2 - \|\mathtt{X}_{pr+p-1}\|^2 \right) = O(1), \text{ a.s.}$$

Then, by (A.4), we have

(A.7)
$$\sum_{t=1}^{k} \left(\frac{2a}{t^{1-2a}} \|\mathtt{X}_{t-1}\|^2 - \frac{2}{t^{1-2a}} \mathtt{X}_{t-1}^\top \bar{\mathtt{U}}_t \mathtt{X}_{t-1} \right) < \infty, \text{ a.s.,}$$

Given $S_0 > 0$, define $\mathtt{S}_k = S_0 - \sum_{t=1}^{k} \left(\frac{2a}{t^{1-2a}} \|\mathtt{X}_{t-1}\|^2 - \frac{2}{t^{1-2a}} \mathtt{X}_{t-1}^\top \bar{\mathtt{U}}_t \mathtt{X}_{t-1} \right)$ and $\mathtt{V}_k = (k+1)^{2a} \|\mathtt{X}_k\|^2 + \mathtt{S}_k$. Hence by (A.3), we have

$$\mathbb{E}\left[\mathtt{V}_k | \mathcal{F}_{k-1}\right] \leq \left(1 + O\left(\frac{1}{k^{\min\{c, 2-2\mu\}}}\right)\right) \mathtt{V}_{k-1} + O\left(\frac{1}{k^{2h-2a}}\right), \text{ a.s.}$$

Then, define $\mathtt{k}_0 = \inf\{k : \mathtt{S}_k < 0\}$. We have

$$\begin{aligned}
&\mathbb{E}\left[\mathtt{V}_{\min\{k, \mathtt{k}_0\}} \middle| \mathcal{F}_{k-1} \right] \\
\leq& \mathtt{V}_{\mathtt{k}_0} I_{\{\mathtt{k}_0 \leq k\}} + \left(1 + O\left(\frac{1}{k^{\min\{c, 2-2\mu\}}}\right)\right) \mathtt{V}_{k-1} I_{\{\mathtt{k}_0 > k\}} + O\left(\frac{1}{k^{2h-2a}}\right) \\
\leq& \left(1 + O\left(\frac{1}{k^{\min\{c, 2-2\mu\}}}\right)\right) \mathtt{V}_{\min\{k-1, \mathtt{k}_0\}} + O\left(\frac{1}{k^{2h-2a}}\right).
\end{aligned}$$

By Theorem 1.3.2 of [8], $\mathtt{V}_{\min\{k,\mathtt{k}_0\}}$ converges to a finite value almost surely. Note that $\mathtt{V}_k = \mathtt{V}_{\min\{k,\mathtt{k}_0\}}$ in the set

$$\{\mathtt{k}_0 = \infty\} = \{\inf_k \mathtt{S}_k \geq 0\} = \left\{ \sum_{t=1}^{k} \left(\frac{2a}{t^{1-2a}} \left\| \mathtt{X}_{t-1} \right\|^2 - \frac{2}{t^{1-2a}} \mathtt{X}_{t-1}^\top \bar{\mathtt{U}}_t \mathtt{X}_{t-1} \right) < S_0 \right\}.$$

Then, by the arbitrariness of S_0 and (A.7), $\mathtt{V}_k$ converges to a finite value almost surely, which implies the almost sure boundedness of $(k+1)^{2a} \left\| \mathtt{X}_k \right\|^2$. Hence, one can get $\mathtt{X}_k = O\left(\frac{1}{k^a}\right)$ almost surely.

Case 2: $2h - 2a \leq 1$**.** In this case, we have

$$\text{(A.8)} \quad \mathbb{E}\left[\frac{(k+1)^{2h-1}}{(\ln(k+1))^2} \left\| \mathtt{X}_k \right\|^2 \middle| \mathcal{F}_{k-1} \right]$$
$$\leq \left(1 + \frac{2h-1}{k} + O\left(\frac{1}{k^{\min\{c, 2-2\mu\}}} \right) \right) \frac{k^{2h-1}}{(\ln k)^2} \left\| \mathtt{X}_{k-1} \right\|^2$$
$$- \frac{2}{k^{2-2h} (\ln k)^2} \mathtt{X}_{k-1}^\top \bar{\mathtt{U}}_k \mathtt{X}_{k-1} + O\left(\frac{1}{k (\ln k)^2} \right)$$
$$\leq \left(1 + O\left(\frac{1}{k^{\min\{c, 2-2\mu\}}} \right) \right) \frac{k^{2h-1}}{(\ln k)^2} \left\| \mathtt{X}_{k-1} \right\|^2 + \frac{2a}{k^{2-2h} (\ln k)^2} \left\| \mathtt{X}_{k-1} \right\|^2$$
$$- \frac{2}{k^{2-2h} (\ln k)^2} \mathtt{X}_{k-1}^\top \bar{\mathtt{U}}_k \mathtt{X}_{k-1} + O\left(\frac{1}{k (\ln k)^2} \right), \text{ a.s.}$$

Then, similar to the case of $2h - 2a > 1$, we have $\mathtt{X}_k = O\left(\frac{\ln k}{k^{h-1/2}} \right)$ almost surely.

We further promote the almost sure convergence rate for the case of $2h - 2a < 1$. Since $\mathtt{X}_k = O\left(\frac{\ln k}{k^{h-1/2}} \right)$ almost surely, one can get

$$\text{(A.9)} \quad (k+1)^{2h-1} \left\| \mathtt{X}_k \right\|^2$$
$$\leq k^{2h-1} \left\| \mathtt{X}_{k-1} \right\|^2 + 2(k+1)^{2h-1} \mathtt{W}_k^\top \left(I_n - \frac{\bar{\mathtt{U}}_k}{k} + O\left(\frac{1}{k^c} \right) \right) \mathtt{X}_{k-1} + \frac{2h-1}{k^{2-2h}} \left\| \mathtt{X}_{k-1} \right\|^2$$
$$- \frac{2}{k^{2-2h}} \mathtt{X}_{k-1}^\top \bar{\mathtt{U}}_k \mathtt{X}_{k-1} + (k+1)^{2h-1} \left(\left\| \mathtt{W}_k \right\|^2 - \mathbb{E}\left[\left\| \mathtt{W}_k \right\|^2 \middle| \mathcal{F}_{k-1} \right] \right) + O\left(\frac{1}{k} \right).$$

By Theorem 1.3.10 of [8], it holds that

$$\sum_{t=1}^{k} 2(t+1)^{2h-1} \mathtt{W}_t^\top \left(I_n - \frac{\bar{\mathtt{U}}_t}{t} + O\left(\frac{1}{t^c} \right) \right) \mathtt{X}_{t-1}$$
$$= \sum_{t=1}^{k} ((t+1)^h \mathtt{W}_t)^\top \left(2(t+1)^{h-1} \left(I_n - \frac{\bar{\mathtt{U}}_t}{t} + O\left(\frac{1}{t^c} \right) \right) \mathtt{X}_{t-1} \right)$$
$$= O(1) + o\left(\sum_{t=1}^{k} \frac{1}{t^{2-2h}} \left\| \mathtt{X}_t \right\|^2 \right), \text{ a.s.}$$

By Theorem 1.3.9 of [8] with $\alpha = 1$, one can get

$$\sum_{t=1}^{k} (t+1)^{2h-1} \left(\left\| \mathtt{W}_t \right\|^2 - \mathbb{E}\left[\left\| \mathtt{W}_t \right\|^2 \middle| \mathcal{F}_{k-1} \right] \right)$$
$$= \sum_{t=1}^{k} (t+1)^{2h} \left(\left\| \mathtt{W}_t \right\|^2 - \mathbb{E}\left[\left\| \mathtt{W}_t \right\|^2 \middle| \mathcal{F}_{k-1} \right] \right) \cdot \frac{1}{t+1} = O(\ln k), \text{ a.s.}$$

Similar to (A.7), we have

$$\sum_{t=1}^{k}\left(\frac{2a}{t^{2-2h}}\left\|\mathtt{X}_{t-1}\right\|^2-\frac{2}{t^{2-2h}}\mathtt{X}_{t-1}^{\top}\bar{\mathtt{U}}_t\mathtt{X}_{t-1}\right)\le o(\ln k),\ \text{ a.s.}$$

Hence, by (A.9),

$$\begin{aligned}
&(k+1)^{2h-1}\left\|\mathtt{X}_k\right\|^2\\
\le&\left\|\mathtt{X}_0\right\|^2+\sum_{t=1}^{k}2(t+1)^{2h-1}\mathtt{W}_t^{\top}\left(I_n-\frac{\bar{\mathtt{U}}_t}{t}+O\left(\frac{1}{t^c}\right)\right)\mathtt{X}_{t-1}\\
&-\sum_{t=1}^{k}\frac{1+2a-2h}{t^{2-2h}}\left\|\mathtt{X}_{t-1}\right\|^2+\sum_{t=1}^{k}\left(\frac{2a}{t^{2-2h}}\left\|\mathtt{X}_{t-1}\right\|^2-\frac{2}{t^{2-2h}}\mathtt{X}_{t-1}^{\top}\bar{\mathtt{U}}_t\mathtt{X}_{t-1}\right)\\
&+\sum_{t=1}^{k}(t+1)^{2h-1}\left(\left\|\mathtt{W}_t\right\|^2-\mathbb{E}\left[\left\|\mathtt{W}_t\right\|^2\middle|\mathcal{F}_{k-1}\right]\right)+O(\ln k)\\
\le& o\left(\sum_{t=1}^{k}\frac{1}{t^{2-2h}}\left\|\mathtt{X}_t\right\|^2\right)-(1+2a-2h)\sum_{t=1}^{k}\frac{1}{t^{2-2h}}\left\|\mathtt{X}_t\right\|^2+O(\ln k)=O(\ln k),\ \text{ a.s.,}
\end{aligned}$$

which implies $\mathtt{X}_k=O\left(\frac{\sqrt{\ln k}}{k^{h-1/2}}\right)$. The lemma is thereby proved. ∎

REFERENCES

[1] D. Alistarh, D. Grubic, J. Li, R. Tomioka, and M. Vojnovic, *QSGD: Communication-efficient SGD via gradient quantization and encoding,* in Proceedings of Advances in Neural Information Processing Systems, Long Beach, CA, 2017.

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

[3] R. Carli and F. Bullo, *Quantized coordination algorithms for rendezvous and deployment*, SIAM J. Control Optim., 48 (2009), pp. 1251–1274.

[4] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri, *Communication constraints in the average consensus problem*, Automatica, 44 (2008), pp. 671–684.

[5] M. Carpentiero, V. Matta, and A. H. Sayed, *Distributed Adaptive Learning Under Communication Constraints*, IEEE Open Journal of Signal Processing, 5 (2023), pp. 321–358.

[6] H. F. Chen, *Stochastic Approximation and Its Applications*, Kluwer Academic Publishers, Dordrecht, 2003.

[7] D. Gan and Z. Liu, *Distributed order estimation of ARX model under cooperative excitation condition*, SIAM J. Control Optim., 60 (2022), pp. 1519–1545.

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

[9] F. Gustafsson and R. Karlsson, *Generating dithering noise for maximum likelihood estimation from quantized data*, Automatica, 49 (2013), pp. 554—560.

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

[11] S. Kar, G. Hug, J. Mohammadi, and J. M. Moura, *Distributed state estimation and energy management in smart grids: A consensus+innovations approach*, IEEE J. Sel. Top. Signal Process., 8 (2014), pp. 1022–1038.

[12] S. Kar, J. M. Moura, and H. V. Poor, *Distributed linear parameter estimation: Asymptotically efficient adaptive strategies*, SIAM J. Control Optim., 51 (2013), pp. 2200–2229.

[13] S. Kar, J. M. Moura, and K. Ramanan, *Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication*, IEEE Trans. Inform. Theory, 58 (2012), pp. 3575–3605.

[14] J. M. Ke, Y. L. Zhao, and J. F. Zhang, *Signal comparison average consensus algorithm under binary-valued communications*, in Proceedings of the IEEE Conference on Decision and Control, Singapore, 2023.

[15] X. Lao, W. Du, and C. Li, *Quantized distributed estimation with adaptive combiner*, IEEE Trans. Signal Inf. Process. Netw., 8 (2022), pp. 187–200.
[16] J. Li and G. AlRegib, *Distributed estimation in energy-constrained wireless sensor networks*, IEEE Trans. Signal Process., 57 (2009), pp. 3746–3758.
[17] T. Li, M. Fu, L. Xie, and J. F. Zhang, *Distributed consensus with limited communication data rate*, IEEE Trans. Automat. Control, 56 (2011), pp. 279–292.
[18] S. Liu, T. Li, and L. Xie, *Distributed consensus for multiagent systems with communication delays and limited data rate*, SIAM J. Control Optim., 49 (2011), pp. 2239–2262.
[19] D. Jakovetic, M. Vukovic, D. Bajovic, A. K. Sahu, and S. Kar, *Distributed recursive estimation under heavy-tail communication noise*, SIAM J. Control Optim., 61 (2023), pp. 1582–1609.
[20] Y. Meng, T. Li, and J. F. Zhang, *Finite-level quantized synchronization of discrete-time linear multiagent systems with switching topologies*, SIAM J. Control Optim., 55 (2017), pp. 275–299.
[21] N. Michelusi, G. Scutari, and C. S. Lee, *Finite-bit quantization for distributed algorithms with linear convergence*, IEEE Trans. Inform. Theory, 68 (2022), pp. 7254–7280.
[22] R. Nassif, S. Vlaski, M. Carpentiero, V. Matta, M. Antonini, and A. H. Sayed, *Quantization for decentralized learning under subspace constraints*, IEEE Trans. Signal Process., 71 (2023), pp. 2320–2335.
[23] A. K. Sahu, D. Jakovetić, and S. Kar, *$\mathcal{CIRFE}$: A distributed random fields estimator*, IEEE Trans. Signal Process., 66 (2018), pp. 4980–4995.
[24] A. H. Sayed, *Diffusion adaptation over networks*, in Academic Press Library in Signal Processing, R. Chellapa and S. Theodoridis, eds., 2014, pp. 323–454.
[25] M. O. Sayin and S. S. Kozat, *Compressive diffusion strategies over distributed networks for reduced communication load*, IEEE Trans. Signal Process., 62 (2014), pp. 5308–5323.
[26] A. N. Shiryaev, *Probability*, 2nd ed., Springer, New York, 1996.
[27] S. Yuan, C. P. Yu, and J. Sun, *Hybrid adaptive event-triggered consensus control with intermittent communication and control updating*, J. Syst. Sci. Complex., (2024), https://doi.org/10.1007/s11424-024-3563-8.
[28] S. Yang, W. Xu, W. He, and J. Cao, *Distributed generalized Nash equilibrium seeking: event-triggered coding-decoding-based secure communication*, Sci. China Inf. Sci., 67 (2024), 172205.
[29] J. Wang, T. Li, and X. Zhang, *Decentralized cooperative online estimation with random observation matrices, communication graphs and time delays*, IEEE Trans. Inform. Theory, 67 (2021), pp. 4035–4059.
[30] J. M. Wang, J. W. Tan, and J. F. Zhang, *Differentially private distributed parameter estimation*, J. Syst. Sci. Complex., 36 (2023), pp. 187–204.
[31] L. Y. Wang, J. F. Zhang, and G. Yin, *System identification using binary sensors*, IEEE Trans. Automat. Control, 48 (2003), pp. 1892-1907.
[32] T. Wang, H. Zhang, and Y. L. Zhao, *Consensus of multi-agent systems under binary-valued measurements and recursive projection algorithm*, IEEE Trans. Automat. Control, 65 (2019), pp. 2678–2685.
[33] Z. Wang, C. Jin, W. He, M. Xiao, G. P. Jiang, and J. D. Cao, *Event-triggered impulsive synchronization of heterogeneous neural networks*, Sci. China Inf. Sci., 5 (2024), 129202.
[34] X. Wu, B. Mao, X. Wu, and J. Lü, *Dynamic event-triggered leader-follower consensus control for multiagent systems*, SIAM J. Control Optim., 60 (2022), pp. 189–209.
[35] S. L. Xie and H. R. Li, *Distributed LMS estimation over networks with quantised communications*, Internat. J. Control, 86 (2013), pp. 478–492.
[36] S. Xie and L. Guo, *Analysis of normalized least mean squares-based consensus adaptive filters under a general information condition*, SIAM J. Control Optim., 56 (2018), pp. 3404–3431.
[37] Q. Zhang and J. F. Zhang, *Distributed parameter estimation over unreliable networks with Markovian switching topologies*, IEEE Trans. Automat. Control, 57 (2012), pp. 2545–2560.
[38] Y. L. Zhao, T. Wang, and W. Bi, *Consensus protocol for multiagent systems with undirected topologies and binary-valued communications*, IEEE Trans. Automat. Control, 64 (2019), pp. 206–221.
[39] S. Zhu, C. Chen, J. Xu, X. Guan, L. Xie, and K. H. Johansson, *Mitigating quantization effects on distributed sensor fusion: A least squares approach*, IEEE Trans. Signal Process., 66 (2018), pp. 3459–3474.
[40] S. Zhu, Y. C. Soh, and L. Xie, *Distributed parameter estimation with quantized communication via running average*, IEEE Trans. Signal Process., 63 (2015), pp. 4634–4646.
[41] V. A. Zorich, *Mathematical Analysis I*, 2nd ed., Springer, New York, 2015.