SIAM J. CONTROL OPTIM.
Vol. 60, No. 2, pp. 992–1017

EVENT-TRIGGERED DISTRIBUTED ESTIMATION WITH DECAYING COMMUNICATION RATE*

XINGKANG HE†, YU XING‡, JUNFENG WU§, AND KARL H. JOHANSSON†

Abstract. We study distributed estimation of a high-dimensional static parameter vector through a group of sensors whose communication network is modeled by a fixed directed graph. Different from existing time-triggered communication schemes, an event-triggered asynchronous scheme is investigated in order to reduce communication while preserving estimation convergence. A distributed estimation algorithm with a single step size is first proposed based on an event-triggered communication scheme with a time-dependent decaying threshold. With the event-triggered scheme, each sensor sends its estimate to neighbor sensors only when the difference between the current estimate and the last sent-out estimate is larger than the triggering threshold. Different sensors can have different step sizes and triggering thresholds, enabling the parameter estimation process to be conducted in a fully distributed way. We prove that the proposed algorithm has mean-square and almost-sure convergence, respectively, under an integrated condition of sensor network topology and sensor measurement matrices. The condition is satisfied if the topology is a balanced digraph containing a spanning tree and the system is collectively observable. The collective observability is the possibly mildest condition, since it is a spatially and temporally collective condition of all sensors and allows sensor measurement matrices to be time-varying, stochastic, and nonstationary. Moreover, we provide estimates for the convergence rates, which are related to the step size as well as the triggering threshold. Furthermore, as an essential metric of sensor communication intensity in the event-triggered distributed algorithms, the communication rate is proved to decay to zero with a certain speed almost surely as time goes to infinity. In addition, we show that it is feasible to tune the threshold and the step size such that requirements of algorithm convergence and communication rate decay are satisfied simultaneously. We also show that given the step size, adjusting the decay speed of the triggering threshold can lead to a tradeoff between the convergence rate of the estimation error and the decay speed of the communication rate. Specifically, increasing the decay speed of the threshold would make the communication rate decay faster but reduce the convergence rate of the estimation error. Numerical simulations are provided to illustrate the developed results.

Key words. distributed estimation, sensor network, event-triggered communications, communication rate

AMS subject classifications. 93E12, 93A14, 93A15, 93E10

DOI. 10.1137/21M1405083

1. Introduction.

Networked systems monitoring complex environments generate a tremendous amount of data, which can be used in the estimation of system states or unknown parameters. Parameter estimation is one of the most important tools in control theory, signal processing, and machine learning with extensive applications in sensor networks, weather prediction, cyber-physical systems, environmental

---

*Received by the editors March 15, 2021; accepted for publication (in revised form) December 19, 2021; published electronically April 19, 2022.
https://doi.org/10.1137/21M1405083
Funding: This work was supported by the Knut and Alice Wallenberg Foundation and the Swedish Research Council. The work of the third author was supported by the National Natural Science Foundation of China grant 61790573.

†Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden (xingkang@kth.se, kallej@kth.se).

‡Corresponding author. Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden (yuxing2@kth.se).

§School of Data Science, Chinese University of Hong Kong (Shenzhen), Shenzhen, Guangdong Province, P. R. China; College of Control Science and Engineering, Zhejiang University, Hangzhou, Zhejiang Province, P. R. China (junfengwu@cuhk.edu.cn).

monitoring, transportation, etc. The development of computational and energy efficient algorithms, which are able to handle imperfect data from sensor networks, is drawing more and more attention.

1.1. Motivation and challenges. To estimate a high-dimensional parameter vector (such as temperature over a large environment) is usually infeasible by a local algorithm for each sensor only with local measurements. Thus, it is necessary to design a collaborative estimation algorithm for sensors to obtain accurate estimates of the parameter. Centralized and distributed architectures are the two dominating approaches to sensor network estimation. In centralized architecture, a data center runs estimation algorithms based on all sensor data in order to estimate a state or system parameter. In this way, optimality in certain sense can be ensured, such as the Kalman filter achieving the minimum variance unbiased estimate for linear dynamical systems under Gaussian noise. However, with large-scale deployment of sensors, centralized architecture is no longer efficient. Therefore, we need to develop effective and efficient distributed estimation algorithms based on local sensor data and local sensor communication.

In distributed parameter estimation, messages shared between sensors play important roles. Since sensor measurements only contain partial parameter information, it is not sufficient to design a global parameter estimation algorithm only by communicating sensor measurements. Quite a few distributed estimators are proposed by requiring that sensors send their estimates to neighbor sensors at *every* time instant of measurement sampling. Diffusion-based distributed estimators are proposed in [4] and [5] by solving least-squares and least-mean squares problems over networks. In [33], a distributed estimator based on sensor beliefs is proposed over a strongly connected communication network. In [21], a gossip distributed estimator is proposed to handle random failures of network links. The results are extended in [24] to nonlinear systems under imperfect communication channels and extended to more general system models and communication networks in [23], where an adaptive learning method is proposed in order to achieve asymptotic efficiency. A robust distributed parameter estimator is given in [38] for systems with Markovian switching networks and uncertain measurement models. However, since the above references require that each sensor communicates with its neighbors persistently, there will be severe energy consumption and serious channel congestion if the measurement sampling rate is high or the parameter is high-dimensional. Under constrained communication resources, although the above algorithms can still work by requiring that sensors communicate periodically, their estimation performance could be much degraded.

Therefore, it is vital to develop a performance-guaranteed distributed estimation algorithm based on an effective and efficient sensor communication mechanism. Nevertheless, there are two challenges. First, for potentially time-varying, stochastic, and nonstationary sensor measurement models with weak observability (e.g., collective observability), how to guarantee the performance of the distributed algorithm under limited communications is challenging. Second, under the influence of noisy measurements, how to quantify the communication frequency of sensors is difficult.

1.2. Related work. To mitigate the issue of sensor communications, there are some approaches, including data quantization, compressed sensing, adaptive sampling, and event-triggered communications. In the following discussion, we focus our attention on adaptive sampling and event-triggered communications, since they can make better use of posterior information compared to the former two approaches.

In parameter estimation under constraint on the total sensing effort, adaptive sampling helps improve performance by strategically allocating sensing effort in future data collection based on the information extracted from data collected previously. In [17], a sequential adaptive sampling-and-refinement procedure called distilled sensing is proposed to detect and estimate parameters. The reference [3] investigates how to estimate the support set of a sparse parameter by making full use of structural information. For a class of physically constrained sensing models, the limitations and advantages of adaptive sampling are analyzed in [10], and it is shown that a constrained adaptive sampling method can substantially improve estimation performance. Moreover, the adaptive sampling problem for a class of continuous-time Markov state processes is studied in [32]. The above and related literature require either structural information or prior distribution of parameters, which is difficult to satisfy in monitoring complex environments. In addition, many methods in the literature are introduced in centralized architecture. It is difficult to extend these methods to distributed architecture with large-scale sensor networks and high-dimensional system parameters.

In contrast, an event-triggered scheme is suitable to distributed architecture under constrained communication resource, since it can efficiently determine when a sensor should share data to other sensors without prior knowledge of parameter structure. In such a way, event-triggered sensor communications are usually asynchronous and aperiodic and, thus different from traditional time-triggered communications. A number of centralized estimators with event-triggered measurement schedulers are proposed in the literature, such as [34, 36] for state estimation of dynamical systems and [11, 16, 37] for static parameter estimation. Regarding event-triggered parameter estimation, the authors in [37] propose a measurement scheduler such that the asymptotic estimation performance is optimized. A stochastic measurement scheduler is studied in [16] to compensate for the loss of the Gaussianity of the system, which ensures the maximum-likelihood parameter estimator. An event-triggered scheduler for finite impulse response systems with binary measurements is proposed in [11], where communication rate is analyzed. The above centralized event-triggered measurement schedulers are usually not suitable to the distributed architecture, because scheduling local measurements between neighboring sensors is insufficient to design a global parameter estimator when each sensor can only observe partial elements of the parameter vector. In the distributed architecture for event-triggered estimation, there are several methods for dynamical systems. For example, an approach based on linear matrix inequality is studied in [29] for the cooperative estimation and control of a dynamical system under an event-triggered communication protocol. An event-triggered distributed Kalman filter is proposed in [1] and proved to be stable in terms of meanssquare boundedness of the estimation error in each sensor. For linear dynamical systems under state equality constraints, an event-triggered projected distributed Kalman filter is studied in [18] with guaranteed estimation error stability under collective system observability. For more related works on event-triggered distributed estimation for dynamical systems, we refer readers to [12] and the references therein. Another related topic is event-triggered distributed optimization, which concerns the distributed design of optimization algorithms such that the global optimal solution is reached by each computational node. In this direction, most algorithms are proposed in noise-free or disturbance-free settings and analyzed to show the influence of event-triggered mechanisms on algorithm convergence [2, 26]. An event-triggered property for continuous-time systems, which is called Zeno behavior (i.e., an infinite number of events occur in a finite amount of time), is also of interest in some papers [8].

However, there are few event-triggered distributed optimization or estimation algorithms handling imperfect data from stochastic environments. In addition, the above studies analyze the influence of event-triggered threshold to estimation performance, but the tradeoff between communication frequency and estimation performance is not well established in the distributed architecture. To the best of our knowledge, there is no result of distributed parameter estimation on communication rate, which is an essential metric of sensor network communication level.

1.3. Contributions. In this paper, we study the event-triggered distributed parameter estimation problem for the sake of reducing sensor communication while preserving convergence of estimators. The main contributions are summarized in the following:

1. We propose a recursive event-triggered distributed parameter algorithm with a single step size (Algorithm 3.1) for a group of sensors with noisy measurements. The algorithm has several advantages: First, it is fully distributed in the sense that each sensor only relies on the local information and has its own step size and triggering threshold. Second, without requiring exact noise distribution or statistics, the event-triggered scheme enables sensors to communicate in an efficient way such that each sensor sends its estimate to neighboring sensors only when the difference between the current estimate and the last sent-out estimate is larger than the triggering threshold. Third, the algorithm is scalable to large-scale sensor networks, since its update is independent of the network size.

2. We prove that the proposed algorithm with a properly designed step size and triggering threshold achieves mean-square convergence (Theorem 3.6). In addition, we provide the estimate for the convergence rate and establish its connection to network structure and system observability. The results are obtained under an integrated condition of sensor network topology and sensor measurement matrices. The condition is satisfied if the topology is a balanced digraph containing a spanning tree and the system is collectively observable. The collective observability is the possibly mildest condition, since it allows each sensor to observe partial elements of the parameter, sensor measurement matrices to be time-varying, stochastic, and nonstationary, and noise processes to be martingale difference sequences. Under some extra conditions on step size and triggering threshold, we prove the algorithm's output is asymptotically convergent to the true parameter almost surely (a.s.) with an estimated convergence rate (Theorem 3.8).

3. We prove that the communication rate is decaying to zero a.s. as time goes to infinity (Theorem 3.12) with a quantified speed. Moreover, we show that it is feasible to tune the threshold and the step size such that the requirements of the algorithm convergence and the communication rate decay are satisfied simultaneously. We also show that given the step size, adjusting the decay speed of the triggering threshold can lead to a tradeoff between the convergence rate of the estimation error and the decay speed of the communication rate. Specifically, increasing the decay speed of the threshold would make the communication rate decay faster but reduce the convergence rate of the estimation error. To the best of our knowledge, this is the first result on communication rate in the direction of event-triggered distributed algorithms for estimating static parameters or solutions [2, 8]. Our result indicates that while ensuring successful parameter estimation, the algorithm enables the alleviation of channel burden and resource consumption in sensor communication compared to existing time-triggered approaches [5, 21, 23, 25, 33, 35, 38] which require a persistent positive communication rate.

The results of this paper are significantly different from the literature. Regarding the step size, we remove the requirement that all sensors share the same setting [9, 19, 21, 23–25, 35, 38], as well as the requirements of the concrete forms [9, 19, 21, 23–25] and the monotonicity in [35]. The stationarity condition of measurement matrices in [38] is removed. Since sensor measurements and measurement matrices are not shared between neighbors, our algorithm can be more suitable to scenarios with a privacy requirement than the diffusion estimators in [4, 6]. Moreover, we generalize sensor communication topologies from undirected graphs [19, 21, 23, 24] to a class of directed graphs, such as balanced digraphs containing a spanning tree.

1.4. Paper organization. In section 2, we formulate the considered problem of event-triggered distributed parameter estimation and introduce some graph preliminaries, the communication model, and the mathematical model of sensors. In order to solve this problem, in section 3 an event-triggered distributed estimation algorithm is proposed and analyzed in terms of estimation error convergence (mean-square and almost-sure convergence) and communication rate. For reading convenience, the proofs for the results in section 3 are provided in section 4. Numerical simulations in two examples are provided in section 5 to illustrate the developed results. Section 6 concludes the paper.

Notations. Denote $\mathbb{R}^{n\times m}$ the set of real-valued matrices with n rows and m columns, with $\mathbb{R}^n = \mathbb{R}^{n\times 1}$ and $\mathbb{R}^1 = \mathbb{R}$. Let $\mathbb{R}^+$ and $\mathbb{N}^+$ be the sets of positive real-valued scalars and integers, respectively, with $\mathbb{N} = \mathbb{N}^+ \cup \{0\}$. Denote Ω and $\emptyset$ the universal set and empty set, respectively. I_n stands for the n-dimensional square identity matrix. $\mathbf{1}_n$ stands for the n-dimensional vector with all elements being one. The superscript "T" represents the transpose. $\mathbb{E}\{x\}$ denotes the mathematical expectation of the random variable x, and blockdiag$\{\cdot\}$ represents the diagonalizations of block elements. $A\otimes B$ is the Kronecker product of A and B. $\|x\|$ is the 2-norm of a vector x, and $\|A\|$ is the induced norm, i.e., $\|A\| = \sup_{\|x\|=1} \|Ax\|$, where $A \in \mathbb{R}^{n\times m}$, $x \in \mathbb{R}^m$. The mentioned scalars, vectors, and matrices of this paper are all real-valued. Let $\sigma(\cdot)$ be the σ-algebra operator which generates the smallest σ-algebra. For a real-valued matrix or vector sequence $\{a(t)\}$ and a real number sequence $\{b(t)\}$, the operator $a(t) = O(b(t))$ means that there is a constant $c \geq 0$ such that for each element sequence of $\{a_i(t)\}$, $\lim_{t\to\infty} |a_i(t)/b(t)| \leq c$, and the operator $a(t) = o(b(t))$ means $\lim_{t\to\infty} |a_i(t)/b(t)| = 0$.

2. Preliminaries and problem formulation.

In this section, we formulate the problem of event-triggered distributed parameter estimation and introduce some graph preliminaries, the communication model, and the mathematical model of sensors.

2.1. Graph preliminaries. In this paper, the communication between N sensors of a network is modeled as a digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, where $\mathcal{V} = \{1, 2, \ldots, N\}$ is the node set and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set. A directed edge $(i, j) \in \mathcal{E}$ if and only if there is a communication link from j to i, where j is called the parent node and i is called the child node. The matrix $\mathcal{A} = [a_{i,j}]_{i,j=1}^N$ is the weighted adjacency matrix with $a_{i,j} \geq 0$, where $a_{i,j} > 0$, if $(i, j) \in \mathcal{E}$. The parent neighbor set and child neighbor set of node i are denoted by $\{j \in \mathcal{V}|(i, j) \in \mathcal{E}\} \triangleq \mathcal{N}_i$ and $\{j \in \mathcal{V}|(j, i) \in \mathcal{E}\} \triangleq \mathcal{N}_i^c$, respectively. Suppose that the graph has no self loop, which means $a_{i,i} = 0$ for any $i \in \mathcal{V}$. $\mathcal{G}$ is called a balanced digraph if $\sum_{j=1}^N a_{i,j} = \sum_{i=1}^N a_{j,i}$ for all $i \in \mathcal{V}$. $\mathcal{G}$ is called an undirected graph if $\mathcal{A}$ is symmetric. The Laplacian matrix of $\mathcal{G}$ is denoted by $\mathcal{L} = \mathcal{D} - \mathcal{A}$, where $\mathcal{D} = \mathrm{diag}\{\sum_{j=1}^N a_{1,j}, \ldots, \sum_{j=1}^N a_{N,j}\}$. The mirror graph of the digraph $\mathcal{G}$ is an undirected graph, denoted by $\bar{\mathcal{G}} = (\mathcal{V}, \mathcal{E}_{\bar{\mathcal{G}}}, \mathcal{A}_{\bar{\mathcal{G}}})$ with $\mathcal{A}_{\bar{\mathcal{G}}} = [\bar{a}_{i,j}]_{i,j=1}^N$, $\bar{a}_{i,j} = \bar{a}_{j,i} = (a_{i,j} + a_{j,i})/2$ [30]. $\mathcal{G}$ is called strongly connected if for any pair nodes (i_1, i_l), $l > 1$, there exists a path from i_l to i_1 consisting of edges in the set $\{(i_m, i_{m+1}) \in \mathcal{E} | m = 1, 2, \ldots, l-1\}$. We call $\mathcal{G}$ is connected if it is strongly connected and undirected. A directed tree is a digraph, where each node except the root has exactly one parent node. A spanning tree of $\mathcal{G}$ is a directed tree whose node set is $\mathcal{V}$ and whose edge set is a subset of $\mathcal{E}$.

2.2. Problem setup. Consider an unknown high-dimensional parameter vector $\theta \in \mathbb{R}^M$ observed by $N > 0$ sensors with the following model:

$$y_i(t) = H_i(t)\theta + v_i(t), \quad i = 1, 2, \ldots, N, \tag{2.1}$$

where $y_i(t) \in \mathbb{R}^{m_i}$ is the measurement vector, $v_i(t) \in \mathbb{R}^{m_i}$ is the measurement noise, and $H_i(t) \in \mathbb{R}^{m_i \times M}$ represents the known measurement matrix of sensor i, all at time t.

The communication rate is an essential metric of communication intensity in the event-triggered distributed algorithms. Its mathematical definition over the digraph $\mathcal{G}$ is in the following.

DEFINITION 2.1. *For the digraph $\mathcal{G}$, in a given time interval $[0, t] \cap \mathbb{N}$, the* **communication rate** $\lambda_c(t)$ *is given by*

$$\lambda_c(t) = \frac{\sum_{i \in \mathcal{V}} K_i(t) |\mathcal{N}_i^c|}{t \sum_{i \in \mathcal{V}} |\mathcal{N}_i^c|}, \tag{2.2}$$

where $K_i(t)$ is the accumulated triggering (data-sending) times of node i in $[0, t] \cap \mathbb{N}$ and $|\mathcal{N}_i^c|$ is the child neighbor number of node i.

According to this definition, $\lambda_c(t) \in [0, 1]$. When $\lambda_c(t) \equiv 1$, communication occurs all the time, which is the case for the time-triggered distributed estimation algorithms [5, 21, 23, 33, 35, 38]. When $\lambda_c(t) \equiv 0$, there is no sensor communication over the network.

The problem considered in this paper is how to design a distributed estimation algorithm with an event-triggered communication scheme and find conditions such that the output of the algorithm is asymptotically convergent to the parameter vector with a convergence rate while the communication rate of the sensor network is decaying to zero as time goes to infinity.

3. Main results. The formulated problem is solved in this section. First, an event-triggered distributed algorithm is proposed to estimate the parameter. Then, the mean-square and almost-sure convergence of the algorithm are studied. Additionally, the decay speed of the communication rate is analyzed. The proofs of these results are provided in section 4.

3.1. Event-triggered distributed estimation. To solve the problem posed in section 2, we propose the event-triggered distributed estimation algorithm in Algorithm 3.1 for each sensor $i \in \mathcal{V}$, where the initial estimate $x_i(0) = x_{i,0} \in \mathbb{R}^M$ is fixed. The time complexity of Algorithm 3.1 is $O(n^2 m_i + n|\mathcal{N}_i|)$. An example of Algorithm 3.1 for a sensor network with 7 nodes is provided in Figure 1.

Remark 3.1. We have a few remarks on Algorithm 3.1.

1. The notation $x_i(t)$ denotes the estimate of θ by sensor i at time t. The notation $\tau_{k_i(t)}$ denotes the time of $k_i(t)$th triggering instant until time t. The amount

Algorithm 3.1 Event-Triggered Distributed Estimation

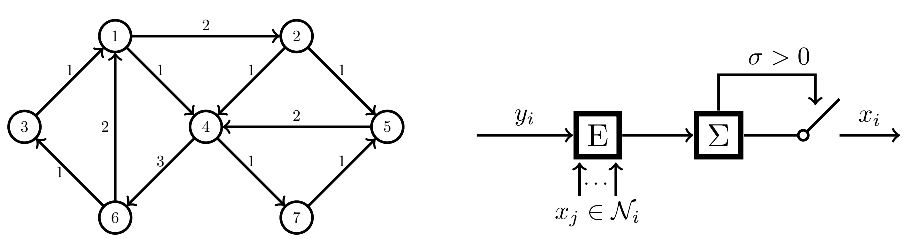
```
Input: x_i(0), {y_i(t)}_{t≥0}, {H_i(t)}_{t≥0}, {α_i(t)}_{t≥0}, {f_i(t)}_{t≥0}.
Output: {x_i(t)}_{t≥0}.
for t = 0, 1, ... do
    // Data transmission
    if t = 0 then
        Send x_i(0) to each child neighbor sensor j ∈ N_i^c, and let x_i(τ_0) = x_i(0) and k_i(0) = 1.
    else if ‖x_i(t) − x_i(τ_{k_i(t−1)})‖ > f_i(t) then
        Send x_i(t) to each child neighbor sensor j ∈ N_i^c, and let k_i(t) = k_i(t − 1) + 1 and τ_{k_i(t)} := t.
    else
        k_i(t) = k_i(t − 1).
    end if

    // Data receiving
    for j ∈ N_i do
        if Receive estimate x_j(t) from parent neighbor sensor j ∈ N_i then
            Let k_j(t) = k_j(t − 1) + 1 and τ_{k_j(t)} := t.
        else
            k_j(t) = k_j(t − 1).
        end if
    end for

    // Estimate update
```

$$
\begin{aligned}
x_i(t+1) = & x_i(t) + \alpha_i(t) H_i^{\mathsf{T}}(t)(y_i(t) - H_i(t)x_i(t)) \\
& + \alpha_i(t) \sum_{j \in \mathcal{N}_i} a_{i,j}(x_j(\tau_{k_j(t)}) - x_i(t)).
\end{aligned} \tag{3.1}
$$

```
end for
```



(a) Communication topology of a sensor network with 7 nodes.

(b) The schematic of Algorithm 3.1 for sensor i.

Fig. 1. *An example of Algorithm* 3.1 *for a sensor network with* 7 *nodes. The communication of sensors forms a weighted balanced digraph with a spanning tree, as shown in* (a) *where the numbers on edges stand for weights. The schematic of Algorithm* 3.1 *for sensor* $i \in \{1, 2, \ldots, 7\}$ *is shown in* (b), *consisting of a parameter estimator* E *and a communication scheduler* Σ. *The estimator has inputs of local measurement* y_i *and parent neighbor estimates* $x_j \in \mathcal{N}_i$. *The scheduler* Σ *sends estimate* x_i *to child neighbor nodes when the event is triggered, i.e.,* $\sigma > 0$, *where* $\sigma := \|x_i(t) - x_i(\tau_{k_i(t-1)})\| - f_i(t)$.

of $k_i(t)$ shows how many events have been triggered for sensor i until time t. The scalar $\alpha_i(t)$ is the step size of sensor i at time t, and the scalar $f_i(t) \geq 0$, $i \in \mathcal{V}$, is the event-triggered threshold of sensor i at time t. Both $\{\alpha_i(t)\}_{t\geq 0}$ and $\{f_i(t)\}_{t\geq 0}$ are to be designed.

2. The event-triggered scheme $\|x_i(t) - x_i(\tau_{k_i(t-1)})\| > f_i(t)$ is to determine whether the estimate $x_i(t)$ is worth sharing with child nodes by comparing it with the last sent-out estimate $x_i(\tau_{k_i(t-1)})$. Since the estimate $x_i(t)$ provides the global parameter information, in weak observability conditions the scheme outweighs the existing event-triggered schemes transmitting local measurements [11, 16, 37]. Moreover, compared to the existing schemes [16, 34, 36, 37], the proposed scheme is built on a more general stochastic framework without requiring knowledge of accurate noise distribution or statistics.
3. To ensure the convergence of Algorithm 3.1, the triggering threshold $f_i(t)$ should decay to zero fast enough, as required in Assumption 3.1. However, to avoid sensors communicating frequently all the time, $f_i(t)$ should not decay too fast, as required in Assumption 3.2. Remark 3.11 will show that it is feasible to satisfy the assumptions on $f_i(t)$ simultaneously. Moreover, as shown in Remark 3.13, the decay speed of $f_i(t)$ can lead to a tradeoff between the decaying speed of the communication rate and the convergence rate of the estimation error.
4. Regarding the time instants $\{\tau_{k_i(t)}\}$, it holds that $\tau_{k_i(t)} = \inf_{t>\tau_{k_i(t-1)}} t$ if $\|x_i(t) - x_i(\tau_{k_i(t-1)})\| > f_i(t)$ otherwise $\tau_{k_i(t)} = \tau_{k_i(t-1)}$. Thus, $\{\tau_{k_i(t)}\}$ is determined only by past and current information.

Remark 3.2. We have a few remarks on the advantages of Algorithm 3.1 in comparison to existing algorithms.

1. An advantage of Algorithm 3.1 compared to the diffusion estimation algorithms in [4, 6] is that Algorithm 3.1 does not require the local measurements and measurement matrices to be shared between sensors. Thus, Algorithm 3.1 is more suitable to scenarios with a privacy requirement and limited communication bandwidth.
2. Algorithm 3.1 does not require any global knowledge of the system and thus is fully distributed. For example, it removes the requirement in [19, 21, 23, 24, 35, 38] that all sensors share the same step size. Moreover, Algorithm 3.1 is able to handle open sensor networks where some sensors may break down or new sensors are plugged in, which however is intractable for the algorithms [21] requiring the total sensor number and the measurement matrices of all sensors.
3. Since Algorithm 3.1 can tremendously reduce the redundant transmissions of estimates, it can require less communication than the existing time-triggered distributed algorithms [5, 22, 23, 35, 38] for convergence.

To ease notation, we write $\tau_{k_i(t)}$ as τ_k^i in the below text, i.e., $\tau_k^i := \tau_{k_i(t)}$. The subscript k of τ_k^i is kept to emphasize the number of triggering times, but the reader should keep in mind that k in τ_k^i depends on time t and sensor i. With this definition, $x_i(\tau_k^i) := x_i(\tau_{k_i(t)})$. Then we rewrite (3.1) in the following way:

$$\begin{aligned}(3.2)\qquad x_i(t+1) =& x_i(t) + \alpha_i(t)H_i^{\mathsf{T}}(t)(y_i(t) - H_i(t)x_i(t)) \\ &+ \alpha_i(t)\sum_{j\in\mathcal{N}_i} a_{i,j}(x_j(t) - x_i(t)) + \alpha_i(t)\sum_{j\in\mathcal{N}_i} a_{i,j}(x_j(\tau_k^j) - x_j(t)).\end{aligned}$$

3.2. Convergence and convergence rate. In order to ensure that Algorithm 3.1 in the previous subsection provides accurate parameter estimates, in this subsection we find conditions such that the output of Algorithm 3.1, i.e., $x_i(t)$, is asymptotically convergent to θ with an estimated convergence rate. To proceed, we introduce the following notations:

$$\begin{aligned}
\bar{\alpha}(t) &= \mathrm{blockdiag}\left\{\alpha_1(t)I_M, \ldots, \alpha_N(t)I_M\right\},\\
\bar{D}_H(t) &= \mathrm{blockdiag}\left\{H_1^\mathsf{T}(t), \ldots, H_N^\mathsf{T}(t)\right\},\\
\tau_k &= \left[\tau_k^1, \ldots, \tau_k^N\right]^\mathsf{T}, \qquad f_{\max}(t) = \max_{i\in\mathcal{V}} f_i(t), \qquad f_{\min}(t) = \min_{i\in\mathcal{V}} f_i(t),\\
V(t) &= [v_1^\mathsf{T}(t), \ldots, v_N^\mathsf{T}(t)]^\mathsf{T}, \qquad X(t) = \left[x_1^\mathsf{T}(t), \ldots, x_N^\mathsf{T}(t)\right]^\mathsf{T},\\
X(\tau_k) &= \left[x_1^\mathsf{T}(\tau_k^1), \ldots, x_N^\mathsf{T}(\tau_k^N)\right]^\mathsf{T}, \qquad Y(t) = \left[y_1^\mathsf{T}(t), \ldots, y_N^\mathsf{T}(t)\right]^\mathsf{T}.
\end{aligned} \tag{3.3}$$

Given the notations in (3.3), the compact form of (3.2) is given in the following:

$$\begin{aligned}
X(t+1) =& X(t) - \bar{\alpha}(t)(\mathcal{L}\otimes I_M)X(t) + \bar{\alpha}(t)\bar{D}_H(t)(Y(t) - \bar{D}_H^\mathsf{T}(t)X(t))\\
&+ \bar{\alpha}(t)(\mathcal{A}\otimes I_M)(X(\tau_k) - X(t)).
\end{aligned} \tag{3.4}$$

In the basic probability space $(\Omega, \mathcal{F}, P)$, define the filtration $\mathcal{F}(t) := \sigma(\bar{D}_H(s), V(s), 0 \le s \le t)$, $t \ge 0$, and $\mathcal{F}(-1) := \{\emptyset, \Omega\}$. With the notations in (3.3), the following assumption is needed in this paper.

Assumption 3.1.

(i.a) There exists a sequence $\{\alpha(t)\}$ such that $\alpha_i(t)/\alpha(t) \to 1$ as $t \to \infty$ for all $1 \le i \le N$. In addition, $\alpha(t) > 0$, $\alpha(t) \to 0$, $\sum_{t=1}^{\infty} \alpha(t) = \infty$, and

$$\frac{1}{\alpha(t+1)} - \frac{1}{\alpha(t)} \to \alpha_0 \ge 0.$$

(i.b) $\sum_{t=1}^{\infty} \alpha(t)^{2(1-\delta)} < \infty$ for some $\delta \in [0, 1/2)$.

(ii.a) The two sequences $\{\bar{D}_H(t)\}$ and $\{V(t)\}$ introduced in (3.3) are independent.

(ii.b) $\{V(t)\}$ is a martingale difference sequence, and there is a scalar $\rho > 2$ such that

$$\sup_{t\in\mathbb{N}} \mathbb{E}\{\|V(t)\|^\rho | \mathcal{F}(t-1)\} := c_V < \infty, \text{ a.s.}$$

(ii.c) $\sup_{t\in\mathbb{N}} \|\bar{D}_H(t)\|^2 \le D < \infty$ a.s. for some positive constant D, and there exist $h \in \mathbb{N}^+$, $\lambda \in \mathbb{R}^+$ such that for any $m \in \mathbb{N}$ and some $\delta \in [0, 1/2)$,

$$\lambda_{\min}\left[\sum_{t=mh}^{(m+1)h-1} \mathbb{E}\left\{\bar{\mathcal{L}} \otimes I_M + \bar{D}_H(t)\bar{D}_H^\mathsf{T}(t) | \mathcal{F}(mh-1)\right\}\right] \ge \lambda + h\alpha_0\delta, \text{ a.s.,} \tag{3.5}$$

where $\bar{\mathcal{L}} = (\mathcal{L} + \mathcal{L}^\mathsf{T})/2$.

(iii.a) $f_{\max}(t)/\alpha^{\delta}(t) \to 0$, $t \to \infty$, for some $\delta \in [0, 1/2)$.

(iii.b) $\sum_{t=1}^{\infty}((\alpha(t))^{1-\delta} f_{\max}(t)) < \infty$, for some $\delta \in [0, 1/2)$.

Remark 3.3. Some remarks on Assumption 3.1 are given:

- In (i.a), $\alpha(t)$ provides a reference rate of step sizes. It is necessary that step sizes of sensors are in the same order. If there exist i and j such that $\alpha_i(t)/\alpha_j(t) \to 0$, then the information provided by sensor i vanishes in terms of $\alpha_j(t)$. The requirements in (i.a) and (i.b) for choosing the step size $\alpha(t)$ are quite general, where $\alpha(t) = a/(t+1)$ for some $a > 0$ in [24] is a special case.

- In (ii.c), (3.5) indicates network connectivity and system observability. Note that the eigenvalue on the left-hand side of (3.5) is nonnegative. It is positive under proper conditions of network connectivity and system observability, such as in Proposition 3.4. In order to ensure a certain convergence rate of Algorithm 3.1, the eigenvalue has to be large enough. If the step size is slowly decreasing such that $\alpha_0 = 0$, then the second term on the right-hand side of (3.5) is zero. In other cases, since α_0 depends only on the step size, one can design the latter so that α_0 is small enough and (3.5) holds.
- Condition (iii.a) requires decaying thresholds $\{f_i(t)\}$ to ensure that the estimates can still be shared over an infinite horizon. Otherwise, in a convergent algorithm (its output may not converge to θ), since the difference between two adjacent estimates is decreasing to zero, the events would no longer be triggered after some finite time.
- Condition (iii.b) is required for almost-sure convergence of the algorithm, meaning that the triggering thresholds have to decrease fast enough to ensure certain convergence rate of the algorithm.

The following proposition shows that the integrated condition (3.5) in Assumption 3.1 is satisfied under certain network connectivity and system observability.

PROPOSITION 3.4. *Suppose that $\mathcal{G}$ is a balanced digraph containing a spanning tree and that there exist $h \in \mathbb{N}^+$ and $\tilde{\lambda} \in \mathbb{R}^+$ such that for any $m \in \mathbb{N}$,*

$$\lambda_{\min}\left[\sum_{j=1}^{N}\sum_{t=mh}^{(m+1)h-1}\mathbb{E}\left\{H_j^{\mathsf{T}}(t)H_j(t)\middle|\mathcal{F}(mh-1)\right\}\right] \geq \tilde{\lambda}, \quad a.s.; \tag{3.6}$$

then for any $m \in \mathbb{N}$,

$$\lambda_{\min}\left[\sum_{t=mh}^{(m+1)h-1}\mathbb{E}\left\{\bar{\mathcal{L}}\otimes I_M + \bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t)|\mathcal{F}(mh-1)\right\}\right] \geq \beta, \quad a.s.,$$

where $\beta = \min\{h\lambda_2(\bar{\mathcal{L}}), \tilde{\lambda}/N\} > 0$.

Proof. Since $\mathcal{G}$ is a balanced digraph, according to [30, Theorem 7], $\bar{\mathcal{L}} = (\mathcal{L}^T + \mathcal{L})/2$ is the Laplacian matrix of the undirected mirror graph $\bar{\mathcal{G}}$, i.e., $\mathcal{L}_{\bar{\mathcal{G}}} = \bar{\mathcal{L}}$. Then by [28, Theorem 2.8], $\bar{\mathcal{L}}$ has a unique eigenvalue zero, and $\lambda_2(\bar{\mathcal{L}}) > 0$. Hence for a vector $x \in \mathbb{R}^{NM}$ such that $x^{\mathsf{T}}(\bar{\mathcal{L}}\otimes I_M)x = 0$, it must have the form $\mathbf{1}_N \otimes z$, $z \in \mathbb{R}^M$. For another vector x, which does not have this form, we know that $x^{\mathsf{T}}(\bar{\mathcal{L}}\otimes I_M)x \geq \lambda_2(\bar{\mathcal{L}})\|x_\perp\|^2 > 0$, where $x_\perp$ is the difference of x minus its projection on the subspace $\{\mathbf{1}_N \otimes z : z \in \mathbb{R}^M\}$. Now for vector $x = \mathbf{1}_N \otimes z$, $z \in \mathbb{R}^M$,

$$\begin{aligned}
&x^{\mathsf{T}}\left(\sum_{t=mh}^{(m+1)h-1}\mathbb{E}\{\bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t)|\mathcal{F}(mh-1)\}\right)x \\
&= \sum_{t=mh}^{(m+1)h-1}\mathbb{E}\left\{\sum_{j=1}^{N} z^{\mathsf{T}}H_j^{\mathsf{T}}(t)H_j(t)z|\mathcal{F}(mh-1)\right\} \\
&= z^{\mathsf{T}}\left(\sum_{t=mh}^{(m+1)h-1}\mathbb{E}\left\{\sum_{j=1}^{N} H_j^{\mathsf{T}}(t)H_j(t)|\mathcal{F}(mh-1)\right\}\right)z \geq \tilde{\lambda}\|z\|^2,
\end{aligned}$$

where the last inequality follows from (3.6).

Consider the orthogonal decomposition of an arbitrary unit vector $x = x_\perp + \mathbf{1}_N \otimes z$, $z \in \mathbb{R}^M$. Then it holds that $\|x\|^2 = \|x_\perp\|^2 + N\|z\|^2$. It follows from the above analysis that

$$x^\mathsf{T}\left(\sum_{t=mh}^{(m+1)h-1} \mathbb{E}\left\{\bar{\mathcal{L}} \otimes I_M + \bar{D}_H(t)\bar{D}_H^\mathsf{T}(t)|\mathcal{F}(mh-1)\right\}\right)x$$

$$\geq h\lambda_2(\bar{\mathcal{L}})\|x_\perp\|^2 + \tilde{\lambda}\|z\|^2 \overset{(a)}{=} h\lambda_2(\bar{\mathcal{L}}) + (\tilde{\lambda} - Nh\lambda_2(\bar{\mathcal{L}}))\|z\|^2 \overset{(b)}{\geq} \min\{h\lambda_2(\bar{\mathcal{L}}), \tilde{\lambda}/N\} = \beta,$$

where (a) follows from $\|x\|^2 = \|x_\perp\|^2 + N\|z\|^2 = 1$, and (b) is due to $\|z\|^2 \in [0, 1/N]$. Since x is an arbitrary unit vector, the conclusion holds. ∎

Remark 3.5. The observability condition (or persistent excitation condition) in (3.6) is the possibly mildest, since it is a spatially and temporally collective condition of all sensors and allows the measurement matrices to be time-varying, stochastic, and nonstationary. Similar conditions are given in [15] and [35] under centralized and distributed settings, respectively. In the literature, time-invariant measurement matrices and stationary measurement matrices are studied in [21, 23] and [38], respectively. Under the conditions in Proposition 3.4, (3.5) is fulfilled if $\beta = \min\{h\lambda_2(\bar{\mathcal{L}}), \tilde{\lambda}/N\} \geq q\lambda + h\alpha_0\delta$.

THEOREM 3.6 (mean-square convergence). *Under Assumption* 3.1(i.a), (ii)*, and* (iii.a)*, the estimation error, $e_0(t) = X(t) - \mathbf{1}_N \otimes \theta$, converges to zero in mean square with a rate $o(\alpha^{2\delta}(t))$, i.e.,*

$$\lim_{t\to\infty} \frac{\mathbb{E}\{\|e_0(t)\|^2\}}{\alpha^{2\delta}(t)} = 0, \tag{3.7}$$

where $\delta \in [0, 1/2)$ satisfies Assumption 3.1(ii.c) *and* (iii.a).

Proof. See section 4.2. ∎

Remark 3.7. Under the conditions of Theorem 3.6, a larger λ in (3.5) leads to a larger δ meaning a faster convergence rate. From Proposition 3.4, the value of λ is determined by the network structure $\mathcal{L}$ and the level of system observability $\sum_{t=mh}^{(m+1)h-1} \mathbb{E}\{\bar{D}_H(t)\bar{D}_H^\mathsf{T}(t)|\mathcal{F}(mh-1)\}$. Thus, in order to obtain a faster convergence rate, the system designer can offline adjust the system and network structure, such as choosing sensors with larger signal-to-noise ratio, new sensor deployment, etc. Moreover, if we increase the decay speed of triggering threshold $f_i(t)$, then according to (iii) of Assumption 3.1, a larger δ will be available such that the convergence rate of the estimation error increases.

In order to analyze the decay speed of the communication rate over the whole sensor network in section 3.3, we need to establish the almost-sure convergence of Algorithm 3.1. As is known, there is a gap between the almost-sure and mean-square convergence of a random variable sequence unless the uniform integrability and certain moment conditions are satisfied [31]. For example, if $\sum_{t=0}^{\infty} \mathbb{E}\{\|e_0(t)\|^2\} < \infty$ is satisfied, almost-sure convergence will hold according to Theorem 6.8 in [31]. Although we provide the mean-square convergence rate in Theorem 3.6 with $o(\alpha^{2\delta}(t))$, the condition $\sum_{t=0}^{\infty} \mathbb{E}\{\|e_0(t)\|^2\} < \infty$ is not directly satisfied due to $\delta \in [0, 1/2)$. In the following theorem, we show that if some slightly stronger conditions on the step size and the triggering threshold are satisfied, Algorithm 3.1 will have almost-sure convergence with a certain convergence rate.

THEOREM 3.8 (almost-sure convergence). *Under Assumption* 3.1*, the estimation error, $e_0(t) = X(t) - \mathbf{1}_N \otimes \theta$, converges to zero a.s. with a rate $o(\alpha^\delta(t))$, i.e.,*

$$\mathbb{P}\left\{\lim_{t\to\infty} \frac{e_0(t)}{\alpha^\delta(t)} = 0\right\} = 1, \tag{3.8}$$

where $\delta \in [0, \frac{1}{2})$ satisfies Assumption 3.1 (i.b), (ii.c)*, and* (iii).

Proof. See section 4.3. ◻

Remark 3.9. Theorems 3.6 and 3.8 show that the mean-square and almost-sure convergence rates of Algorithm 3.1 are $o(\alpha^{2\delta}(t))$ and $o(\alpha^\delta(t))$, respectively. Since $\delta \in [0, 1/2)$, the conclusions in Theorems 3.6 and 3.8 reduce to the mean-square and almost-sure convergence [35, 38] when $\delta = 0$.

3.3. Communication rate. After studying the estimation performance of Algorithm 3.1, in this subsection we find conditions such that the Communicatio rate of sensors using the algorithm decays to zero with a certain speed while ensuring the estimation error convergence. To proceed, we need some extra conditions on the step size $\alpha_i(t)$ and the triggering threshold $f_i(t)$.

Assumption 3.2. There are two monotonically nonincreasing sequences, $\{\bar{f}(t)\}_{t=0}^\infty$ and $\{\beta(t)\}_{t=0}^\infty$, and a scalar $\mu \in [1/2, 1)$ such that the following hold:
(i) $\bar{f}(t) \le f_{\min}(t)$ for any $t \in \mathbb{N}$.
(ii) $\liminf_{t\to\infty} \bar{f}(t + t^\mu)/\bar{f}(t) > 0$.
(iii) $\beta(t) = O(\alpha(t)^{1-2(1-\delta)/\rho})$, where δ and ρ are given in Theorem 3.8 and Assumption 3.1 (ii.b), respectively.
(iv) $\liminf_{t\to\infty} \frac{\bar{f}(t)}{t^\mu \beta(t)} > 0$.

Remark 3.10. In (i) and (iii) of Assumption 3.2, although monotonicity is required for $\bar{f}(t)$ and $\beta(t)$, triggering threshold $f_i(t)$ and step size $\alpha_i(t)$ are not necessarily monotonic. Conditions (ii) and (iv) are satisfied if $\bar{f}(t)$ decays slowly, such as the case in Remark 3.11. The parameter μ can reflect the decay speed of $\bar{f}(t)$. One can find a larger μ if $\bar{f}(t)$ decays more slowly.

Remark 3.11. The requirements on the step size $\alpha_i(t)$ and the threshold $f_i(t)$ in Assumptions 3.1 and 3.2 can be satisfied at the same time. For instance, let $\alpha(t) = \alpha_i(t) = t^{-1}$, $\bar{f}(t) = f_i(t) = t^{-\epsilon_0}$, and $\beta(t) = t^{-1+2/\rho}$. Then the requirements are satisfied if $1 - \epsilon_0 - \mu - 2/\rho > 0$, $\mu \in [1/2, 1)$, and $\epsilon_0 > \delta \ge 0$, where δ indicates the convergence rate of the estimation error according to Theorems 3.6 and 3.8. It is feasible to satisfy these conditions simultaneously if the parameter ρ in Assumption 3.1(ii.b) is large enough.

THEOREM 3.12 (communication rate decay). *Under Assumptions* 3.1 *and* 3.2*, the communication rate $\lambda_c(t)$ of Algorithm* 3.1 *converges to zero a.s. with a rate $o(t^{-\gamma})$ for any $\gamma \in [0, \frac{2\mu}{2\mu+1})$, i.e.,*

$$\mathbb{P}\left\{\lim_{t\to\infty} \lambda_c(t) t^\gamma = 0\right\} = 1. \tag{3.9}$$

Proof. See section 4.4. ◻

Remark 3.13. (tradeoff). Given the step size $\alpha_i(t)$, if we reduce the decay speed of the triggering threshold $f_i(t)$, we are able to obtain a larger parameter μ according to Remark 3.10. Then according to Theorem 3.12, the communication rate $\lambda_c(t)$ will have a faster decay speed. However, the convergence rate of the estimation error will be reduced since δ in Theorems 3.6 and 3.8 becomes smaller according to (iii) of Assumption 3.1. Thus, the decay speed of the triggering threshold can lead to a tradeoff between the convergence rate of the estimation error and the decay speed of the communication rate. This tradeoff can be illustrated from the example discussed in Remark 3.11: One can decrease ϵ_0 to reduce the decay speed of $f_i(t) = t^{-\epsilon_0}$; then a larger μ is feasible from $1 - \epsilon_0 - \mu - 2/\rho > 0$ such that a faster decay speed of the communication rate is available. However, it would lead to a smaller δ due to $\epsilon_0 > \delta \geq 0$, meaning the convergence rate of the estimation error is compromised.

4. Proofs of the main results. The proofs of the main results in the last section are provided in this section.

4.1. Convergence of linear recursion. To show the convergence of Algorithm 3.1, we first study the following linear recursion:

$$e(t+1) = e(t) + \alpha(t)(Q(t) + \Delta(t))e(t) + \alpha(t)(\varepsilon'(t) + \varepsilon''(t)), \tag{4.1}$$

where $e(t), \varepsilon'(t), \varepsilon''(t) \in \mathbb{R}^q$, $Q(t), \Delta(t) \in \mathbb{R}^{q\times q}$, and $\alpha(t) \in \mathbb{R}$ is the step size, $t \in \mathbb{N}$. Some assumptions are introduced below.

Assumption 4.1.
(i) $\{\alpha(t)\}$ satisfies that $\alpha(t) > 0$, $\alpha(t) \to 0$, $\sum_{t=1}^{\infty} \alpha(t) = \infty$, and

$$\frac{1}{\alpha(t+1)} - \frac{1}{\alpha(t)} \to \alpha_0 \geq 0. \tag{4.2}$$

(ii.a) $\{Q(t), t \in \mathbb{N}\}$ is a sequence of random matrices, and there exists a constant $\pi_1 \in \mathbb{R}^+$ such that

$$\sup_{t\in\mathbb{N}} \|Q(t)\| \leq \pi_1, \text{ a.s.} \tag{4.3}$$

In addition, there exist $h \in \mathbb{N}^+$, $\lambda \in \mathbb{R}^+$ such that for any $m \in \mathbb{N}$,

$$\lambda_{\max}\left[\sum_{t=mh}^{(m+1)h-1} \mathbb{E}\left\{Q(t) + Q^{\mathsf{T}}(t)|\mathcal{F}(mh-1)\right\}\right] \leq -\lambda, \text{ a.s.}, \tag{4.4}$$

where $\mathcal{F}(t) := \sigma(Q(s), \Delta(s), \varepsilon'(s), 0 \leq s \leq t)$, and $\mathcal{F}(-1) = \{\emptyset, \Omega\}$.
(ii.b) $\{\Delta(t), t \in \mathbb{N}\}$ is a sequence of random matrices such that for $t \in \mathbb{N}$

$$\|\Delta(t)\| \leq g_1(t), \text{ a.s.},$$

where $g_1(t) : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}^+$ is a measurable function satisfying $g_1(t) \to 0$ as $t \to \infty$ and $\sup_{t\geq 0} g_1(t) < \infty$.
(iii.a) $\{\varepsilon'(t), \mathcal{F}(t)\}$ is a martingale difference sequence, i.e., $\mathbb{E}\{\varepsilon'(t)|\mathcal{F}(t-1)\} = 0$ for all $t \in \mathbb{N}$, and $\sup_{t\in\mathbb{N}} \mathbb{E}\{\|\alpha^{\delta}(t)\varepsilon'(t)\|^2|\mathcal{F}(t-1)\} \leq c_\varepsilon$ a.s. with c_ε a positive constant and some $\delta \in [0, 1/2)$.
(iii.b) $\varepsilon''(t) \in \mathcal{F}(t)$, and $\|\varepsilon''(t)\| \leq g_2(t)$ a.s., $t \in \mathbb{N}$, where $g_2(t) : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}^+$ is a measurable function satisfying $g_2(t) \to 0$ as $t \to \infty$ and $\sup_{t\geq 0} g_2(t) < \infty$.

Remark 4.1. In the above assumption, the first three conditions of (i) are standard for step sizes, and (4.2) is a mild condition used in the characterization of the convergence rate of an algorithm [7]. The boundedness of $Q(t)$ in (4.3) is assumed for simplicity but could be extended to uniform conditional boundedness with respect to $\mathcal{F}(t-1)$.

Similarly, the dominance condition in (ii.b) could also be extended to certain conditional dominance by $g_1(t)$ with respect to $\mathcal{F}(t-1)$. Condition (4.4) is a counterpart of the persistent excitation condition for centralized algorithms [14]. (iii.a) is a common noise assumption and guarantees that $\sum_{t=1}^{\infty}\alpha(t)\varepsilon'(t)=\sum_{t=1}^{\infty}\alpha^{1-\delta}(t)(\alpha^{\delta}(t)\varepsilon'(t))<\infty$ a.s. when $\sum_{t=1}^{\infty}\alpha^{2(1-\delta)}(t)<\infty$. In addition, (iii.b) ensures $\varepsilon''(t)\to 0$. These two conditions, $\sum_{t=1}^{\infty}\alpha(t)\varepsilon'(t)<\infty$ and $\varepsilon''(t)\to 0$, are standard for almost-sure convergence of stochastic approximation algorithms [7]. Extensions of (iii.a) and (iii.b) are left to future work.

THEOREM 4.2. *Under Assumption* 4.1*, $e(t)$ in* (4.1)*, starting with any fixed initial condition, converges in mean square, i.e.,*

$$\lim_{t\to\infty}\mathbb{E}\{\|e(t)\|^2\}=0.$$

In addition, suppose that the conditions below hold:
(a) $\sum_{t=1}^{\infty}\alpha(t)^{2(1-\delta)}<\infty$*, where δ is defined in Assumption* 4.1(iii.a)*,*
(b) $\sum_{t=1}^{\infty}\alpha(t)g_2(t)<\infty$*,*
(c) *$\varepsilon'(t)$ can be written as $\varepsilon'(t)=u(t)w(t)$, where $u(t)\in\mathbb{R}^{q\times q}$ and $w(t)\in\mathbb{R}^q$ such that $\{w(t),t\in\mathbb{N}\}$ is independent of $\{Q(t),\Delta(t),u(t),t\in\mathbb{N}\}$ and $\{w(t),\mathcal{F}(t)\}$ is a martingale difference sequence.*
Then $e(t)$ starting with any fixed initial condition converges to zero a.s., i.e.,

$$\mathbb{P}\left\{\lim_{t\to\infty}e(t)=0\right\}=1.$$

Remark 4.3. As mentioned in Remark 4.1, the additional condition (a) is to ensure $\sum_{t=1}^{\infty}\alpha(t)\varepsilon'(t)<\infty$ a.s. Conditions (b) and (c) are introduced to deal with the possible dependence of measurement noise, $\epsilon'(t)$ and $\epsilon''(t)$, on $Q(t)$ and $\Delta(t)$. This does not exist in classic linear recursion, where $Q(t)$ and $\Delta(t)$ are both deterministic [7]. Possible extensions of these two conditions will be investigated in the future.

Proof sketch of Theorem 4.2. Due to space limitation, we provide a proof sketch here. Readers can refer to [20] for details.

Denote $\Phi(t,s):=(I+\alpha(t)(Q(t)+\Delta(t)))(I+\alpha(t-1)(Q(t-1)+\Delta(t-1)))\cdots(I+\alpha(s)(Q(s)+\Delta(s)))$, $t\geq s\geq 0$, and $\Phi(t,s)=I$, $0\leq t<s$.

To prove Theorem 4.2, we first show that under Assumption 4.1(i), (ii.a), and (ii.b), there exists a positive integer $\tilde{m}$ such that for $t\geq\tilde{m}h$ and $s\geq 0$,

$$\mathbb{E}\left\{\Phi^{\mathsf{T}}(t,s)\Phi(t,s)|\mathcal{F}(s-1)\right\}\leq c_2\exp\left(-c_1\sum_{k=s}^{t}\alpha(k)\right)I,\quad\text{a.s.,}\tag{4.5}$$

where c_1 and c_2 are positive constants.

To verify this conclusion, one can first bound $\Phi^{\mathsf{T}}(t,s)\Phi(t,s)$ in one period of length h, i.e., $\Phi^{\mathsf{T}}\big((m+1)h-1,mh\big)\Phi\big((m+1)h-1,mh\big)$, $m\in\mathbb{N}$, using Assumption 4.1(ii) and the following facts from Assumption 4.1(i)

$$\frac{\alpha(t+i)}{\alpha(t+j)}=1+O\big(\alpha(t+i)\big),\tag{4.6}$$

$$\alpha(t+i)-\alpha(t+j)=o\big(\alpha(t+j)\big),\tag{4.7}$$

where $0\leq i\neq j\leq k$ and $k\geq 1$ is a fixed integer. Then utilizing (4.3) and Assumption 4.1(i), one is able to generalize the exponential bound to the case where t is large enough and $0\leq s\leq t$.

From (4.1) and the definition of $\Phi(t,s)$, write

$$e(t+1)=\Phi(t,0)e(0)+\sum_{k=0}^{t}\alpha(k)\Phi(t,k+1)(\varepsilon'(k)+\varepsilon''(k)).$$

To prove the mean-square convergence, we split an upper bound of $\mathbb{E}\{\|e(t+1)\|^2\}$ into three parts:

$$\begin{aligned}&\mathbb{E}\{\|e(t+1)\|^2\}\\&\le 3\Bigg(\mathbb{E}\{\|\Phi(t,0)e(0)\|^2\}+\mathbb{E}\left\{\left\|\sum_{k=0}^{t}\alpha(k)\Phi(t,k+1)\varepsilon'(k)\right\|^2\right\}\\&\quad+\mathbb{E}\left\{\left\|\sum_{k=0}^{t}\alpha(k)\Phi(t,k+1)\varepsilon''(k)\right\|^2\right\}\Bigg)\\&:=3((I)+(II)+(III)).\end{aligned}$$

It follows from (4.5) and Assumption 4.1(i) that $(I)\to 0$ as $t\to\infty$. For (II), (4.5) and Assumption 4.1(iii.a) yield that

$$(II)\le c_2c_\varepsilon\sum_{k=0}^{t}\alpha^{2(1-\delta)}(k)\ \exp\left(-c_1\sum_{i=k+1}^{t}\alpha(i)\right),$$

where c_2 and c_ε are given in (4.5) and Assumption 4.1(iii.a), respectively. In fact, using properties of $\alpha(t)$ given in Assumption 4.1(i), one can show that the above bound tends to zero as $t\to\infty$, implying $(II)\to 0$. Similarly, $(III)\to 0$ follows from (4.5), Assumption 4.1(i), and Assumption 4.1(iii.b).

To show the almost-sure convergence of $e(t)$, we first study the convergence of sequence $\{\|e(mh)\|^2, m\in\mathbb{N}\}$. By expanding $\mathbb{E}\{\|e((m+1)h)\|^2|\mathcal{F}(mh-1)\}$ and utilizing similar arguments as above, under the additional assumption (c), a bound can be obtained as follows:

$$\begin{aligned}&\mathbb{E}\{\|e((m+1)h)\|^2|\mathcal{F}(mh-1)\}\\&\le\|e(mh)\|^2+c_3\Bigg(\sum_{k=mh}^{(m+1)h-1}\alpha^{2(1-\delta)}(k)+\sum_{k=mh}^{(m+1)h-1}\alpha^2(k)g_2^2(k)\\&\quad+(\|e(mh)\|+c_4)\sum_{k=mh}^{(m+1)h-1}\alpha(k)g_2(k)\Bigg)\end{aligned}$$

for large enough m, where c_3 and c_4 are positive constants. Since we have proved that $\mathbb{E}\{\|e(t)\|^2\}\to 0$ as $t\to\infty$, $\mathbb{E}\{\|e(mh)\|\}$ is bounded. Hence, under the additional assumptions (a)–(b), Lemma 1.2.2 in [7] ensures that $e(mh)$ converges a.s. as $m\to\infty$, implying that $e(mh)\to 0$ a.s.

Note that for $m\ge 0$ and $0<s<h$,

$$\begin{aligned}\|e(mh+s)\|\le\|\Phi(mh+s-1,mh)e(mh)\|+\left\|\sum_{k=mh}^{mh+s-1}\alpha(k)\Phi(mh+s,k+1)\varepsilon'(k)\right\|\\+\left\|\sum_{k=mh}^{mh+s-1}\alpha(k)\Phi(mh+s,k+1)\varepsilon''(k)\right\|.\end{aligned}$$

From Assumption 4.1(iii.a), the additional assumption (a), and the Borel–Cantelli lemma, one can show that $\alpha(k)\|\varepsilon'(k)\| \to 0$. Therefore, the right side of the above inequality converges to zero a.s. for fixed s as $m \to \infty$, from $e(mh) \to 0$, $\alpha(k)\|\varepsilon'(k)\| \to 0$, and Assumption 4.1(ii.a) and (iii.b). This implies $e(t) \to 0$ a.s.

4.2. Proof of Theorem 3.6. From (3.4) and $\mathcal{L}\mathbf{1}_N = \mathbf{0}_N$, the estimation error $e_0(t) = X(t) - \mathbf{1}_N \otimes \theta$ evolves as below;

$$\begin{aligned}
e_0(t+1) &= e_0(t) - \bar{\alpha}(t)(\mathcal{L} \otimes I_M + \bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t))e_0(t) \\
&\quad + \bar{\alpha}(t)\left(\bar{D}_H(t)V(t) + (\mathcal{A} \otimes I_M)(X(\tau_k) - X(t))\right) \\
&= e_0(t) - \alpha(t)\left(\frac{\bar{\alpha}(t)}{\alpha(t)}\right)(\mathcal{L} \otimes I_M + \bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t))e_0(t) \\
&\quad + \alpha(t)\left(\frac{\bar{\alpha}(t)}{\alpha(t)}\right)\left(\bar{D}_H(t)V(t) + (\mathcal{A} \otimes I_M)(X(\tau_k) - X(t))\right) \\
&= e_0(t) - \alpha(t)(I_{MN} + o(1))(\mathcal{L} \otimes I_M + \bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t))e_0(t) \\
&\quad + \alpha(t)(I_{MN} + o(1))\left(\bar{D}_H(t)V(t) + (\mathcal{A} \otimes I_M)(X(\tau_k) - X(t))\right),
\end{aligned}$$

where $o(1)$ is a deterministic infinitesimal.

Let $e(t) = e_0(t)/\alpha^{\delta}(t)$, $\delta \in [0, 1/2)$, so

$$\begin{aligned}
&e(t+1) \\
&= \left(\frac{\alpha(t)}{\alpha(t+1)}\right)^{\delta} \frac{e_0(t+1)}{\alpha^{\delta}(t)} \\
&= \left(\frac{\alpha(t)}{\alpha(t+1)}\right)^{\delta} \frac{1}{\alpha^{\delta}(t)}\Big(\big(I_{MN} \\
&\quad - \alpha(t)(I_{MN} + o(1))(\mathcal{L} \otimes I_M + \bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t))\big)e_0(t) \\
&\quad + \alpha(t)(I_{MN} + o(1))\big(\bar{D}_H(t)V(t) + (\mathcal{A} \otimes I_M)(X(\tau_k) - X(t))\big)\Big) \\
&= \left(\frac{\alpha(t)}{\alpha(t+1)}\right)^{\delta}\Bigg(\big(I_{MN} \\
&\quad - \alpha(t)(I_{MN} + o(1))(\mathcal{L} \otimes I_M + \bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t))\big)e(t) \\
&\quad + \alpha(t)(I_{MN} + o(1))\left(\frac{\bar{D}_H(t)V(t)}{\alpha^{\delta}(t)} + (\mathcal{A} \otimes I_M)\frac{X(\tau_k) - X(t)}{\alpha^{\delta}(t)}\right)\Bigg)
\end{aligned}$$

$$\begin{aligned}
&= \left(1 + \delta\frac{\alpha(t) - \alpha(t+1)}{\alpha(t+1)} + O\left(\left(\frac{\alpha(t) - \alpha(t+1)}{\alpha(t+1)}\right)^2\right)\right) \\
&\quad \Bigg(\big(I_{MN} - \alpha(t)(I_{MN} + o(1))(\mathcal{L} \otimes I_M + \bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t))\big)e(t) \\
&\quad + \alpha(t)(I_{MN} + o(1))\left(\frac{\bar{D}_H(t)V(t)}{\alpha^{\delta}(t)} + (\mathcal{A} \otimes I_M)\frac{X(\tau_k) - X(t)}{\alpha^{\delta}(t)}\right)\Bigg)
\end{aligned} \tag{4.8}$$

$$\begin{aligned}
&= \left(1 + \delta\frac{\alpha(t) - \alpha(t+1)}{\alpha(t+1)} + O\left(\left(\frac{\alpha(t) - \alpha(t+1)}{\alpha(t+1)}\right)^2\right)\right) \\
&\quad \big(I_{MN} - \alpha(t)(I_{MN} + o(1))(\mathcal{L} \otimes I_M + \bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t))\big)e(t) \\
&\quad + \big(1 + O(\alpha(t))\big)\alpha(t)(I_{MN} + o(1))
\end{aligned} \tag{4.9}$$

$$
\begin{aligned}
&\left(\frac{\bar{D}_H(t)V(t)}{\alpha^\delta(t)} + (\mathcal{A}\otimes I_M)\frac{X(\tau_k)-X(t)}{\alpha^\delta(t)}\right)\\
=&\Bigg\{I_{MN}+\alpha(t)\\
&\Bigg[-(I_{MN}+o(1))(\mathcal{L}\otimes I_M+\bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t))+\delta\frac{\alpha(t)-\alpha(t+1)}{\alpha(t)\alpha(t+1)}I_{MN}\\
&-\delta\frac{\alpha(t)-\alpha(t+1)}{\alpha(t+1)}(I_{MN}+o(1))(\mathcal{L}\otimes I_M+\bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t))\\
&+O\left(\left(\frac{\alpha(t)-\alpha(t+1)}{\alpha(t+1)}\right)^2\right)\left(\frac{1}{\alpha(t)}I_{MN}\right.\\
&\left.-(I_{MN}+o(1))(\mathcal{L}\otimes I_M+\bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t))\right)\Bigg]\Bigg\}e(t)\\
&+\alpha(t)\big(1+O(\alpha(t))\big)(I_{MN}+o(1))\\
&\left(\frac{\bar{D}_H(t)V(t)}{\alpha^\delta(t)}+(\mathcal{A}\otimes I_M)\frac{X(\tau_k)-X(t)}{\alpha^\delta(t)}\right),
\end{aligned}
$$

where (4.8) follows from Taylor's expansion and (4.6), and (4.9) is from (4.7).

Now let

$$
\begin{aligned}
Q(t)&=-(\mathcal{L}\otimes I_M+\bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t))+\alpha_0\delta I_{MN},\\
\Delta(t)&=o(1)(\mathcal{L}\otimes I_M+\bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t))+\delta\left(\frac{\alpha(t)-\alpha(t+1)}{\alpha(t)\alpha(t+1)}-\alpha_0\right)I_{MN}\\
&\quad-\delta\frac{\alpha(t)-\alpha(t+1)}{\alpha(t+1)}(I_{MN}+o(1))\big(\mathcal{L}\otimes I_M+\bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t)\big)\\
&\quad+O\left(\left(\frac{\alpha(t)-\alpha(t+1)}{\alpha(t+1)}\right)^2\right)\left(\frac{1}{\alpha(t)}I_{MN}\right.\\
&\quad\left.-(I_{MN}+o(1))(\mathcal{L}\otimes I_M+\bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t))\right),
\end{aligned}
$$

$$
\begin{aligned}
\varepsilon'(t)&=\big(1+O(\alpha(t))\big)(I_{MN}+o(1))\frac{\bar{D}_H(t)V(t)}{\alpha^\delta(t)},\\
\varepsilon''(t)&=\big(1+O(\alpha(t))\big)(I_{MN}+o(1))(\mathcal{A}\otimes I_M)\frac{X(\tau_k)-X(t)}{\alpha^\delta(t)},
\end{aligned}
$$

and we can write $e_0(t)/\alpha^\delta(t)$ in the form (4.1). Note that Assumption 3.1(i.a) is identical to Assumption 4.1(i). Since

$$
Q(t)+Q^{\mathsf{T}}(t)=-(\mathcal{L}+\mathcal{L}^{\mathsf{T}})\otimes I_M-2\bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t)+2\alpha_0\delta I_{MN},
$$

Assumption 4.1(ii.a) holds under Assumption 3.1(ii.c).

From (4.7), it holds that

$$
\begin{aligned}
\Delta(t)&=o(1)(\mathcal{L}\otimes I_M+\bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t))+\delta\left(\frac{\alpha(t)-\alpha(t+1)}{\alpha(t)\alpha(t+1)}-\alpha_0\right)I_{MN}\\
&\quad-\delta\frac{\alpha(t)-\alpha(t+1)}{\alpha(t+1)}(I_{MN}+o(1))\big(\mathcal{L}\otimes I_M+\bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t)\big)
\end{aligned}
$$

$$+ O\left(\frac{\alpha(t)-\alpha(t+1)}{\alpha(t+1)}\right) I_{MN}$$
$$- O\left(\left(\frac{\alpha(t)-\alpha(t+1)}{\alpha(t+1)}\right)^2\right)(I_{MN}+o(1))\left(\mathcal{L}\otimes I_M + \bar{D}_H(t)\bar{D}_H^{\mathsf{T}}(t)\right),$$

so by $\sup_{t\in\mathbb{N}}\|\bar{D}_H(t)\|^2 \le D$ and Assumption 3.1(i.a) we know that $\Delta(t)$ can be bounded by a deterministic function that converges to zero as $t\to\infty$.

Assumption 3.1(ii.a) and (ii.b) ensure that $\{\varepsilon'(t),\mathcal{F}(t)\}$ is a martingale difference sequence, and

$$\begin{aligned}
&\mathbb{E}\{\|\alpha^{\delta}(t)\varepsilon'(t)\|^2|\mathcal{F}(t-1)\}\\
&= \mathbb{E}\left\{\left\|\big(1+O(\alpha(t))\big)(I_{MN}+o(1))\bar{D}_H(t)V(t)\right\|^2\Big|\mathcal{F}(t-1)\right\}\\
&\le c_1\mathbb{E}\{\|\bar{D}_H(t)V(t)\|^2|\mathcal{F}(t-1)\}\\
&\le c_1 D\mathbb{E}\{\|V(t)\|^2|\mathcal{F}(t-1)\}\\
&\le c_1 D\left(\mathbb{E}\{\|V(t)\|^{\rho}|\mathcal{F}(t-1)\}\right)^{\frac{2}{\rho}} \le c(c_V)^{\frac{2}{\rho}}D,
\end{aligned}$$

where c_1 is a positive constant, and the last inequality is obtained from conditional Lyapunov inequality.

Finally, $\|\varepsilon''(t)\| \le c_2 f_{\max}(t)/\alpha^{\delta}(t) \to 0$ for some positive constant c_2, from Assumption 3.1(iii.a). Therefore, Theorem 4.2 implies the conclusion.

4.3. Proof of Theorem 3.8. Following section 4.2, we only need to validate the additional assumptions in Theorem 4.2. Note that $\sum_{t=1}^{\infty}\alpha(t)^{2(1-\delta)}<\infty$ is given in Assumption 3.1(i.b), and $\sum_{t=1}^{\infty}\alpha(t)g_2(t)<\infty$ holds from Assumption 3.1(iii.b), by noticing $g_2(t) := c_2 f_{\max}(t)/\alpha^{\delta}(t)$. Finally, letting $u(t) := \big(1+O(\alpha(t))\big)(I_{MN} + o(1))\bar{D}_H(t)/\alpha^{\delta}(t)$ and $w(t) := V(t)$, we know that (c) holds under Assumption 3.1(ii.a) and (ii.b).

4.4. Proof of Theorem 3.12. Recall that $\tau_k^i := \tau_{k_i(t)}$ is the kth triggering instant of sensor $i\in\mathcal{V}$ in the time interval $[0,t]\cap\mathbb{N}$. Denote the set $\Gamma = \{i\in\mathcal{V}|\tau_k^i\to\infty \text{ as } t\to\infty\}$, which is the set of sensors whose number of communications goes to infinity as time goes to infinity. If $\Gamma=\emptyset$, then there are positive integers n_0 and N_0 such that for $t\ge n_0$, $\max_{i\in\mathcal{V}}\sup_t \tau_k^i \le N_0$ surely. From the definition of communication rate (2.2), for $t\ge n_0$ and $\gamma\in[0,\frac{2\mu}{2\mu+1})$ for all $\mu\in[1/2,1)$, it holds that

$$\lim_{t\to\infty}\lambda_c(t)t^{\gamma} \le \lim_{t\to\infty}\frac{N_0}{t^{1-\gamma}} = 0.$$

In the case, the conclusion holds.

Next, we consider the case that $\Gamma\ne\emptyset$. According to (3.2), for any sensor $i\in\Gamma$ and time $t\ge\tau_k^i$, we have

$$\begin{aligned}
&x_i(t+1)-x_i(\tau_k^i)\\
&=x_i(t)-x_i(\tau_k^i)+\alpha_i(t)\sum_{j\in\mathcal{N}_i}a_{i,j}(x_j(t)-x_i(t))\\
&\quad+\alpha_i(t)H_i^{\mathsf{T}}(t)(y_i(t)-H_i(t)x_i(t))+\alpha_i(t)\sum_{j\in\mathcal{N}_i}a_{i,j}(x_j(\tau_k^j)-x_j(t)).
\end{aligned}\tag{4.10}$$

From Assumption 3.1(i.a), there is a constant $c_0>0$ such that $\alpha_i(t)\le c_0\alpha(t)$. Taking norm operator on both sides of (4.10) yields

$$\begin{aligned}
&\left\|x_i(t+1)-x_i(\tau_k^i)\right\| \\
&\le \left\|x_i(t)-x_i(\tau_k^i)\right\| + c_0\alpha(t)\left\|\sum_{j\in\mathcal{N}_i} a_{i,j}(x_j(t)-x_i(t))\right\| \\
&\quad + c_0\alpha(t)\left\|H_i^{\mathsf{T}}(t)(y_i(t)-H_i(t)x_i(t))\right\| + c_0\alpha(t)\left\|\sum_{j\in\mathcal{N}_i} a_{i,j}(x_j(\tau_k^j)-x_j(t))\right\| \\
&:= \left\|x_i(t)-x_i(\tau_k^i)\right\| + (I) + (II) + (III). \qquad (4.11)
\end{aligned}$$

Consider (I) in (4.11). Denote $d_0 = \max_{i\in\mathcal{V}} \sum_{j\in\mathcal{N}_i} a_{i,j}$; then we have

$$\begin{aligned}
c_0\alpha(t)\left\|\sum_{j\in\mathcal{N}_i} a_{i,j}(x_j(t)-x_i(t))\right\| &= c_0\alpha(t)\left\|\sum_{j\in\mathcal{N}_i} a_{i,j}(x_j(t)-\theta+\theta-x_i(t))\right\| \\
&\le 2d_0c_0\alpha(t)\max_{j\in\mathcal{V}}\|x_j(t)-\theta\| \le \bar{c}_1\alpha(t)^{1+\delta}, \qquad (4.12)
\end{aligned}$$

where $\bar{c}_1$ is a positive scalar, which could be different in different sample trajectories, and the last inequality is obtained from Theorem 3.8.

Consider (II) in (4.11). For all $\varepsilon_0 > 0$, by (ii.b) of Assumption 3.1 and the Markov inequality, we obtain

$$\begin{aligned}
\sum_{t=0}^{\infty}\mathbb{P}\{\alpha(t)^{\frac{2(1-\delta)}{\rho}}\|V(t)\| \ge \varepsilon_0\} &\le \frac{1}{\varepsilon_0^{\rho}}\sum_{t=0}^{\infty}\alpha(t)^{2(1-\delta)}\mathbb{E}\{\|V(t)\|^{\rho}\} \\
&\le \frac{\bar{c}_V}{\varepsilon_0^{\rho}}\sum_{t=0}^{\infty}\alpha(t)^{2(1-\delta)} < \infty. \qquad (4.13)
\end{aligned}$$

Hence by the Borel–Cantelli lemma, we have

$$\lim_{t\to\infty}\alpha(t)^{\frac{2(1-\delta)}{\rho}}\|v_i(t)\| = 0, \text{ a.s.} \qquad (4.14)$$

Under (ii.c) of Assumption 3.1, $\sup_{t\ge 0}\left\|H_i^{\mathsf{T}}(t)\right\| < \infty$; then from (4.14) and Theorem 3.8, there are positive scalar $\bar{c}_2, \bar{c}_3$, which could be different in different sample trajectories, such that

$$\begin{aligned}
&c_0\alpha(t)\left\|H_i^{\mathsf{T}}(t)(y_i(t)-H_i(t)x_i(t))\right\| \\
=&c_0\alpha(t)\left\|H_i^{\mathsf{T}}(t)H_i(t)(\theta-x_i(t)) + H_i^{\mathsf{T}}(t)v_i(t)\right\| \\
\le&c_0\alpha(t)\max_{i\in\mathcal{V}}\sup_{t\ge 0}\left\|H_i^{\mathsf{T}}(t)H_i(t)\right\|\max_{i\in\mathcal{V}}\|x_i(t)-\theta\| + \bar{c}_2\alpha(t)^{1-\frac{2(1-\delta)}{\rho}} \\
\le&\bar{c}_3\alpha(t)^{1+\delta} + \bar{c}_2\alpha(t)^{1-\frac{2(1-\delta)}{\rho}}. \qquad (4.15)
\end{aligned}$$

Consider (III) in (4.11). Since $\left\|x_j(\tau_k^j)-x_j(t)\right\| \le f_j(t)$,

$$c_0\alpha(t)\left\|\sum_{j\in\mathcal{N}_i} a_{i,j}(x_j(\tau_k^j)-x_j(t))\right\| \le c_0d_0\alpha(t)f_{\max}(t). \qquad (4.16)$$

From inequalities (4.11), (4.12), (4.15), and (4.16), it follows that

$$\begin{aligned}
\left\|x_i(t+1)-x_i(\tau_k^i)\right\| \le& \left\|x_i(t)-x_i(\tau_k^i)\right\| + (\bar{c}_1+\bar{c}_3)\alpha(t)^{1+\delta} \\
&+ \bar{c}_2\alpha(t)^{1-\frac{2(1-\delta)}{\rho}} + c_0d_0\alpha(t)f_{\max}(t).
\end{aligned}$$

Under Assumption 3.2, there is a monotonically nonincreasing sequence $\beta(t) = O(\alpha(t)^{1-2(1-\delta)/\rho})$. Thus, there is a scalar $\bar{c}_4 > 0$, which could be different in different sample trajectories, such that

$$\left\|x_i(t+1) - x_i(\tau_k^i)\right\| \leq \left\|x_i(t) - x_i(\tau_k^i)\right\| + \bar{c}_4\beta(t).$$

Denote $L_k^i := \tau_{k+1}^i - \tau_k^i$ the interval length between the $(k+1)$th triggering time and the kth triggering time; then we have

$$\begin{aligned}\left\|x_i(\tau_{k+1}^i) - x_i(\tau_k^i)\right\| &= \left\|x_i(\tau_k^i + L_k^i) - x_i(\tau_k^i)\right\| \\ &\leq \left\|x_i(\tau_k^i + L_k^i - 1) - x_i(\tau_k^i)\right\| + \bar{c}_4\beta(\tau_k^i + L_k^i - 1) \\ &\quad \vdots \\ &\leq \bar{c}_4 \sum_{s=\tau_k^i}^{\tau_k^i + L_k^i - 1} \beta(s) \leq \bar{c}_4 L_k^i \beta(\tau_k^i),\end{aligned}$$

where the last inequality is obtained from the monotonicity of $\beta(t)$.

A necessary condition to guarantee that the event is triggered for sensor i is

$$\bar{c}_4 L_k^i \beta(\tau_k^i) > f_i(\tau_{k+1}^i) \Longleftrightarrow L_k^i > \frac{f_i(\tau_k^i + L_k^i)}{\bar{c}_4\beta(\tau_k^i)}. \tag{4.17}$$

Then we make the claim taht

$$\liminf_{k\to\infty} \frac{L_k^i}{(\tau_k^i)^\mu} > 0, \tag{4.18}$$

where $\mu \in [1/2, 1)$ is introduced in Assumption 3.2. The proof of claim (4.18) is given by contradiction. Suppose claim (4.18) does not hold, i.e., $\liminf_{k\to\infty} \frac{L_k^i}{(\tau_k^i)^\mu} = 0$. Then $\{k\}_{k=0}^\infty$ has a subsequence $\{k_j\}_{j=0}^\infty$ such that

$$\lim_{j\to\infty} \frac{L_{k_j}^i}{(\tau_{k_j}^i)^\mu} = 0, \tag{4.19}$$

which means there is a finite integer $J > 0$ such that $L_{k_j}^i \leq (\tau_{k_j}^i)^\mu$ for any $j \geq J$. It follows from (4.17) and the monotonicity of $\bar{f}(t)$ that

$$L_{k_j}^i > \frac{\bar{f}(\tau_{k_j}^i + L_{k_j}^i)}{\bar{c}_4\beta(\tau_{k_j}^i)} = \frac{\bar{f}(\tau_{k_j}^i + L_{k_j}^i)}{\bar{c}_4\bar{f}(\tau_{k_j}^i)}\frac{\bar{f}(\tau_{k_j}^i)}{\beta(\tau_{k_j}^i)} \geq \frac{\bar{f}(\tau_{k_j}^i + (\tau_{k_j}^i)^\mu)}{\bar{c}_4\bar{f}(\tau_{k_j}^i)}\frac{\bar{f}(\tau_{k_j}^i)}{\beta(\tau_{k_j}^i)} \geq \frac{\bar{c}_5}{\bar{c}_4}\frac{\bar{f}(\tau_{k_j}^i)}{\beta(\tau_{k_j}^i)},$$

where $\bar{c}_5 > 0$ exists due to Assumption 3.2(ii). From Assumption 3.2(iv), there is $\bar{c}_6 > 0$ such that $\frac{\bar{f}(t)}{\beta(t)} > \bar{c}_6 t^\mu$ for any $t \in \mathbb{N}$. Then $L_{k_j}^i > \frac{\bar{c}_5\bar{c}_6}{\bar{c}_4}(\tau_{k_j}^i)^\mu$ for $j \geq J$, which contradicts (4.19). Thus, claim (4.18) holds.

According to (4.18), there is $\bar{M} > 0$ such that $L_k^i > \bar{M}(\tau_k^i)^\mu =: g(\tau_k^i)$ for any $k \in \mathbb{N}$. It follows that

$$g(\tau_{k+1}^i) - g(\tau_k^i) = \bar{M}(\tau_k^i + L_k^i)^\mu - \bar{M}(\tau_k^i)^\mu \geq \bar{M}(\tau_k^i + \bar{M}(\tau_k^i)^\mu)^\mu - \bar{M}(\tau_k^i)^\mu.$$

According to Newton's generalized binomial theorem and $\mu \in [1/2, 1)$, there are $T_1 > 0$ and $\tilde{M} > 0$ such that for $\tau_k^i \geq T_1$, we have $\bar{M}(\tau_k^i + \bar{M}(\tau_k^i)^\mu)^\mu - \bar{M}(\tau_k^i)^\mu \geq \tilde{M}(\tau_k^i)^{2\mu-1}$. Therefore, for $\tau_k^i \geq T_1$, it holds that

$$L_k^i > g(\tau_k^i), \quad g(\tau_{k+1}^i) - g(\tau_k^i) \geq \tilde{M}(\tau_k^i)^{2\mu-1}. \tag{4.20}$$

Next, we prove the decay speed of the communication rate. For any sensor $i \in \Gamma$, it follows from the definition that $\tau_k^i \to \infty$ as $t \to \infty$. Then there is an integer $s > 0$ such that $\tau_s^i \geq T_1$. Let $t > \tau_s^i$; then we split the interval $[0, t] \cap \mathbb{N}$ into two subintervals, namely, $[0, \tau_s^i] \cap \mathbb{N}$ and $(\tau_s^i, t] \cap \mathbb{N}$. Denote $\bar{s}$ the triggering times of sensor i in $(\tau_s^i, t] \cap \mathbb{N}$. Given the above finite s, for any $\gamma \in [0, \frac{2\mu}{2\mu+1})$, it is straightforward to see that

$$\frac{s}{t^{1-\gamma}} \to 0, \quad t \to \infty. \tag{4.21}$$

Recall that τ_k^i is the kth triggering instant of sensor $i \in \mathcal{V}$ in the time interval $[0, t] \cap \mathbb{N}$; thus we have $\tau_k^i \geq k$. Then for $\tau_l^i \geq \tau_s^i \geq T_1$, it follows from (4.20) that

$$\begin{aligned} g(\tau_l^i) &= g(\tau_s^i) + \sum_{k=s}^{l-1} (g(\tau_{k+1}^i) - g(\tau_k^i)) \\ &\geq g(\tau_s^i) + \tilde{M} \sum_{k=s}^{l-1} (\tau_k^i)^{2\mu-1} \\ &\geq g(\tau_s^i) + \tilde{M} \sum_{k=1}^{l-1} k^{2\mu-1} - \tilde{M} \sum_{k=1}^{s-1} k^{2\mu-1} \\ &\geq g(\tau_s^i) + \tilde{M}_2 l^{2\mu} - \tilde{M} \sum_{k=1}^{s-1} k^{2\mu-1}, \\ &\geq \tilde{M}_3 l^{2\mu}, \end{aligned} \tag{4.22}$$

where $\tilde{M}_2, \tilde{M}_3 > 0$, and the last inequality is obtained by using the sum formula in [13, page 1].

Next, we consider $\bar{s}/t^{1-\gamma}$. According to (4.20), (4.22), and $\tau_l^i \geq \tau_s^i \geq T_1$ for $l \geq s$, it follows that

$$t \geq t - \tau_s^i \geq \sum_{l=s}^{s+\bar{s}-1} L_l^i \geq \sum_{l=s}^{s+\bar{s}-1} g(\tau_l^i) \geq \tilde{M}_3 \sum_{l=s}^{s+\bar{s}-1} l^{2\mu} \geq \tilde{M}_4 (s + \bar{s} - 1)^{2\mu+1},$$

where $\tilde{M}_4 > 0$ and the last inequality is obtained by using the sum formula in [13, page 1].

Due to $\gamma \in [0, \frac{2\mu}{2\mu+1})$, we use l'Hôspital's rule to obtain

$$\frac{\bar{s}}{t^{1-\gamma}} \leq \frac{\bar{s}}{\left(\tilde{M}_4 (s + \bar{s} - 1)^{2\mu+1}\right)^{1-\gamma}} \stackrel{\bar{s} \to \infty}{\longrightarrow} 0. \tag{4.23}$$

Due to $K_i(t) = s + \bar{s}$, it follows that for any $i \in \Gamma$,

$$\frac{K_i(t)}{t^{1-\gamma}} \stackrel{t \to \infty}{\longrightarrow} 0. \tag{4.24}$$

According to the definition of communication rate in Definition 2.1, it follows that for any $\gamma \in [0, \frac{2\mu}{2\mu+1})$,

$$\lambda_c(t)t^\gamma \le \sum_{j\in\mathcal{V}} |\mathcal{N}_j^c| \frac{K_j(t)}{t^{1-\gamma}} = \sum_{j\in\mathcal{V}\setminus\Gamma} |\mathcal{N}_j^c| \frac{K_j(t)}{t^{1-\gamma}} + \sum_{j\in\Gamma} |\mathcal{N}_j^c| \frac{K_j(t)}{t^{1-\gamma}} \to 0, \quad t\to\infty, \tag{4.25}$$

where the first item goes to zero since $K_j(t)$ is upper bounded for $j \in \mathcal{V}\setminus\Gamma$, and the second item goes to zero according to (4.24).

5. Numerical simulations. In this section, we provide two examples to illustrate the effectiveness of Algorithm 3.1 and the developed theoretical results.

5.1. Example 1. In this example, we consider the sensor network in Figure 1 with $N = 7$ sensors. Suppose the parameter vector to be estimated is $\theta = [\theta_1, \theta_2]^\mathsf{T}$, where $\theta_1 = -1$ and $\theta_2 = 2$. The sensor measurement matrices and the initial estimates are in the following:

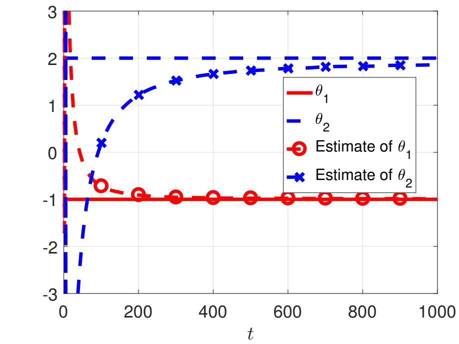
$$H_1 = [1,0]^\mathsf{T}, \quad H_2 = [0,1]^\mathsf{T}, \quad H_7 = H_5 = H_3 = H_1, \quad H_6 = H_4 = H_2,$$
$$x_1(0) = [0,-100]^\mathsf{T}, \quad x_7(0) = x_5(0) = x_3(0) = x_1(0),$$
$$x_2(0) = [100,0]^\mathsf{T}, \quad x_6(0) = x_4(0) = x_2(0).$$

Suppose the time interval is from $t = 0$ to $t = 1000$. The noise of each sensor follows a Gaussian process with mean zero and standard deviation 0.1. The noise processes are independent in time and space.

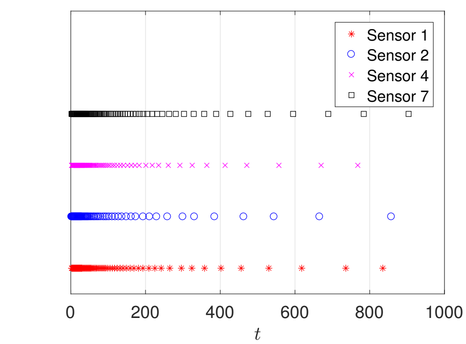
Under the above setting, we conduct a Monte Carlo experiment with $M_0 = 100$ runs for Algorithm 3.1 with $\alpha_i(t) = t^{-0.7}$ and $f_i(t) = t^{-0.5}$ for $i = 1, 2, \ldots, 7$. To evaluate the mean-square error (MSE), we define

$$\mathrm{MSE}(t) = \frac{1}{NM_0} \sum_{j=1}^{M_0} \sum_{i=1}^{N} \left\| x_i^j(t) - \theta \right\|^2, \tag{5.1}$$

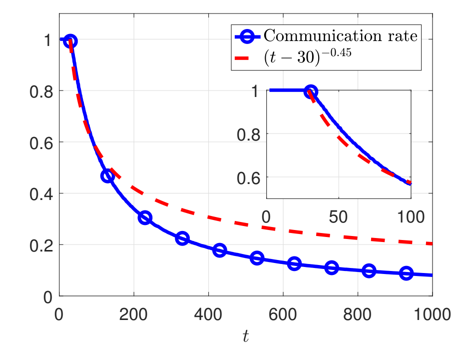
where $x_i^j(t)$ is the estimate of θ by sensor i at time t in the jth run. The simulation results are provided in Figure 2. From Figure 2(a), the average estimate of all sensors is asymptotically convergent to the true parameter vector. The event-triggered communication triggering instants of sensors 1, 2, 4, and 7 are provided in Figure 2(b), where we can see less and less communication occur as time goes on. The communication rate is 0.08 in the interval $t = [0, 1000] \cap \mathbb{N}$. The dynamics of the communication rate in the given interval is provided in Figure 2(c), where the communication rate remains 1 from $t = 1$ to $t = 30$, meaning that sensors persistently communicate with each other. This is because at the initial time, much informative data can be used to update the sensor estimates. As time goes on, the communication rate is tending to zero, since sensors only transmit informative data which is becoming less. Moreover, in order to illustrate the convergence rate of the communication rate, we provide the dynamics of $(t-30)^{-0.45}$ for $t > 30$. From Figure 2(c), the communication rate asymptotically decays to zero faster than $(t-30)^{-0.45}$, corresponding to the results in Theorem 3.12. The mean-square convergence of the algorithm is illustrated in Figure 2(d), corresponding to Theorem 3.6. Moreover, by choosing three triggering thresholds, i.e., $f_i(t) = t^{-0.8}, t^{-0.6}, t^{-0.4}$, we provide Figure 3 for illustrating the influence of triggering threshold to communication rate and MSE. From this figure, for the threshold with a faster decreasing speed, the corresponding communication rate decays more slowly, while the MSE decays more quickly. This indicates that the threshold leads to a tradeoff between communication rate and MSE, corresponding to Remark 3.13.



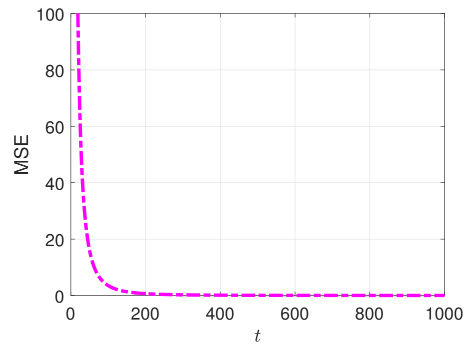
(a) Element-wise asymptotic convergence.



(b) Communication triggering instants of each sensor.

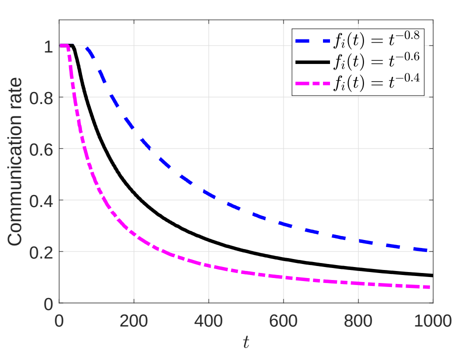


(c) Dynamics of communication rate $\lambda_c(t)$.

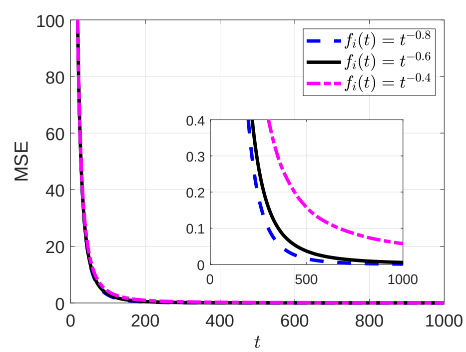


(d) Mean-square convergence.

FIG. 2. *Simulation results of Algorithm* 3.1.



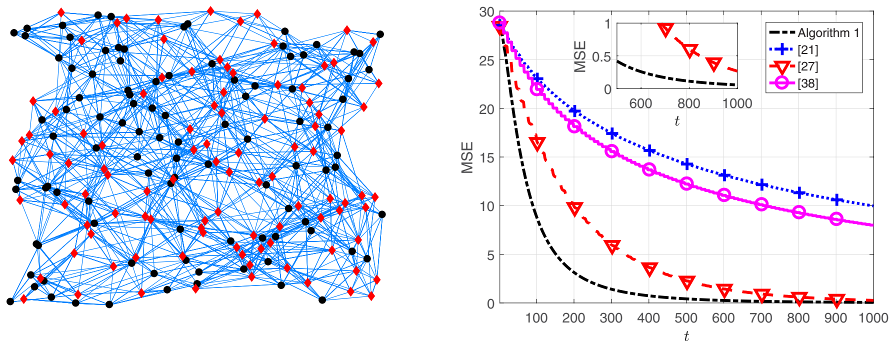
(a) Dynamics of communication rate $\lambda_c(t)$.



(b) Mean-square convergence.

FIG. 3. *Communication rate and MSE of Algorithm* 3.1 *under three triggering thresholds.*

5.2. Example 2. In this example, we compare the proposed algorithm with three existing algorithms over a sensor network whose size is larger than that in Example 1. Consider an undirected connected sensor network with 200 nodes for estimating a target position $\theta = [1, 2, 5]$. The sensor network topology is generated as a random geometric graph and provided in Figure 4(a), where two types of sensors are deployed with the same number (i.e., 100) and denoted by black circles and red



(a) Sensor network.

(b) MSE performance of four algorithms.

Fig. 4. *Comparison of four distributed algorithms over a sensor network with* 200 *nodes.*

diamonds. Assume for any $i = 1, 2, \ldots, 200$ the weight $a_{i,j} = 1$ for $j \in \mathcal{N}_i$. The measurement matrices of black-circle and red-diamond sensors are assumed to be $H_A = [0, 0, 1]$ and $H_B = [1, 0, 0; 0, 1, 0]$, respectively. Suppose the sensor noise follows a standard Gaussian process and is independent in both space and time.

We compare Algorithm 3.1 with three typical distributed estimation algorithms in the literature: the generalized linear unconstrained algorithm from [21], the diffusion least-mean squares algorithm from [27], and the distributed parameter algorithm from [38]. The parameters in our algorithm are $\alpha_i(t) = (t + 100)^{-0.7}$ and $f_i(t) = t^{-0.5}$ for $i = 1, 2, \ldots, 200$. The parameters in the algorithm of [21] are $\alpha(t) = 10/(t + 1)^{0.7}$, $\beta(t) = 0.1/(t+1)^{0.7}$, and $K = (\sum_{i=1}^{200} H_i^T H_i)^{-1}$, where H_i is the measurement matrix of sensor i. The parameters in the algorithm of [38] are $b(t) = (t+100)^{-0.7}$ and $a_{i,j} = 1$ for $i, j = 1, 2, \ldots, 200$. The parameters in the algorithm of [27] are $\mu(t) = (t+100)^{-0.7}$ and $c_{i,j} = 1/|\mathcal{N}_i + 1|$ for $j \in \mathcal{N}_i \bigcup \{i\}$, $i = 1, 2, \ldots, 200$. The initial parameter estimate is zero for each algorithm. We compare our event-triggered algorithm with the above three time-triggered algorithms under the same communication rate $\lambda_c = 0.09$. It means that for the three time-triggered algorithms, each sensor receives the messages from neighbors for every $11 \approx 1/\lambda_c$ steps; before that they use the latest messages from neighbors to run the algorithms. Under this setting, we conduct a Monte Carlo experiment for running the four algorithms simultaneously with 100 runs. With the MSE notation in (5.1), the performance of these algorithms are illustrated in Figure 4(b). It shows that the outputs of all algorithms are convergent to the true parameter vector, and our event-triggered distributed algorithm outweighs the other three algorithms in convergence speed under the same communication rate constraint.

6. Conclusion. In this paper, a distributed parameter estimation problem over a sensor network with event-triggered communications was studied. First, a fully distributed estimation algorithm was proposed based on an event-triggered communication scheme which determines when a sensor should share the parameter estimates with neighboring sensors. Then, under mild conditions, some main estimation properties of the algorithm including mean-square and almost-sure convergence were analyzed. The convergence rates were also estimated. Under some extra conditions, it was proved that the communication rate of the whole network using the proposed algorithm decays to zero almost a.s. time goes to infinity, which indicates that a tremendous amount of redundant communications are avoided. It was also shown that adjusting the decay speed of the triggering threshold can lead to a tradeoff between the convergence rate of the estimation error and the decay speed of the communication rate. Future work can be done by considering more general models of systems and networks, such as nonlinear measurement models and time-varying, unbalanced, or stochastic graphs.

REFERENCES

[1] G. Battistelli, L. Chisci, and D. Selvi, *A distributed Kalman filter with event-triggered communication and guaranteed stability*, Automatica J. IFAC, 93 (2018), pp. 75–82.

[2] X. Cao and T. Basar, *Decentralized online convex optimization with event-triggered communications*, IEEE Trans. Signal Process., 69 (2021), pp. 284–299.

[3] R. M. Castro and E. Tánczos, *Adaptive sensing for estimation of structured sparse signals*, IEEE Trans. Inform. Theory, 61 (2015), pp. 2060–2080.

[4] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, *Diffusion recursive least-squares for distributed estimation over adaptive networks*, IEEE Trans. Signal Process., 56 (2008), pp. 1865–1877.

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

[6] F. S. Cattivelli and A. H. Sayed, *Distributed nonlinear Kalman filtering with applications to wireless localization*, in Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, 2010, pp. 3522–3525.

[7] H.-F. Chen, *Stochastic Approximation and Its Applications*, Kluwer, Boston, MA, 2002.

[8] W. Chen and W. Ren, *Event-triggered zero-gradient-sum distributed consensus optimization over directed networks*, Automatica J. IFAC, 65 (2016), pp. 90–97.

[9] Y. Chen, S. Kar, and J. M. Moura, *Resilient distributed field estimation*, SIAM J. Control Optim., 58 (2020), pp. 1429–1456.

[10] M. A. Davenport, A. K. Massimino, D. Needell, and T. Woolf, *Constrained adaptive sensing*, IEEE Trans. Signal Process., 64 (2016), pp. 5437–5449.

[11] J.-D. Diao, J. Guo, and C.-Y. Sun, *Event-triggered identification of FIR systems with binary-valued output observations*, Automatica J. IFAC, 98 (2018), pp. 95–102.

[12] X. Ge, Q.-L. Han, X.-M. Zhang, L. Ding, and F. Yang, *Distributed event-triggered estimation over sensor networks: A survey*, IEEE Trans. Cybernet, 50 (2019), pp. 1306–1320.

[13] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, New York, 2014.

[14] L. Guo, *Estimating time-varying parameters by the Kalman filter based algorithm: Stability and convergence*, IEEE Trans. Automat. Control, 35 (1990), pp. 141–147.

[15] L. Guo, *Stability of recursive stochastic tracking algorithms*, SIAM J. Control Optim., 32 (1994), pp. 1195–1225.

[16] D. Han, K. You, L. Xie, J. Wu, and L. Shi, *Optimal parameter estimation under controlled communication over sensor networks.*, IEEE Trans. Signal Process., 63 (2015), pp. 6473–6485.

[17] J. Haupt, R. M. Castro, and R. Nowak, *Distilled sensing: Adaptive sampling for sparse detection and estimation*, IEEE Trans. Inform. Theory, 57 (2011), pp. 6222–6235.

[18] X. He, C. Hu, Y. Hong, L. Shi, and H.-T. Fang, *Distributed Kalman filters with state equality constraints: Time-based and event-triggered communications*, IEEE Trans. Automat. Control, 65 (2020), pp. 28–43.

[19] X. He, Q. Liu, J. Wu, and K. H. Johansson, *Distributed parameter estimation under event-triggered communications*, in Proceeding of the Chinese Control Conference, IEEE, 2019, pp. 5888–5893.

[20] X. He, Y. Xing, J. Wu, and K. H. Johansson, *Event-Triggered Distributed Estimation with Decaying Communication Rate*, Preprint, arxiv:2013.06035 [eess. SY], 2021 https://arxiv.org/abs/2103.06035.

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

[22] S. Kar and J. M. Moura, *Gossip and distributed Kalman filtering: Weak consensus under weak detectability*, IEEE Trans. Signal Process., 59 (2011), pp. 1766–1784.

[23] 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.

[24] 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.

[25] J. Lei and H.-F. Chen, *Distributed estimation for parameter in heterogeneous linear time-varying models with observations at network sensors*, Commun. Inf. Syst., 15 (2015), pp. 423–451.
[26] H. Li, S. Liu, Y. C. Soh, and L. Xie, *Event-triggered communication and data rate constraint for distributed optimization of multiagent systems*, IEEE Trans. Systems Man Cybernet Systems, 48 (2017), pp. 1908–1919.
[27] C. G. Lopes and A. H. Sayed, *Diffusion least-mean squares over adaptive networks: Formulation and performance analysis*, IEEE Trans. Signal Process., 56 (2008), pp. 3122–3136.
[28] M. Mehran and E. Magnus, *Graph Theoretic Methods in Multiagent Networks*, Princeton University Press, Princeton, NJ, 2010.
[29] M. Muehlebach and S. Trimpe, *Distributed event-based state estimation for networked systems: An Lmi approach*, IEEE Trans. Automat. Control, 63 (2017), pp. 269–276.
[30] R. Olfati-Saber and R. M. Murray, *Consensus problems in networks of agents with switching topology and time-delays*, IEEE Trans. Automat. Control, 49 (2004), pp. 1520–1533.
[31] A. S. Poznyak, *Advanced Mathematical Tools for Automatic Control Engineers: Stochastic Techniques*, Elsevier, New York, 2009.
[32] M. Rabi, G. V. Moustakides, and J. S. Baras, *Adaptive sampling for linear state estimation*, SIAM J. Control Optim., 50 (2012), pp. 672–702.
[33] K. R. Rad and A. Tahbaz-Salehi, *Distributed parameter estimation in networks*, in Proceedings of the IEEE Conference on Decision and Control, 2010, pp. 5050–5055.
[34] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, *Kalman filtering with intermittent observations*, IEEE Trans. Automat. Control, 49 (2004), pp. 1453–1464.
[35] 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.
[36] J. Wu, Q.-S. Jia, K. H. Johansson, and L. Shi, *Event-based sensor data scheduling: Trade-off between communication rate and estimation quality*, IEEE Trans. Automat. Control, 58 (2012), pp. 1041–1046.
[37] K. You, L. Xie, and S. Song, *Asymptotically optimal parameter estimation with scheduled measurements*, IEEE Trans. Signal Process., 61 (2013), pp. 3521–3531.
[38] 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.