

Learning-Based Distributed Detection-Estimation in Sensor Networks with Unknown Sensor Defects

Qing Zhou, Di Li, *Student Member, IEEE*, Soumya Kar, *Member, IEEE*, Lauren Huie,
H. Vincent Poor, *Fellow, IEEE*, Shuguang Cui, *Fellow, IEEE*

Abstract—We consider the problem of distributed estimation of an unknown deterministic scalar parameter (the target signal) in a wireless sensor network (WSN), where each sensor receives a single snapshot of the field. We assume that the observation at each node randomly falls into one of two modes: a valid or an invalid observation mode. Specifically, mode one corresponds to the desired signal plus noise observation mode (*valid*), and mode two corresponds to the pure noise mode (*invalid*) due to node defect or damage. With no prior information on such local sensing modes, we introduce a learning-based distributed procedure, called the mixed detection-estimation (MDE) algorithm, based on iterative closed-loop interactions between mode learning (detection) and target estimation. The online learning step re-assesses the validity of the local observations at each iteration, thus refining the ongoing estimation update process. The convergence of the MDE algorithm is established analytically. Asymptotic analysis shows that, in the high signal-to-noise ratio (SNR) regime, the MDE estimation error converges to that of an ideal (centralized) estimator with perfect information about the node sensing modes. This is in contrast to the estimation performance of a naive average consensus based distributed estimator (without mode learning), whose estimation error blows up with an increasing SNR.

Index Terms—Distributed estimation, robust inference, distributed learning, sensor networks, order statistics.

I. INTRODUCTION

A Key issue in wireless sensor network (WSN) design is to attain a meaningful network-wide consensus on knowledge based on unreliable locally sensed data [2]–[5]. Due to the limited sensing capability and other unpredictable physical factors, such local observations may be invalid. For each single sensor, without jointly analyzing its observation with the other nodes, the validity of the data is not detectable. The traditional solution is to fuse data at a special powerful node named the fusion center. By collecting the data from all of the sensors, the fusion center could make a jointly optimal decision. If such a centralized solution is not possible,

the distributed sensing problem arises [1], [6]–[13], where each sensor exchanges its local data with the neighbors, and merges the new information to its local estimate, in order to achieve the estimation accuracy of a centralized counterpart [14]–[16]. The existing research literature on relevant network-based distributed estimation may be broadly categorized into three classes. The first intensively studied family of distributed sensing problems consists of the so-called distributed network consensus or agreement problems and its variants [8], [9], [17]–[19], of which a popular type is the distributed averaging problem, where a group of agents want to compute a liner function of a set of values distributed across the agent network, in particular, the average of their observations [20], [21]. The second well-studied family of distributed sensing problems consists of distributed/decentralized estimation of parameters/processes in collaborative multi-agent networks with a single snapshot of the field, i.e., each agent obtains a single real or vector valued observation of the field at the beginning and no new observations are sampled over time. For example, in [22], [23] the authors studied estimation in static networks, where the sensors take a single snapshot of the field and then initiate distributed optimization to fuse the local estimates. The third well-studied family of distributed sensing problems consists of general time-sequential distributed estimation procedures for parameter inference in multi-agent networks in which agents access time-series observation data sequentially over time. In this family, two main approaches were proposed: the so-called consensus+innovation approach [24]–[26] and the diffusion approach [27]–[29]. We also mention the important and relevant literature on distributed detection and classification in multi-agent networks such as those based on the running consensus approach [30], [31], the diffusion approach [32]–[34], the consensus+innovation approach [35]–[37], and also [38]–[41].

In this paper, different from prior distributed approaches which focus solely on estimation or detection, we propose a mixed distributed detection-estimation algorithm with online interactions between detection and estimation. We assume that the observation process at each node randomly falls into one of the two modes, i.e., a valid observation mode vs. an invalid observation mode, where the valid observation is the desired signal plus noise and the invalid observation is just the pure noise. The rationale behind this stochastic observation model is that the sensor might be damaged during deployment or physically blocked by certain objects between the sensor and the target; but the communication part in the sensor node still works. In this case, the sensor cannot observe a valid

Q. Zhou is with Qualcomm Inc., San Diego, CA 92121 USA (e-mail:zhouqing02@gmail.com).

D. Li and S. Cui are with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: dili@tamu.edu; cui@tamu.edu). S. Cui is also a Distinguished Adjunct Professor at King Abdulaziz University in Saudi Arabia.

S. Kar is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: soumyak@andrew.cmu.edu).

L. Huie is with the Air Force Research Lab, Rome, NY 13441 USA (e-mail:Lauren.Huie@rl.af.mil).

H. V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Part of the work has been presented in GLOBECOM'11 and ICASSP'12 [1], [2].

observation but pure noise, and the sensor itself cannot detect the validity of the observation on its own and keep executing the standard procedure in the network as a normal node. With the above setup, the traditional distributed consensus algorithms [42]–[44] could reach a naive averaging estimate of the target signal, by locally averaging the neighbor observations. However, the stochastic property of the observation modes may cause unreliable performance as shown later in the paper.

To address the above issue, a mixed detection-estimation (MDE) algorithm is introduced in this paper, which is a learning-based distributed procedure with closed-loop iterative interactions between the distributed mode learning and target estimation. In the MDE algorithm, the mode learning part detects the validity of the local observation iteratively as it performs the distributed estimation task. In each round of iteration, each node locally detects the observation validity with the maximum a posteriori probability (MAP) criterion based on the knowledge of the local current estimate of the target together with the local observation. The local estimate is then refined with the detected validities of the local observations and other exchanged information from the neighbors using a consensus + innovations type mechanism. By alternatively detecting validity and estimating the target, the sensor network can achieve a global consensus among all nodes. We analytically establish the convergence of the MDE algorithm. With asymptotic performance analysis, we show that in the high SNR regime, the local detection error on the observation mode converges to zero and the MDE estimation error converges to that of an ideal estimator with perfect information about the node defect status. The adaptive learning property of the MDE algorithm achieves a reliable estimation performance, in contrast to the unsatisfactory estimation performance of a naive average consensus based algorithm in the high SNR regime.

The rest of this paper is organized as follows. In Section II, we describe the network model first, and then present a naive averaging based estimation scheme and an ideal centralized estimation scheme as benchmarks. Section III presents the MDE algorithm. Section IV summarizes the main results, with some intermediate results proved in Section V. Section VI formally proves the convergence of MDE. In Section VII, we further analyze the performance of MDE, with some asymptotic analysis established in Section VIII. Simulation results are presented in Section IX. Finally, Section X concludes the paper.

II. NETWORK MODEL

Let \mathcal{N}_i and Ω_i , $i \in \{1, 2, \dots, n\}$, denote sensor node i and the set of its neighbors respectively. The received signal at \mathcal{N}_i is $y_i = h_i\theta + w_i$, where $h_i \in \{0, 1\}$ is an unknown validity index of the observation at node \mathcal{N}_i : i.e., $h_i = 1$ indicates that y_i is a valid observation and $h_i = 0$ indicates the invalid observation case. In addition, w_i 's are independent Gaussian white noises with zero mean and variance σ^2 . Although the exact instantiations of the h_i 's are unknown, we assume that h_i 's are i.i.d. Bernoulli random variables and the probability $p_1 \triangleq \Pr\{h_i = 1\}$ is known *a priori*. We denote the variance of

h_i as $\sigma_{h_i}^2$, i.e., $\sigma_{h_i}^2 = p_1(1-p_1)$. We are interested in estimating θ using an iterative distributed procedure, in which each node \mathcal{N}_i may only use its neighbors' current state information for updating its local estimate (state) at time t . We assume that θ is a deterministic unknown target of real value.

Denote by \mathbf{y} the network observation vector, i.e.,

$$\mathbf{y} = \mathbf{h}\theta + \mathbf{w}, \quad (1)$$

with $\mathbf{y} = [y_1, y_2, \dots, y_n]^T$, $\mathbf{h} = [h_1, h_2, \dots, h_n]^T$, and $\mathbf{w} = [w_1, w_2, \dots, w_n]^T$. With this observation model, the sufficient statistic for estimation is \mathbf{y} , and the optimal estimator is given by a maximum a posteriori (MAP) estimator. However, the complexity of MAP is too high to implement in practice. In order to reduce the complexity, we consider the linear estimator model. We note that, a straight-forward approach based on naive averaging could be cast as

$$\hat{\theta}_{\text{Naive}} = \frac{\mathbf{1}^T \mathbf{y}}{np_1}, \quad (2)$$

which yields a linear minimum variance unbiased estimator (LMVUE) with the property that $\hat{\theta}_{\text{Naive}} \rightarrow \theta$ almost surely as $n \rightarrow \infty$. The variance (which coincides with the mean-squared error) of $\hat{\theta}_{\text{Naive}}$ may be expressed as

$$\text{Var}(\hat{\theta}_{\text{Naive}}) = \frac{1}{n} \left[\frac{\sigma^2}{p_1^2} (1 + \text{SNR}\sigma_{h_i}^2) \right], \quad (3)$$

where SNR is defined as $\frac{\theta^2}{\sigma^2}$.

Although this naive estimate is quite straight-forward in terms of implementation [4], [20], [21], we observe from (3) that the precision is poor in the high SNR regime, where in particular, the mean-squared error (MSE) blows up with an increasing SNR. On the other extreme, if we assume that \mathbf{h} is perfectly known, we may generate an *ideal* estimate $\hat{\theta}_{\text{Ideal}}$ of x by eliminating the invalid observations, i.e.,

$$\hat{\theta}_{\text{Ideal}} = \frac{\sum_{\{i:h_i=1\}} y_i}{\sum_{\{i:h_i=1\}} h_i} = \frac{\sum_{i=1}^n h_i y_i}{\sum_{i=1}^n h_i}. \quad (4)$$

The above estimate is also unbiased, with $\hat{\theta}_{\text{Ideal}} \rightarrow \theta$ almost surely as $n \rightarrow \infty$, and its variance may be expressed as

$$\text{Var}(\hat{\theta}_{\text{Ideal}}) = \text{E}(\text{Var}(\hat{\theta}_{\text{Ideal}}|\mathbf{h})) + \text{Var}(\text{E}(\hat{\theta}_{\text{Ideal}}|\mathbf{h})) = \psi\sigma^2, \quad (5)$$

where $\psi = \sum_{k=1}^n \frac{1}{k} \binom{n}{k} p_1^k (1-p_1)^{n-k}$, the derivation of which is given in Appendix A. We note that ψ is not related to SNR and is on the order of $\frac{1}{n}$. For example, when $p_1 = 0.5$, we have $\psi \approx \frac{2-2^{-n}}{n+1}$. A key difference from the naive estimate in (2) is that the variance of the ideal estimate stays constant over SNR, i.e., the estimation error does not scale up with the SNR.

From the MSE viewpoint, the ideal estimate is in fact optimal as long as the observation noise is Gaussian. However, such a scheme may not be implementable as it requires the perfect knowledge of \mathbf{h} , which is unknown *a priori*. In Section III, we introduce a learning-based distributed estimation procedure, the MDE algorithm, based on the iterative detection of \mathbf{h} and estimate refinement of θ . Our results indicate that not only \mathbf{h} could be detected with high accuracy by the MDE algorithm, but also does the estimation performance (in terms

of MSE) approach that of the ideal estimate $\hat{\theta}_{\text{ideal}}$ in the high SNR regime.

III. DISTRIBUTED MDE ALGORITHM

In this section, we present the MDE algorithm for the problem of interest. In each iteration of the MDE algorithm, each node first locally detects the value of h_i by using its current local estimate of θ and its local observation. This initially detected observation validity index is used to update some intermediate parameters, which are subsequently forwarded to the neighboring nodes. This leads to an estimate refinement process, which feeds back new information to improve the validity detection in the next iteration. The algorithm at sensor i is presented as follows.

Step 1. Initialization at time 1

$$\hat{\theta}_i^+(1) = \hat{\theta}_i^-(1) = \frac{\sum_{j \in \Omega_i} y_j}{|\Omega_i| p_1}, \quad \hat{y}_i(1) = y_i. \quad (6)$$

Step 2. Detection of h_i at time $t > 1$

$$(\hat{\theta}_i^+(t))^2 - 2y_i \hat{\theta}_i^+(t) \underset{\hat{h}_i^+(t)=0}{\overset{\hat{h}_i^+(t)=1}{\leq}} 2\sigma^2 \ln \frac{p_1}{p_0}, \quad (7)$$

$$(\hat{\theta}_i^-(t))^2 - 2y_i \hat{\theta}_i^-(t) \underset{\hat{h}_i^-(t)=0}{\overset{\hat{h}_i^-(t)=1}{\leq}} 2\sigma^2 \ln \frac{p_1}{p_0}, \quad (8)$$

where $p_0 = 1 - p_1$.

Step 3. Calculation of intermediate parameters u, v , and the estimation of \bar{y} at time t

$$u_i^+(t) = u_i^+(t-1) - \beta(t) \sum_{j \in \Omega_i} (u_i^+(t-1) - u_j^+(t-1)) + \alpha(t) (y_i \hat{h}_i^+(t) - u_i^+(t-1)), \quad (9)$$

$$v_i^+(t) = v_i^+(t-1) - \beta(t) \sum_{j \in \Omega_i} (v_i^+(t-1) - v_j^+(t-1)) + \alpha(t) (\hat{h}_i^+(t) - v_i^+(t-1)), \quad (10)$$

$$u_i^-(t) = u_i^-(t-1) - \beta(t) \sum_{j \in \Omega_i} (u_i^-(t-1) - u_j^-(t-1)) + \alpha(t) (y_i \hat{h}_i^-(t) - u_i^-(t-1)), \quad (11)$$

$$v_i^-(t) = v_i^-(t-1) - \beta(t) \sum_{j \in \Omega_i} (v_i^-(t-1) - v_j^-(t-1)) + \alpha(t) (\hat{h}_i^-(t) - v_i^-(t-1)), \quad (12)$$

$$\hat{y}_i(t) = \hat{y}_i(t-1) - \beta(t) \sum_{j \in \Omega_i} (\hat{y}_i(t-1) - \hat{y}_j(t-1)), \quad (13)$$

where $u_i^+(0) = y_i \hat{h}_i^+(1)$, $v_i^+(0) = \hat{h}_i^+(1)$, $u_i^-(0) = y_i \hat{h}_i^-(1)$, $v_i^-(0) = \hat{h}_i^-(1)$, and $\alpha(t)$ and $\beta(t)$ satisfy the following four conditions:

- $0 < \alpha(t) < 1$ and $0 < \beta(t) < 1$,
- $\alpha(t) \rightarrow 0$, $\beta(t) \rightarrow 0$,
- $\beta(t)/\alpha(t) \rightarrow \infty$,
- $\sum_{t=1}^{\infty} \alpha(t) = \infty$, $\sum_{t=1}^{\infty} \beta(t) = \infty$.

Step 4. Estimation update of θ

$$\hat{\theta}_i(t+1) = \begin{cases} \hat{\theta}_i^+(t+1), & \hat{y}_i(t) \geq 0 \\ \hat{\theta}_i^-(t+1), & \hat{y}_i(t) < 0 \end{cases} \quad (14)$$

where $\hat{\theta}_i^+(t+1) = \max \left\{ \frac{u_i^+(t)}{v_i^+(t)+\delta}, 0 \right\}$ and $\hat{\theta}_i^-(t+1) = \min \left\{ \frac{u_i^-(t)}{v_i^-(t)+\delta}, 0 \right\}$, with δ as an arbitrary small positive constant, to prevent the denominator from being zero.

We then repeat steps 2 to 4 until $\left| \frac{u_i^+(t)}{v_i^+(t)+\delta} - \frac{u_i^+(t-1)}{v_i^+(t-1)+\delta} \right| < \epsilon$, $\left| \frac{u_i^-(t)}{v_i^-(t)+\delta} - \frac{u_i^-(t-1)}{v_i^-(t-1)+\delta} \right| < \epsilon$, and $|\hat{y}_i(t) - \hat{y}_i(t-1)| < \epsilon$, $\forall i$, where ϵ is a predefined small positive error tolerant parameter.

Basically, the algorithm starts with a linear minimum variance unbiased estimator (LMVUE) among 1-hop neighbors as the initial estimator in step 1. In step 2, each node locally detects (re-assesses) the value of h_i using the current local estimate of θ and y_i . The validity indices, thus obtained, are used to update intermediate parameters that are subsequently forwarded to the neighboring nodes, leading to the state update in step 3, where each node refines its local parameters, i.e., $u_i^+(t)$, $u_i^-(t)$, $v_i^+(t)$, and $v_i^-(t)$, based on new information from its neighbors using a consensus + innovations type mechanism. (The consensus potential governs how neighboring observations are assimilated to seek agreement among agents, whereas, the local innovation potential may be viewed as a refinement capturing the agent's local observation and its instantaneous validity measure.) Finally, a new estimate is obtained from $u_i^+(t)$, $u_i^-(t)$, $v_i^+(t)$, and $v_i^-(t)$, and a new iteration starts if needed. In the next section, we investigate the convergence of this iterative procedure. We also emphasize that the conditions on $\alpha(t)$ and $\beta(t)$ listed above are not hard to satisfy. For example, we may choose $\alpha(t) = \delta_a/t$, and $\beta(t) = \delta_b/t^{1-\epsilon}$, with $\epsilon \in (0, 1)$, δ_a and δ_b as small positive real constants.

IV. MAIN RESULTS

In this section, we present the main results, with the proofs given in the subsequent sections.

Theorem 1: Let the inter-sensor communication network be connected¹, and assume that $\alpha(t)$ and $\beta(t)$ in (9)-(13) satisfy the following four conditions:

- $0 < \alpha(t) < 1$ and $0 < \beta(t) < 1$,
- $\alpha(t) \rightarrow 0$, $\beta(t) \rightarrow 0$,
- $\beta(t)/\alpha(t) \rightarrow \infty$,
- $\sum_{t=1}^{\infty} \alpha(t) = \infty$, $\sum_{t=1}^{\infty} \beta(t) = \infty$.

Then, the estimate sequence $\{\hat{\theta}_i(t)\}$ at each node \mathcal{N}_i converges almost surely as

$$\lim_{t \rightarrow \infty} \hat{\theta}_i(t) = \begin{cases} \max \left\{ \frac{\sum_{j=1}^n \hat{h}_j^+ y_j}{\sum_{j=1}^n \hat{h}_j^+ + n\delta}, 0 \right\}, & \text{on the event } \{\bar{y} \geq 0\} \\ \min \left\{ \frac{\sum_{j=1}^n \hat{h}_j^- y_j}{\sum_{j=1}^n \hat{h}_j^- + n\delta}, 0 \right\}, & \text{on the event } \{\bar{y} < 0\} \end{cases}, \forall i, \quad (15)$$

where $\hat{h}_i^{(\cdot)} \in \{0, 1\}$ denotes the limiting value of the convergent sequence $\{\hat{h}_i^{(\cdot)}(t)\}$, in which we use (\cdot) to denote either + or -; \bar{y} is the arithmetic mean of all y_i 's. Note that $\hat{h}_i^{(\cdot)}$ is,

¹The network is said to be connected if there exists a path (possibly multi-hop) between any pair of nodes.

in general, random given the stochasticity of the h_i s and the y_i 's.

The proof of Theorem 1 is presented in Section VI. Theorem 1 shows that the estimate sequence $\{\hat{\theta}_i(t)\}$ at each node converges to a unique (stochastic) limit, denoted by $\hat{\theta}_{\text{Ideal}}$, as $t \rightarrow \infty$, which implies that the nodes in the network achieve agreement over the estimate of the unknown parameter θ , i.e., realizing the network consensus. Since we consider a general real valued parameter θ , according to the proposed algorithm, the limiting estimate value takes on different forms depending on whether the event $\{\bar{y} \geq 0\}$ or its complement holds, reflecting the possible non-negativity or negativity of the parameter θ respectively. We further prove in Theorem 3 that this converged estimation value is unbiased in the asymptotic regime as SNR goes to infinity.

Theorem 2: If we order the observations $\{y_i\}$ in the increasing order as $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$, and denote the corresponding decisions given in step 2 of the proposed algorithm as $\hat{h}_{(1)}^{(\cdot)}, \hat{h}_{(2)}^{(\cdot)}, \dots, \hat{h}_{(n)}^{(\cdot)}$, we have

$$\hat{h}_{(1)}^+ \leq \hat{h}_{(2)}^+ \leq \dots \leq \hat{h}_{(n)}^+, \quad (16)$$

and

$$\hat{h}_{(1)}^- \geq \hat{h}_{(2)}^- \geq \dots \geq \hat{h}_{(n)}^-, \quad (17)$$

where $\hat{h}_{(i)}^{(\cdot)} \in \{0, 1\}$.

We prove Theorem 2 in Section VII-A. Theorem 2 demonstrates an interesting property of the proposed algorithm: if the observations from different nodes are ordered, the corresponding $\hat{h}_{(i)}^{(\cdot)}$'s are also ordered. Specifically, if the observations are increasingly ordered, $\hat{h}_{(i)}^+$'s have the same increasing order as that of observations, while $\hat{h}_{(i)}^-$'s inherit a decreasing order. Since $\hat{h}_{(i)}^+$'s correspond to (7) with non-negative $\theta_i^+(t)$ and $\hat{h}_{(i)}^-$'s correspond to (8) with non-positive $\theta_i^-(t)$, this intuitively explains why $\hat{h}_{(i)}^+$'s and $\hat{h}_{(i)}^-$'s have different orders.

Theorem 3: For the MDE algorithm, we have

$$\lim_{\text{SNR} \rightarrow \infty} E(\hat{\theta}) = \theta, \quad (18)$$

where $\hat{\theta}$ is the converged value shown in (15). Since the converged value in (15) does not depend on the node index i , the index is dropped.

The proof of Theorem 3 is presented in Section VII-B. Theorem 3 shows that the converged estimation value in (15) is unbiased in the asymptotic regime as $\text{SNR} \rightarrow \infty$.

Theorem 4: For the MDE algorithm, we have

$$\lim_{n \rightarrow \infty, \text{SNR} \rightarrow \infty} \text{Var}(\hat{\theta}) = \text{Var}(\hat{\theta}_{\text{Ideal}}), \quad (19)$$

where $\hat{\theta}_{\text{Ideal}}$ is the ideal estimator defined in (4).

The proof of Theorem 4 is given in Section VIII. Theorem 4 shows that the estimation error variance converges almost surely to that of the ideal estimate $\hat{\theta}_{\text{Ideal}}$ defined in (4), when both node number n and SNR increase.

By combining Theorem 3 and Theorem 4, we see that the performance of our proposed distributed algorithm converges to that of the ideal estimate $\hat{\theta}_{\text{Ideal}}$ defined in (4). Since this ideal estimate is computed based on the assumption that \mathbf{h} is

perfectly known or precisely learned, as an optimal estimation method, its performance is the benchmark of all other estimation algorithms to deal with unknown sensor defects. Theorem 3 and Theorem 4 imply that the proposed distributed algorithm converges to the optimal solution and the validity index \mathbf{h} can be precisely learned, as SNR goes to infinity.

V. INTERMEDIATE RESULTS FOR PROOFS

In this section, we establish some intermediate results to be used later. In the MDE algorithm, we note that the positive and negative parts are symmetric, i.e., $\hat{\theta}_i^+(t)$ vs. $\hat{\theta}_i^-(t)$, $h_i^+(t)$ vs. $h_i^-(t)$, $u_i^+(t)$ vs. $u_i^-(t)$, and $v_i^+(t)$ vs. $v_i^-(t)$. In the following, we use (\cdot) to denote either $+$ or $-$ and the results can be applied to both of these two cases. We denote $\frac{\sum_i u_i^{(\cdot)}(t)}{n}$ and $\frac{\sum_i v_i^{(\cdot)}(t)}{n}$ as $\bar{u}^{(\cdot)}(t)$ and $\bar{v}^{(\cdot)}(t)$, respectively. In the following, Lemma 5 proves that $\bar{u}^{(\cdot)}(t)$ is a bounded sequence. Then we show the limiting relationship between $\bar{u}^{(\cdot)}(t)$ and $u_i^{(\cdot)}(t)$ in Lemma 6, where $\lim_{t \rightarrow \infty} (u_i^{(\cdot)}(t) - \bar{u}^{(\cdot)}(t)) = 0$. Both $u_i^{(\cdot)}(t)$ and $\bar{u}^{(\cdot)}(t)$ in the above results could be replaced by $v_i^{(\cdot)}(t)$ and $\bar{v}^{(\cdot)}(t)$ respectively and the proofs are similar. Then, Lemma 7 proves that $\lim_{t \rightarrow \infty} (\frac{\bar{u}^{(\cdot)}(t+1)}{\bar{v}^{(\cdot)}(t+1)+\delta} - \frac{\bar{u}^{(\cdot)}(t)}{\bar{v}^{(\cdot)}(t)+\delta}) = 0$. After that, the limiting relationship between $\hat{\theta}_i(t)$ and $\frac{\bar{u}^{(\cdot)}(t)}{\bar{v}^{(\cdot)}(t)+\delta}$ is proved in Lemma 8.

Lemma 5: Let the inter-sensor communication network be connected. Thus we have that $\bar{u}^{(\cdot)}(t)$ is a bounded sequence.

Proof: In step 3 of the algorithm, we have

$$u_i^{(\cdot)}(t) = u_i^{(\cdot)}(t-1) + \alpha(t)(y_i \hat{h}_i^{(\cdot)}(t) - u_i^{(\cdot)}(t-1)) - \beta(t) \sum_{j \in \Omega_i} (u_i^{(\cdot)}(t-1) - u_j^{(\cdot)}(t-1)). \quad (20)$$

Taking the average on both sides over all $i \in [1, \dots, n]$, we have the iterative expression of $\bar{u}^{(\cdot)}(t)$ as follows

$$\begin{aligned} \bar{u}^{(\cdot)}(t) &= \\ &\bar{u}^{(\cdot)}(t-1) + \alpha(t) (\{y_i \hat{h}_i^{(\cdot)}(t)\}_{\text{avg}} - \bar{u}^{(\cdot)}(t-1)), \end{aligned} \quad (21)$$

where $\{y_i \hat{h}_i^{(\cdot)}(t)\}_{\text{avg}} = \sum_{i=1}^n (y_i \hat{h}_i^{(\cdot)}(t))/n$.

We rewrite the above equation in another form as

$$\bar{u}^{(\cdot)}(t) = (1 - \alpha(t)) \bar{u}^{(\cdot)}(t-1) + \alpha(t) \{y_i \hat{h}_i^{(\cdot)}(t)\}_{\text{avg}}; \quad (22)$$

and for $\bar{u}^{(\cdot)}(t+1)$, we have

$$\begin{aligned} \bar{u}^{(\cdot)}(t+1) &= (1 - \alpha(t+1)) \bar{u}^{(\cdot)}(t) + \alpha(t+1) \\ &\{y_i \hat{h}_i^{(\cdot)}(t+1)\}_{\text{avg}}. \end{aligned} \quad (23)$$

By substituting (22) into the right-side of (23), we have

$$\begin{aligned} |\bar{u}^{(\cdot)}(t+1)| &= \\ &\left| (1 - \alpha(t+1)) [(1 - \alpha(t)) \bar{u}^{(\cdot)}(t-1) + \alpha(t) \{y_i \hat{h}_i^{(\cdot)}(t)\}_{\text{avg}}] \right. \\ &\left. + \alpha(t+1) \{y_i \hat{h}_i^{(\cdot)}(t+1)\}_{\text{avg}} \right| \end{aligned} \quad (24)$$

$$\begin{aligned}
&= \left| \prod_{j=t-1}^t (1 - \alpha(j+1)) \bar{u}^{(\cdot)}(t-1) + (1 - \alpha(t+1)) \right. \\
&\quad \left. \alpha(t) \{y_i \hat{h}_i^{(\cdot)}(t)\}_{avg} + \alpha(t+1) \{y_i \hat{h}_i^{(\cdot)}(t+1)\}_{avg} \right| \\
&\leq \prod_{j=t-1}^t (1 - \alpha(j+1)) |\bar{u}^{(\cdot)}(t-1)| + (1 - \alpha(t+1)) \\
&\quad \alpha(t) y_{max} + \alpha(t+1) y_{max} \\
&= \prod_{j=t-1}^t (1 - \alpha(j+1)) (|\bar{u}^{(\cdot)}(t-1)| - y_{max}) + y_{max}, \quad (25)
\end{aligned}$$

where $y_{max} = \max_{i=1}^n |y_i|$ is the natural upper bound of $\{y_i \hat{h}_i^{(\cdot)}(t)\}_{avg}$. Iteratively, we deduce that

$$|\bar{u}^{(\cdot)}(t+1)| \leq \prod_{j=1}^t (1 - \alpha(j+1)) (|\bar{u}^{(\cdot)}(1)| - y_{max}) + y_{max}. \quad (26)$$

Note that $1 - a \leq e^{-a}$ for $0 \leq a \leq 1$; thus we have

$$\begin{aligned}
&\prod_{i=1}^t (1 - \alpha(i+1)) (|\bar{u}^{(\cdot)}(1)| - y_{max}) + y_{max} \\
&\leq e^{-\sum_{i=1}^t \alpha(i+1)} (|\bar{u}^{(\cdot)}(1)| - y_{max}) + y_{max}. \quad (27)
\end{aligned}$$

When $t \rightarrow \infty$, we have $e^{-\sum_{i=0}^t \alpha(i+1)} \rightarrow 0$ by the fourth condition of $\alpha(t)$; and then we conclude that $\bar{u}^{(\cdot)}(t)$ is a bounded function. ■

Lemma 6: Let the inter-sensor communication network be connected. We have

$$\begin{aligned}
\lim_{t \rightarrow \infty} (u_i^{(\cdot)}(t) - \bar{u}^{(\cdot)}(t)) &= 0, \quad \forall i \\
\lim_{t \rightarrow \infty} (v_i^{(\cdot)}(t) - \bar{v}^{(\cdot)}(t)) &= 0, \quad \forall i
\end{aligned}$$

with $\bar{u}^{(\cdot)}(t)$ and $\bar{v}^{(\cdot)}(t)$ defined previously.

Proof: This Lemma can be proved by applying Lemma 15 in [44], which is skipped here. ■

In Lemmas 5 and 6, $u_i^{(\cdot)}(t)$ and $\bar{u}^{(\cdot)}(t)$ could be directly replaced by $v_i^{(\cdot)}(t)$ and $\bar{v}^{(\cdot)}(t)$ respectively, and the proofs are similar.

Lemma 7: Let the inter-sensor communication network be connected. Then,

$$\lim_{t \rightarrow \infty} \left(\frac{\bar{u}^{(\cdot)}(t+1)}{\bar{v}^{(\cdot)}(t+1) + \delta} - \frac{\bar{u}^{(\cdot)}(t)}{\bar{v}^{(\cdot)}(t) + \delta} \right) = 0. \quad (28)$$

Proof: We have

$$\begin{aligned}
&\frac{\bar{u}^{(\cdot)}(t+1)}{\bar{v}^{(\cdot)}(t+1) + \delta} - \frac{\bar{u}^{(\cdot)}(t)}{\bar{v}^{(\cdot)}(t) + \delta} \\
&= \frac{(1 - \alpha(t+1)) \bar{u}^{(\cdot)}(t) + \alpha(t+1) \{y_i \hat{h}_i^{(\cdot)}(t+1)\}_{avg}}{(1 - \alpha(t+1)) \bar{v}^{(\cdot)}(t) + \alpha(t+1) \{\hat{h}_i^{(\cdot)}(t+1)\}_{avg} + \delta} \\
&\quad - \frac{\bar{u}^{(\cdot)}(t)}{\bar{v}^{(\cdot)}(t) + \delta}
\end{aligned} \quad (29)$$

$$\begin{aligned}
&= \left\{ \frac{\{y_i \hat{h}_i^{(\cdot)}(t+1)\}_{avg} (\bar{v}^{(\cdot)}(t) + \delta) - \bar{u}^{(\cdot)}(t) \delta}{(1 - \alpha(t+1)) \bar{v}^{(\cdot)}(t) + \alpha(t) \{\hat{h}_i^{(\cdot)}(t+1)\}_{avg} + \delta} \right. \\
&\quad \left. - \frac{\{\hat{h}_i^{(\cdot)}(t+1)\}_{avg} \bar{u}^{(\cdot)}(t)}{(1 - \alpha(t+1)) \bar{v}^{(\cdot)}(t) + \alpha(t) \{\hat{h}_i^{(\cdot)}(t+1)\}_{avg} + \delta} \right\} \\
&\quad \cdot \frac{\alpha(t+1)}{(\bar{v}^{(\cdot)}(t) + \delta)} \quad (30)
\end{aligned}$$

By Lemma 5, both $\bar{v}^{(\cdot)}(t)$ and $\bar{u}^{(\cdot)}(t)$ are bounded (both upper- and lower-bounded). In addition, $\{y_i \hat{h}_i^{(\cdot)}(t+1)\}_{avg}$ and $\{\hat{h}_i^{(\cdot)}(t+1)\}_{avg}$ are naturally bounded. Together with the fact that δ is an arbitrarily small constant, and $\lim_{t \rightarrow \infty} \alpha(t) = 0$, we conclude that $\lim_{t \rightarrow \infty} \left(\frac{\bar{u}^{(\cdot)}(t+1)}{\bar{v}^{(\cdot)}(t+1) + \delta} - \frac{\bar{u}^{(\cdot)}(t)}{\bar{v}^{(\cdot)}(t) + \delta} \right) = 0$. ■

Lemma 8: Let the inter-sensor communication network be connected. Then,

$$\begin{aligned}
\lim_{t \rightarrow \infty} \left(\hat{\theta}_i^+(t+1) - \max \left\{ \frac{\bar{u}^+(t)}{\bar{v}^+(t) + \delta}, 0 \right\} \right) &= 0, \quad \forall i, \\
\lim_{t \rightarrow \infty} \left(\hat{\theta}_i^-(t+1) - \min \left\{ \frac{\bar{u}^-(t)}{\bar{v}^-(t) + \delta}, 0 \right\} \right) &= 0, \quad \forall i,
\end{aligned}$$

where $\bar{u}^{(\cdot)}(t)$ and $\bar{v}^{(\cdot)}(t)$ denote the averaging values of $u_i^{(\cdot)}(t)$ and $v_i^{(\cdot)}(t)$, respectively.

Proof: Recall $\hat{\theta}_i^+(t+1) = \max \left\{ \frac{u_i^+(t)}{v_i^+(t) + \delta}, 0 \right\}$ and $\hat{\theta}_i^-(t+1) = \min \left\{ \frac{u_i^-(t)}{v_i^-(t) + \delta}, 0 \right\}$. We have

$$\begin{aligned}
\lim_{t \rightarrow \infty} \left(\frac{u_i^{(\cdot)}(t)}{v_i^{(\cdot)}(t) + \delta} - \frac{\bar{u}^{(\cdot)}(t)}{\bar{v}^{(\cdot)}(t) + \delta} \right) &= \\
\lim_{t \rightarrow \infty} \left(\frac{u_i^{(\cdot)}(t) \bar{v}^{(\cdot)}(t) - \bar{u}^{(\cdot)}(t) v_i^{(\cdot)}(t) + \delta (u_i^{(\cdot)}(t) - \bar{u}^{(\cdot)}(t))}{(v_i^{(\cdot)}(t) + \delta) (\bar{v}^{(\cdot)}(t) + \delta)} \right) &= 0,
\end{aligned}$$

which is according to Lemma 6. Therefore, the proof is completed. ■

VI. PROOF OF THEOREM 1

In this section, we prove the convergence and derive the limiting value for Theorem 1. Without loss of generality, we prove the case of $\hat{\theta}_i^+(t)$ and skip the proof of $\hat{\theta}_i^-(t)$, which is similar. We first partition the real axis in Subsection VI-A, such that the detection of $\hat{\mathbf{h}}^+(t)$ has the same results when $\hat{\theta}_i^+(t)$'s are in the same interval. Then we derive the smooth moving condition in Subsection VI-B, under which $\hat{\theta}_i^+(t)$ moves on the real axis by passing the partitions sequentially along the iteration process, such that the changing of $\hat{\mathbf{h}}^+(t)$ is successive with time. From the proposed algorithm, we notice that the local estimation is the greater one between 0 and $\frac{u_i^+(t)}{v_i^+(t) + \delta}$, when $\hat{y}_i(t) \geq 0$. As such, we only need to prove the convergence of $\frac{u_i^+(t)}{v_i^+(t) + \delta}$, and then the convergence of $\hat{\theta}_i^+(t)$ is guaranteed. In Subsection VI-C, we complete the proof of Theorem 1.

A. Partitions of the Real Axis

We now seek a suitable scale to study the iteration procedure. We start by exploring step 2 of the proposed algorithm. For each i , we make a hard decision at step 2. We define the region that returns $\hat{h}_i^+(t) = 1$ as the decision region of $\hat{\theta}_i^+(t)$, denoted by \mathcal{D}_i . In particular, if $y_i^2 + 2\sigma^2 \ln \frac{p_1}{p_0} \geq 0$, we have $\mathcal{D}_i = [r_i^-, r_i^+]$ for node i , where $r_i^- = y_i - \sqrt{y_i^2 + 2\sigma^2 \ln \frac{p_1}{p_0}}$, $r_i^+ = y_i + \sqrt{y_i^2 + 2\sigma^2 \ln \frac{p_1}{p_0}}$; otherwise, $\mathcal{D}_i = \emptyset$. Next we partition the real axis into at most $2n+1$ parts by these boundaries of \mathcal{D}_i 's, i.e., r_i^- 's and r_i^+ 's. Here, we say "at most" due to the fact that some of the r_i^+ 's may not exist, e.g., when $y_i^2 + 2\sigma^2 \ln \frac{p_1}{p_0} < 0$ or when multiple boundaries share the same value. Then we name these boundaries in an increasing order of their values as b_1 to b_M and name the partitioned left-open and right-closed intervals as \mathcal{I}_1 to \mathcal{I}_{M+1} , from left to right on the real axis.

B. Smooth Moving Condition

In this subsection, we define the gathering region of $\{\hat{\theta}_i^+(t)\}$ as $\mathcal{G}^+(t)$, which is the range that covers all possible values of $\hat{\theta}_i^+(t)$'s. Then we study the condition for $\mathcal{G}^+(t)$ to move on the axis smoothly during the iteration process. In other words, the gathering region touches those boundaries b_m 's (from $\{b_1, \dots, b_M\}$) sequentially in order without jumping if it passes through the boundaries. Also, for each time, the gathering region $\mathcal{G}^+(t)$ touches at most one of those different boundaries at each iteration. Next, we propose two conditions to guarantee the above situation.

We choose ε that is at least 3 (the reason of choosing 3 is explained at the end of this subsection) times smaller than the narrowest range in \mathcal{K}_j 's, i.e., $3\varepsilon < \min_j \{|\mathcal{K}_j|\}$, where \mathcal{K}_j 's are the intervals partitioned jointly by b_m 's and y_i 's. (such that the number of \mathcal{K}_j 's is larger, the minimum length of \mathcal{K}_j 's is shorter, than \mathcal{I}_m 's).

- By Lemma 8, we have

$$\lim_{t \rightarrow \infty} \left(\hat{\theta}_i^+(t) - \max \left\{ \frac{\bar{u}^+(t-1)}{\bar{v}^+(t-1) + \delta}, 0 \right\} \right) = 0.$$

Thus we could find t_ε , such that for any $t > t_\varepsilon$, we have $\left| \hat{\theta}_i^+(t) - \max \left\{ \frac{\bar{u}^+(t-1)}{\bar{v}^+(t-1) + \delta}, 0 \right\} \right| < \varepsilon$.

- By Lemma 7, we have

$$\lim_{t \rightarrow \infty} \left(\frac{\bar{u}^+(t+1)}{\bar{v}^+(t+1) + \delta} - \frac{\bar{u}^+(t)}{\bar{v}^+(t) + \delta} \right) = 0.$$

Thus we could find t_α , such that for any $t > t_\alpha$, $\left| \frac{\bar{u}^+(t+1)}{\bar{v}^+(t+1) + \delta} - \frac{\bar{u}^+(t)}{\bar{v}^+(t) + \delta} \right| < \varepsilon$.

When $t > \max(t_\varepsilon, t_\alpha)$, we define $\mathcal{G}^+(t) = \left(\max \left\{ \frac{\bar{u}^+(t-1)}{\bar{v}^+(t-1) + \delta}, 0 \right\} - \varepsilon, \max \left\{ \frac{\bar{u}^+(t-1)}{\bar{v}^+(t-1) + \delta}, 0 \right\} + \varepsilon \right)$ as the gathering region of $\hat{\theta}_i^+(t)$, i.e., $\hat{\theta}_i^+(t) \in \mathcal{G}^+(t)$, $\forall i$. Since $\varepsilon < \frac{1}{3} \min_j \{|\mathcal{K}_j|\} \leq \frac{1}{3} \min_m \{|\mathcal{I}_m|\}$, $\mathcal{G}^+(t)$ does not touch or pass two successive b_m 's during two successive iterations as desired. In the sequel, we assume all the iterations under concern satisfy $t > \max(t_\varepsilon, t_\alpha)$.

C. Proof of Theorem 1

We now prove the convergence result stated in Theorem 1.

Proof: In this proof, we first prove that the estimate sequence $\{\hat{\theta}_i^{(\cdot)}(t)\}$ at each node \mathcal{N}_i converges almost surely (a.s.), and the limiting value is given by

$$\lim_{t \rightarrow \infty} \hat{\theta}_i^+(t) = \max \left\{ \frac{\sum_{i=1}^n \hat{h}_i^+ y_i}{\sum_{i=1}^n \hat{h}_i^+ + n\delta}, 0 \right\}, \forall i,$$

$$\lim_{t \rightarrow \infty} \hat{\theta}_i^-(t) = \min \left\{ \frac{\sum_{i=1}^n \hat{h}_i^- y_i}{\sum_{i=1}^n \hat{h}_i^- + n\delta}, 0 \right\}, \forall i.$$

with $\hat{h}_i^{(\cdot)} \in \{0, 1\}$ denoting the limiting value of the convergent sequence $\{\hat{h}_i^{(\cdot)}(t)\}$. We then use the fact that $\lim_{t \rightarrow \infty} \hat{y}_i(t) = \bar{y}$, $\forall i$ [44], to prove the convergence of $\{\hat{\theta}_i(t)\}$.

Without loss of generality, we only prove the positive case, i.e., $\{\hat{\theta}_i^+(t)\}$. In Lemma 8, we have proved that $\lim_{t \rightarrow \infty} (\hat{\theta}_i^+(t) - \max\{\hat{\theta}_{current}^+(t), 0\}) = 0$. Thus, we only need to show that $\max\{\hat{\theta}_{current}^+(t), 0\}$ converges. Since $\mathcal{G}^+(t) = (\max\{\hat{\theta}_{current}^+(t), 0\} - \varepsilon, \max\{\hat{\theta}_{current}^+(t), 0\} + \varepsilon)$, the study on $\max\{\hat{\theta}_{current}^+(t), 0\}$ is equivalent to the study on $\mathcal{G}^+(t)$ in term of convergence. By the smooth moving condition, there is at most one b_k in $\mathcal{G}^+(t)$, $\forall t$. Thus, there are two different moving statuses of $\hat{\theta}_{current}^+(t)$ at each iteration cataloged by the number of boundaries in $\mathcal{G}^+(t)$:

- Case 1: No boundaries belong to $\mathcal{G}^+(t)$, i.e., $b_k \notin \mathcal{G}^+(t)$, $\forall k$. In other words, $\mathcal{G}^+(t)$ belongs to a single interval \mathcal{I}_j , i.e., $\mathcal{G}^+(t) \subseteq \mathcal{I}_j$.
- Case 2: A boundary exists in $\mathcal{G}^+(t)$, i.e., $\exists k, b_k \in \mathcal{G}^+(t)$.

In Appendix B, we provide Lemmas 10 through 14. Specifically, in Lemmas 10, 11, and 12, we prove the convergence of $\max\{\hat{\theta}_{current}^+(t), 0\}$. In particular, we show that $\max\{\hat{\theta}_{current}^+(t), 0\}$ either converges or the moving status switches to the other one for Case 1 and Case 2 in Lemma 10 and Lemma 11, respectively. For the moving status switching, Lemma 12 further shows that the number of switching between Case 1 and Case 2 is finite, which implies the convergence of $\max\{\hat{\theta}_{current}^+(t), 0\}$. In Lemmas 13 and 14, we further derive the limiting values for Case 1 and Case 2, respectively.

Together with the fact that $\lim_{t \rightarrow \infty} \hat{y}_i(t) = \bar{y}$, $\forall i$, the convergence of $\hat{\theta}_i(t)$ is guaranteed, which could be expressed as

$$\lim_{t \rightarrow \infty} \hat{\theta}_i(t) = \begin{cases} \max \left\{ \frac{\sum_{i=1}^n \hat{h}_i^+ y_i}{\sum_{i=1}^n \hat{h}_i^+ + n\delta}, 0 \right\}, & \bar{y} \geq 0 \\ \min \left\{ \frac{\sum_{i=1}^n \hat{h}_i^- y_i}{\sum_{i=1}^n \hat{h}_i^- + n\delta}, 0 \right\}, & \bar{y} < 0 \end{cases}, \forall i. \quad (31)$$

■

VII. PROOFS FOR THEOREM 2 AND THEOREM 3

In this section, we derive the expectation and the variance of local estimate with the proposed algorithm. In Section VI, we have proven that $\hat{\theta}_i^+$ in the MDE algorithm converges to $\max \left\{ \frac{\sum_i \hat{h}_i^+ y_i / n}{\sum_i \hat{h}_i^+ / n + \delta}, 0 \right\}$. Since δ can be arbitrarily small, we approximate the converged value $\hat{\theta}^+$ as $\max \left\{ \frac{\sum_i \hat{h}_i^+ y_i}{\sum_i \hat{h}_i^+}, 0 \right\}$ here. In addition, the converged values $\hat{\theta}^+$'s (even with the

same initial observations) may be different over different network realizations. In particular, the proposed algorithm might lead to random realizations of $\hat{\theta}^+$ and $\hat{\mathbf{h}}^+$, which satisfy

$$(\hat{\theta}^+)^2 - 2y_i\hat{\theta}^+ \underset{\hat{h}_i^+=1}{\overset{\hat{h}_i^+=0}{\geq}} 2\sigma^2 \ln \frac{p_1}{p_0} \quad (32)$$

$$\hat{\theta}^+ = \max \left\{ \frac{\sum_i \hat{h}_i^+ y_i}{\sum_i \hat{h}_i^+}, 0 \right\} \geq 0, \quad (33)$$

where $\hat{\mathbf{h}}^+$ is a random vector denoting $[\hat{h}_1^+, \dots, \hat{h}_n^+]$. In total, there are 2^n possible random values for $\hat{\mathbf{h}}^+$. In order to derive a meaningful result, we adopt order statistics into the rest of the analysis. In Subsection VII-A, we first prove Theorem 2 to establish the shrinking over the dimension of the probability space from 2^n to $2n$, with a more structured format when we order the observations. We then study the expectation of $\hat{\theta}$ in Theorem 3 at Subsection VII-B. We also study the variance $\text{Var}(\hat{\theta}^{(\cdot)})$ of $\hat{\theta}$, whose elements are derived respectively in Subsections VII-C, VII-D, and VII-E.

A. Shrinking the Probability Space of $\hat{\mathbf{h}}^{(\cdot)}$

In this subsection, we prove Theorem 2 to establish the shrinking over the probability space of interest when we order the observations.

Proof: Here we only prove the $\hat{\mathbf{h}}^+$ part, for the proof of the $\hat{\mathbf{h}}^-$ part is similar. We define the decision region of $\hat{h}_{(i)}^+$ as $\mathcal{D}_{(i)}^+$, which is the region of $\hat{\theta}^+$ when $\hat{h}_{(i)}^+ = 1$. By (32), $\mathcal{D}_{(i)}^+$ can be expressed as:

1) If $y_{(i)}^2 + 2\sigma^2 \ln \frac{p_1}{p_0} < 0$, we have $\hat{h}_{(i)}^+ = 0$ for any $\hat{\theta}^+$. Thus, we have $\mathcal{D}_{(i)}^+ = \emptyset$;

2) If $y_{(i)}^2 + 2\sigma^2 \ln \frac{p_1}{p_0} \geq 0$, we have $\mathcal{D}_{(i)}^+ = \left[y_{(i)} - \sqrt{y_{(i)}^2 + 2\sigma^2 \ln \frac{p_1}{p_0}}, y_{(i)} + \sqrt{y_{(i)}^2 + 2\sigma^2 \ln \frac{p_1}{p_0}} \right]$.

The proof here is equivalent to proving that $\mathcal{D}_{(1)}^+ \subseteq \mathcal{D}_{(2)}^+ \subseteq \dots \subseteq \mathcal{D}_{(n)}^+$ is true. Next, we prove the above statement for both of the two cases: $p_1 \geq 0.5$ and $p_1 < 0.5$.

Case 1: $p_1 \geq 0.5$. In this case, we have $2\sigma^2 \ln \frac{p_1}{p_0} \geq 0$ and $\mathcal{D}_{(i)}^+ \neq \emptyset$ for all i . For the upper boundaries of $\mathcal{D}_{(i)}^+$'s, they are increasing with their index i , which could be proven by showing that $r(y) = y + \sqrt{y^2 + 2\sigma^2 \ln \frac{p_1}{p_0}}$ is a monotonic increasing function when $2\sigma^2 \ln \frac{p_1}{p_0} \geq 0$, i.e.,

$$\begin{aligned} r'(y) &= \left(y + \sqrt{y^2 + 2\sigma^2 \ln \frac{p_1}{p_0}} \right)' \\ &= 1 + \frac{y}{\sqrt{y^2 + 2\sigma^2 \ln \frac{p_1}{p_0}}} > 0. \end{aligned} \quad (34)$$

For the lower boundaries of $\mathcal{D}_{(i)}^+$'s, they are all negative. Since $\hat{\theta}^+$ is always positive, the negative part of $\mathcal{D}_{(i)}^+$'s are infeasible. Thus, we redefine

$\mathcal{D}_{(i)}^+ = \left[0, y_{(i)} + \sqrt{y_{(i)}^2 + 2\sigma^2 \ln \frac{p_1}{p_0}} \right]$ in this case. Thus, we conclude that $\mathcal{D}_{(1)}^+ \subseteq \mathcal{D}_{(2)}^+ \subseteq \dots \subseteq \mathcal{D}_{(n)}^+$ when $p_1 \geq 0.5$.

Case 2: $p_1 < 0.5$. In this case, we have $2\sigma^2 \ln \frac{p_1}{p_0} < 0$. Next, we derive the expression of $\mathcal{D}_{(i)}^+$ for different values

of $y_{(i)}$. When $y_{(i)}^2 < -2\sigma^2 \ln \frac{p_1}{p_0}$, we have $\mathcal{D}_{(i)}^+ = \emptyset$; When $y_{(i)}^2 \geq -2\sigma^2 \ln \frac{p_1}{p_0}$, for the case of $y_{(i)} \leq -\sqrt{-2\sigma^2 \ln \frac{p_1}{p_0}}$, $\mathcal{D}_{(i)}^+ = \left[y_{(i)} - \sqrt{y_{(i)}^2 + 2\sigma^2 \ln \frac{p_1}{p_0}}, y_{(i)} + \sqrt{y_{(i)}^2 + 2\sigma^2 \ln \frac{p_1}{p_0}} \right]$ is in the negative field. Since $\hat{\theta}^+$ is always positive, the case of $y_{(i)} \leq -\sqrt{-2\sigma^2 \ln \frac{p_1}{p_0}}$ is infeasible. Thus we only need to consider the case of $y_{(i)} \geq \sqrt{-2\sigma^2 \ln \frac{p_1}{p_0}}$. We then have $\mathcal{D}_{(i)}^+ = \left[y_{(i)} - \sqrt{y_{(i)}^2 + 2\sigma^2 \ln \frac{p_1}{p_0}}, y_{(i)} + \sqrt{y_{(i)}^2 + 2\sigma^2 \ln \frac{p_1}{p_0}} \right]$, where the upper boundary is an increasing sequence over i by the same argument as (34) and the lower boundary is a positive decreasing sequence over i , which could be proven by showing that $j(y) = y - \sqrt{y^2 + 2\sigma^2 \ln \frac{p_1}{p_0}}$ is a monotonic decreasing function when $2\sigma^2 \ln \frac{p_1}{p_0} < 0$, i.e.,

$$\begin{aligned} j'(y) &= \left(y - \sqrt{y^2 + 2\sigma^2 \ln \frac{p_1}{p_0}} \right)' \\ &= 1 - \frac{y}{\sqrt{y^2 + 2\sigma^2 \ln \frac{p_1}{p_0}}} < 0. \end{aligned} \quad (35)$$

Therefore, we have the same conclusion as the previous case, and we conclude that $\mathcal{D}_{(1)}^+ \subseteq \mathcal{D}_{(2)}^+ \subseteq \dots \subseteq \mathcal{D}_{(n)}^+$ as desired. ■

We denote the corresponding convergence vector according to the ordered observations as a random vector $\mathbf{h}^{(\cdot)}$. Although there are totally 2^n possible values for $\mathbf{h}^{(\cdot)}$, only n possible values are in the probability space of \mathbf{h}^+ or \mathbf{h}^- , i.e., $\mathbf{h}_1^+ = [1, 1, \dots, 1]$, $\mathbf{h}_2^+ = [0, 1, \dots, 1]$, ..., $\mathbf{h}_n^+ = [0, 0, \dots, 0, 1]$, and $\mathbf{h}_1^- = [1, 1, \dots, 1]$, $\mathbf{h}_2^- = [1, \dots, 1, 0]$, ..., $\mathbf{h}_n^- = [1, 0, \dots, 0, 0]$, which means that the possible values of \mathbf{h}^+ could only be in the form that starts with successive 0's and followed with successive 1's, with similar rules held for \mathbf{h}^- .

B. Expectation of $\hat{\theta}$

In this subsection, we prove Theorem 3 to derive the expected value of the achieved estimate.

Proof: Without loss of generality, for the n given observations of θ , we denote the k invalid observations as Y_1, Y_2, \dots, Y_k , with $Y_j \sim N(0, \sigma^2)$, $j \in \{1, \dots, k\}$, and the $n - k$ valid observations as $Y_{k+1}, Y_{k+2}, \dots, Y_n$, with $Y_j \sim N(\theta, \sigma^2)$, $j \in \{k+1, \dots, n\}$.

We first prove $\text{sgn}(\hat{y}_i) \xrightarrow{P} \text{sgn}(\theta)$ (where \xrightarrow{P} denotes convergence in probability), $\forall i$, as $\text{SNR} \rightarrow \infty$, where sgn is a function such that $\text{sgn}(x) = +$ when $x \geq 0$ and $\text{sgn}(x) = -$ when $x < 0$. Since $\hat{y}_i \xrightarrow{P} \bar{y}$, $\forall i$ [44], it is enough to show that $\text{sgn}(\bar{y}) \xrightarrow{P} \text{sgn}(\theta)$. The mean of y_i could be expressed as,

$$\bar{y} = \frac{\sum_i y_i}{n} = \frac{k}{n}\theta + \frac{\sum_i w_i}{n}. \quad (36)$$

Since w_i 's are i.i.d. Gaussian white noises with zero mean and variance σ^2 , $\frac{\sum_i w_i}{n}$ is Gaussian random variable with zero mean and variance σ^2/n . Thus, the error probability is given as,

$$\Pr\{\text{sgn}(\bar{y}) \neq \text{sgn}(\theta)\} = Q\left(\frac{\frac{k}{n}\theta}{\frac{\sigma}{\sqrt{n}}}\right) < \frac{1}{2}e^{-\frac{k^2\theta^2}{2\sigma^2 n}}. \quad (37)$$

Thus, $\text{sgn}(\bar{y}) \xrightarrow{P} \text{sgn}(\theta)$, as $\text{SNR} \rightarrow \infty$.

Next, we prove that $E(\hat{\theta}^+) \xrightarrow{P} \theta$ (for the case of $\hat{\theta}^-$, the proof is similar and skipped). Define $\hat{\theta}_c = \frac{\sum_{i=k+1}^n Y_i}{n-k}$. Thus, $E(\hat{\theta}_c) = \theta$. Define the probability of successful estimate as, $P_c^+ = \Pr\{\hat{\theta}^+ = \hat{\theta}_c\}$.

In the following part, we prove that $P_c^+ \rightarrow 1$, as $\text{SNR} \rightarrow \infty$ for both of the two cases: $p_1 \geq 0.5$ and $p_1 < 0.5$. When $p \geq 0.5$, P_c^+ can be expressed with the boundaries of the decision regions:

$$\begin{aligned} P_c^+ &= \Pr \left\{ \max_{j \in \{1, \dots, k\}} \left(Y_j + \sqrt{Y_j^2 + 2\sigma^2 \ln \frac{p_1}{p_0}} \right) \right. \\ &\leq \frac{\sum_{i=k+1}^n Y_i}{n-k} \leq \left. \min_{j \in \{k+1, \dots, n\}} \left(Y_j + \sqrt{Y_j^2 + 2\sigma^2 \ln \frac{p_1}{p_0}} \right) \right\}. \end{aligned} \quad (38)$$

Thus, the union bound of the probability of error, $P_e^+ = 1 - P_c^+$, could be expressed as

$$\begin{aligned} P_e \leq & \Pr \left\{ \min_{j \in \{k+1, \dots, n\}} \left(Y_j + \sqrt{Y_j^2 + 2\sigma^2 \ln \frac{p_1}{p_0}} \right) \leq \frac{\sum_{i=k+1}^n Y_i}{n-k} \right\} \\ & + \Pr \left\{ \frac{\sum_{i=k+1}^n Y_i}{n-k} \leq \max_{j \in \{1, \dots, k\}} \left(Y_j + \sqrt{Y_j^2 + 2\sigma^2 \ln \frac{p_1}{p_0}} \right) \right\} \end{aligned} \quad (39)$$

where both of the above two items go to 0 as $\text{SNR} \rightarrow \infty$.

The proof for the case of $p_1 < 0.5$ is similar. Therefore, we conclude that $\lim_{\text{SNR} \rightarrow \infty} E(\hat{\theta}^+) = E(\hat{\theta}_c) = \theta$. Similarly we could have $\lim_{\text{SNR} \rightarrow \infty} E(\hat{\theta}^-) = E(\hat{\theta}_c) = \theta$. Together with the result in the first part for $\text{sgn}(\bar{y}) \xrightarrow{P} \text{sgn}(\theta)$, as $\text{SNR} \rightarrow \infty$, we have $\lim_{\text{SNR} \rightarrow \infty} E(\hat{\theta}) = \theta$. ■

C. Variance of $\hat{\theta}$

In this subsection, we derive the variance of $\hat{\theta}$. We have ordered the observations as $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$, and we define the corresponding random variables as $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$.

Conditioned on \mathbf{h} , the variance of $\hat{\theta}$ can be derived as

$$\text{Var}(\hat{\theta}) = E(\text{Var}(\hat{\theta} | \mathbf{h})) + \text{Var}(E(\hat{\theta} | \mathbf{h})). \quad (40)$$

The first term on the right-hand side of (40) can be expressed as,

$$\begin{aligned} E(\text{Var}(\hat{\theta} | \mathbf{h})) &= \sum_{k=1}^n \text{Var}(\hat{\theta}^{k+}) \Pr\{\mathbf{h} = \mathbf{h}_k^+\} \\ &+ \sum_{k=1}^n \text{Var}(\hat{\theta}^{k-}) \Pr\{\mathbf{h} = \mathbf{h}_k^-\}, \end{aligned} \quad (41)$$

where $\hat{\theta}^{k+}$ and $\hat{\theta}^{k-}$ are the estimates when $\mathbf{h} = \mathbf{h}_k^+$ and $\mathbf{h} = \mathbf{h}_k^-$, respectively, i.e.,

$$\hat{\theta}^{k+} = \frac{\sum_i \hat{h}_i^+ Y_i}{\sum_i \hat{h}_i^+} = \frac{\sum_{i=k}^n Y_{(i)}}{n-k+1}, \quad (42)$$

$$\hat{\theta}^{k-} = \frac{\sum_i \hat{h}_i^- Y_i}{\sum_i \hat{h}_i^-} = \frac{\sum_{i=1}^{n-k+1} Y_{(i)}}{n-k+1}, \quad (43)$$

and the variances of $\hat{\theta}^{k+}$ and $\hat{\theta}^{k-}$ can be expressed as,

$$\text{Var}(\hat{\theta}^{k+}) = \text{Var} \left(\frac{\sum_{i=k}^n Y_{(i)}}{n-k+1} \right) = \frac{\sum_{i=k}^n \sigma_{(i)}^2}{(n-k+1)^2},$$

$$\text{Var}(\hat{\theta}^{k-}) = \text{Var} \left(\frac{\sum_{i=1}^{n-k+1} Y_{(i)}}{n-k+1} \right) = \frac{\sum_{i=1}^{n-k+1} \sigma_{(i)}^2}{(n-k+1)^2},$$

where $\sigma_{(i)}^2$ is the variance of $Y_{(i)}$, which will be derived in the next subsection.

The second term on the right-hand side of (40) can be expressed as,

$$\begin{aligned} \text{Var}(E(\hat{\theta} | \mathbf{h})) &= E((E(\hat{\theta} | \mathbf{h}))^2) - E^2(E(\hat{\theta} | \mathbf{h})) \\ &= \left[\sum_{k=1}^n E^2(\hat{\theta}^{k+}) \Pr\{\mathbf{h} = \mathbf{h}_k^+\} + \sum_{k=1}^n E^2(\hat{\theta}^{k-}) \Pr\{\mathbf{h} = \mathbf{h}_k^-\} \right] \\ &- \left[\sum_{k=1}^n E(\hat{\theta}^{k+}) \Pr\{\mathbf{h} = \mathbf{h}_k^+\} + \sum_{k=1}^n E(\hat{\theta}^{k-}) \Pr\{\mathbf{h} = \mathbf{h}_k^-\} \right]^2, \end{aligned} \quad (44)$$

where the expectation of $\hat{\theta}^{k(\cdot)}$ can be derived as,

$$E(\hat{\theta}^{k+}) = \frac{E(\sum_{i=k}^n Y_{(i)})}{n-k+1} = \frac{\sum_{i=k}^n \mu_{(i)}}{n-k+1}, \quad (45)$$

with $\mu_{(i)}$ as the mean of $Y_{(i)}$, which will be derived in the next subsection. For the negative part, similarly, we have

$$E(\hat{\theta}^{k-}) = \frac{E(\sum_{i=1}^{n-k+1} Y_{(i)})}{n-k+1} = \frac{\sum_{i=1}^{n-k+1} \mu_{(i)}}{n-k+1}. \quad (46)$$

From the above expressions, we see that both of the two terms on the right-hand side of (40) are constructed by three basic elements, i.e., $\Pr\{\mathbf{h} = \mathbf{h}_k^{(\cdot)}\}$'s, $\mu_{(i)}$'s, and $\sigma_{(i)}^2$'s. In the following subsections, we derive them by exploring the statistics of $Y_{(i)}$.

D. Statistics of $Y_{(i)}$

First, we start from the pdf of Y , where the received signal Y is a random variable, which is the sum of two independent random variables, i.e., $Y = h\theta + W$, where $\Pr(h\theta = \theta) = p_1$ and $\Pr(h\theta = 0) = p_0$, and W is an independent Gaussian random variable with zero mean and variance σ^2 . The pdf of Y can be expressed as

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}} p_0 + \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\theta)^2}{2\sigma^2}} p_1$$

and its cdf is expressed as

$$F_Y(y) = \Phi\left(\frac{y}{\sigma}\right) \cdot p_0 + \Phi\left(\frac{y-\theta}{\sigma}\right) \cdot p_1,$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$.

Next, we derive the cdf of the ordered received signals Y_j 's. The cdf of $Y_{(i)}$ can then be expressed as

$$\begin{aligned} F_{Y_{(i)}}(r) &= \Pr\{Y_{(i)} < r\} \\ &= \Pr\{\text{the number of } Y_j \text{ less than or equal to } r \text{ is at least } i\} \\ &= \sum_{k=i}^n \binom{n}{k} F_Y^k(r) [1 - F_Y(r)]^{n-k}. \end{aligned} \quad (47)$$

The joint pdf of $Y_{(k_1)}, Y_{(k_2)}, \dots, Y_{(k_j)}$, ($1 \leq k_1 < k_2 < \dots < k_j \leq n$; $1 \leq j \leq n$), is, for $y_1 \leq y_2 \leq \dots \leq y_j$,

$$\begin{aligned} & f_{k_1 k_2 \dots k_j}(y_1, y_2, \dots, y_j) \\ &= \frac{n! \cdot F_Y^{k_1-1}(y_1) f_Y(y_1) [F_Y(y_2) - F_Y(y_1)]^{k_2-k_1-1} f_Y(y_2)}{(k_1-1)!(k_2-k_1-1)! \dots (n-k_j)!} \\ & \quad \times \dots [1 - F_Y(y_k)]^{n-k_j} f_Y(y_j) \end{aligned} \quad (48)$$

By the result in [45], the mean of $Y_{(i)}$ can be calculated as

$$\begin{aligned} \mu_{(i)} &= n \binom{n-1}{i-1} \int_{-\infty}^{\infty} x [F_Y(x)]^{i-1} [1 - F_Y(x)]^{n-i} f_Y(x) dx \\ &= n \binom{n-1}{i-1} \int_0^1 F_Y^{-1}(u) u^{i-1} (1-u)^{n-i} du, \end{aligned} \quad (49)$$

and the variance of $Y_{(i)}$ is given as

$$\sigma_{(i)}^2 = E((Y_{(i)})^2) - \mu_{(i)}^2. \quad (50)$$

E. Probability of $\mathbf{h} = \mathbf{h}_k^{(\cdot)}$

Next, we derive the probability that \mathbf{h} equals \mathbf{h}_k^+ . We have

$$\begin{aligned} \Pr\{\mathbf{h} = \mathbf{h}_k^+\} &= \Pr\left\{\hat{\theta} \in \mathcal{D}_{(i)}, i = k, k+1, \dots, n; \right. \\ & \quad \left. \hat{\theta} \notin \mathcal{D}_{(j)}, j = 1, 2, \dots, k-1; \text{sgn}\left(\sum Y_i\right) = +\right\}. \end{aligned} \quad (51)$$

Specifically, when $p_1 \geq 0.5$, we have

$$\begin{aligned} \Pr\{\mathbf{h} = \mathbf{h}_k^+\} &= \\ \Pr\left\{Y_{(k-1)} + \sqrt{Y_{(k-1)}^2 + 2\sigma^2 \ln \frac{p_1}{p_0}} \leq \frac{\sum_{i=k}^n Y_{(i)}}{n-k+1} \right. \\ & \quad \left. \leq Y_{(k)} + \sqrt{Y_{(k)}^2 + 2\sigma^2 \ln \frac{p_1}{p_0}}; \text{sgn}\left(\sum Y_i\right) = +\right\}. \end{aligned} \quad (52)$$

When $p_1 < 0.5$, we have

$$\begin{aligned} \Pr\{\mathbf{h} = \mathbf{h}_k^+\} &= \\ &= \Pr\left\{Y_{(k-1)} + \sqrt{Y_{(k-1)}^2 + 2\sigma^2 \ln \frac{p_1}{p_0}} \leq \frac{\sum_{i=k}^n Y_{(i)}}{n-k+1} \right. \\ & \quad \leq Y_{(k)} + \sqrt{Y_{(k)}^2 + 2\sigma^2 \ln \frac{p_1}{p_0}}; \\ & \quad \left. Y_{(k-1)} \geq -\sqrt{-2\sigma^2 \ln \frac{p_1}{p_0}}; \text{sgn}\left(\sum Y_i\right) = +\right\} \\ &+ \Pr\left\{\frac{\sum_{i=k}^n Y_{(i)}}{n-k+1} \leq Y_{(k)} + \sqrt{Y_{(k)}^2 + 2\sigma^2 \ln \frac{p_1}{p_0}}; \right. \\ & \quad \left. Y_{(k)} \geq -\sqrt{-2\sigma^2 \ln \frac{p_1}{p_0}}; \right. \\ & \quad \left. Y_{(k-1)} < -\sqrt{-2\sigma^2 \ln \frac{p_1}{p_0}}; \text{sgn}\left(\sum Y_i\right) = +\right\} \end{aligned} \quad (53)$$

$$\begin{aligned} &+ \Pr\left\{Y_{(k-1)} - \sqrt{Y_{(k-1)}^2 + 2\sigma^2 \ln \frac{p_1}{p_0}} \geq \frac{\sum_{i=k}^n Y_{(i)}}{n-k+1} \right. \\ & \quad \geq Y_{(k)} - \sqrt{Y_{(k)}^2 + 2\sigma^2 \ln \frac{p_1}{p_0}}; \\ & \quad \left. Y_{(k-1)} \geq -\sqrt{-2\sigma^2 \ln \frac{p_1}{p_0}}; \text{sgn}\left(\sum Y_i\right) = +\right\} \\ &+ \Pr\left\{\frac{\sum_{i=k}^n Y_{(i)}}{n-k+1} \geq Y_{(k)} - \sqrt{Y_{(k)}^2 + 2\sigma^2 \ln \frac{p_1}{p_0}}; \right. \\ & \quad \left. Y_{(k)} \geq -\sqrt{-2\sigma^2 \ln \frac{p_1}{p_0}}; \right. \\ & \quad \left. Y_{(k-1)} < -\sqrt{-2\sigma^2 \ln \frac{p_1}{p_0}}; \text{sgn}\left(\sum Y_i\right) = +\right\}. \end{aligned} \quad (54)$$

The expression for the negative case of \mathbf{h}_k^- is similar, which is omitted here. So far, all the terms in (40) have been calculated. Thus, the closed-form variance could be derived. However, this expression is too complicated to make any intuitive observations. In the next section, we analyze the asymptotic performance of the proposed algorithm, which could lead to some compact and intuitive observations.

VIII. ASYMPTOTIC ANALYSIS

In the previous section, we studied the mean and variance of the limiting value with the proposed algorithm. In this section, we study the asymptotic performance of the proposed algorithm as $n \rightarrow \infty$. We first review the asymptotic theory of order statistics, then we study the asymptotic result of the given estimator. Afterwards, we show that $\text{Var}(\hat{\theta})$ is of the same order as $\text{Var}(\hat{\theta}_{\text{Ideal}})$ when n tends to infinity.

In the asymptotic theory of order statistics [45], the limiting distributions of appropriately standardized sequences of k th order statistics $\{X_{(k)}\}$ as the number of samples n tends to infinity are studied. Generally, the order number k can change as a function of n . If $\lim_{n \rightarrow \infty} k/n$ exists between 0 and 1, but not equal to 0 or 1, the corresponding order statistics $X_{(k)}$ of the sequence $\{X_{(k)}\}$ are called the central order statistics. Otherwise, they are called the extreme order statistics.

In mathematical statistics, central order statistics are used to construct consistent sequences of estimators for quantiles of the unknown distribution $F(u)$ based on the realization of a random vector X . For instance, let x_q be a quantile at level q , ($0 < q < 1$), of the distribution function $F(u)$ with a continuous probability density $f(u)$ and strictly positive in some neighborhood of the point x_q . As such, the sequence of central order statistics $\{X_{(k)}\}$ with order numbers $k = \lceil nq \rceil$, where $\lceil \cdot \rceil$ is the ceiling function, is a sequence of consistent estimators for the quantiles x_q , as $n \rightarrow \infty$ [45].

For a general distribution F with a continuous non-zero density at $F^{-1}(q)$, the q -th sample quantile is asymptotically normally distributed as n tends to infinity, and is approximated by

$$\lim_{n \rightarrow \infty} F_{X_{(\lceil nq \rceil)}}(x) = F_{X_{n,q}}(x), \quad (55)$$

where $X_{n,q} \sim N\left(F^{-1}(q), \frac{q(1-q)}{n[f(F^{-1}(q))]^2}\right)$ [45].

In (42) and (43), we defined $\hat{\theta}^{k(\cdot)}$ when n is finite. Next, we derive the limiting value of $\hat{\theta}^{\lceil nq \rceil(\cdot)}$ when $n \rightarrow \infty$.

Theorem 9: If F_Y is a continuous function, for any $0 < q < 1$ and $\varepsilon > 0$, we have

$$\lim_{n \rightarrow \infty} \Pr \left\{ \left| \hat{\theta}^{\lceil nq \rceil+} - \frac{\int_{F_Y^{-1}(q)}^{+\infty} y f_Y(y) dy}{\int_{F_Y^{-1}(q)}^{+\infty} f_Y(y) dy} \right| \geq \varepsilon \right\} = 0,$$

$$\lim_{n \rightarrow \infty} \Pr \left\{ \left| \hat{\theta}^{\lceil nq \rceil-} - \frac{\int_{-\infty}^{F_Y^{-1}(q)} y f_Y(y) dy}{\int_{-\infty}^{F_Y^{-1}(q)} f_Y(y) dy} \right| \geq \varepsilon \right\} = 0.$$

Proof: Here, we prove the positive part, while the proof of the negative part is similar. By definition, the cdf of $\hat{\theta}^{\lceil nq \rceil+}$ can be expressed as

$$F_{\hat{\theta}^{\lceil nq \rceil+}}(r) = \Pr \left\{ \frac{\sum_{i=\lceil nq \rceil}^n Y_{(i)}}{n - \lceil nq \rceil + 1} < r \right\}. \quad (56)$$

Since $\{Y_{(i)}\}$ is the ordered version of $\{Y_i\}$, we have

$$\Pr \left\{ \frac{\sum_{i=\lceil nq \rceil}^n Y_{(i)}}{n - \lceil nq \rceil + 1} < r \right\} = \Pr \left\{ \frac{\sum_{j \in \Omega_{n,q}} Y_j}{n - \lceil nq \rceil + 1} < r \right\}, \quad (57)$$

where $\Omega_{n,q} = \{j : Y_j \geq Y_{(\lceil nq \rceil)}, j \in \{1, 2, \dots, n\}\}$.

By (55), we have $\lim_{n \rightarrow \infty} F_{Y_{(\lceil nq \rceil)}}(y) = F_{Y_{n,q}}(y)$, where $Y_{n,q} \sim N\left(F^{-1}(q), \frac{q(1-q)}{n[f(F^{-1}(q))]^2}\right)$. Thus, we have

$$\lim_{n \rightarrow \infty} \Pr\{|Y_{(\lceil nq \rceil)} - F^{-1}(q)| \geq \varepsilon\} = 0. \quad (58)$$

Since Y_j 's are i.i.d. random variables, we have

$$\lim_{n \rightarrow \infty} \Pr \left\{ \left| \Pr \left\{ \frac{\sum_{j \in \Omega_{n,q}} Y_j}{n - \lceil nq \rceil + 1} < r \right\} - \Pr \left\{ \frac{\sum_{j \in \Omega_q} Y_j}{n - \lceil nq \rceil + 1} < r \right\} \right| \geq \varepsilon \right\} = 0, \quad (59)$$

where $\Omega_q = \{j : Y_j \geq F^{-1}(q), j \in \{1, 2, \dots, n\}\}$.

Also, since we only consider the random variables Y_j 's where the index is in Ω_q , the cdf of Y_j can be derived from the cdf of Y with a normalization factor $\int_{F_Y^{-1}(q)}^{+\infty} f_Y(y) dy$ as

$$F_{Y_j; Y_j > F_Y^{-1}(q)}(r) = \frac{f_Y(r)}{\int_{F_Y^{-1}(q)}^{+\infty} f_Y(y) dy}, \quad r > F_Y^{-1}(q). \quad (60)$$

Thus we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\sum_{j \in \Omega_q} Y_j}{n - \lceil nq \rceil + 1} &= E(Y_j | j \in \Omega_q) \\ &= \int_{F_Y^{-1}(q)}^{+\infty} r \frac{f_Y(r)}{\int_{F_Y^{-1}(q)}^{+\infty} f_Y(y) dy} dr \\ &= \frac{\int_{F_Y^{-1}(q)}^{+\infty} y f_Y(y) dy}{\int_{F_Y^{-1}(q)}^{+\infty} f_Y(y) dy}, \end{aligned} \quad (61)$$

which is a constant.

Combining the results in (59) and (61), together with the definition of cdf, we obtain the desired result. ■

Next, we prove Theorem 4.

Proof: Without loss of generality, for the n given observations of a positive θ (for the case of negative θ , the proof is

similar), we denote the k invalid observations as Y_1, Y_2, \dots, Y_k , with $Y_j \sim N(0, \sigma^2)$, $j \in \{1, \dots, k\}$, and the $n - k$ valid observations as $Y_{k+1}, Y_{k+2}, \dots, Y_n$, with $Y_j \sim N(\theta, \sigma^2)$, $j \in \{k+1, \dots, n\}$.

Conditioned on \mathbf{h} , the variance of $\hat{\theta}$ can be derived as

$$\text{Var}(\hat{\theta}) = E(\text{Var}(\hat{\theta} | \mathbf{h})) + \text{Var}(E(\hat{\theta} | \mathbf{h})). \quad (62)$$

According to (41), the first term on the right-hand side of (62) can be expressed as,

$$\begin{aligned} E(\text{Var}(\hat{\theta} | \mathbf{h})) &= \sum_{i=1}^n \text{Var}(\hat{\theta}^{i+}) \Pr\{\mathbf{h} = \mathbf{h}_i^+\} \\ &\quad + \sum_{i=1}^n \text{Var}(\hat{\theta}^{i-}) \Pr\{\mathbf{h} = \mathbf{h}_i^-\}, \end{aligned} \quad (63)$$

where

$$\text{Var}(\hat{\theta}^{i+}) = \frac{\sum_{j=i}^n \sigma_{(j)}^2}{(n - i + 1)^2}, \quad (64)$$

$$\text{Var}(\hat{\theta}^{i-}) = \frac{\sum_{j=1}^{n-i+1} \sigma_{(j)}^2}{(n - i + 1)^2}, \quad (65)$$

with $\sigma_{(j)}^2$ as the variance of $Y_{(j)}$, which converges to $\frac{\frac{i}{n}(1-\frac{i}{n})}{n[f(F^{-1}(\frac{i}{n}))]^2}$ when n goes to infinity by (55).

According to (52), $\Pr\{\mathbf{h} = \mathbf{h}_i^+\}$ is exponentially decreasing over SNR when $i \neq k$, due to the Gaussian assumption. Similarly, we also have that $\Pr\{\mathbf{h} = \mathbf{h}_i^-\}$ is exponentially decreasing over SNR. By combining (64) and (65) with (55), we have that the linear rate of $\text{Var}(\hat{\theta}^{i(\cdot)})$ changing over SNR is lower than the exponential rate of $\Pr\{\mathbf{h} = \mathbf{h}_i^{(\cdot)}\}$ decreasing over SNR when $\mathbf{h} \neq \mathbf{h}_k^+$. Thus, only the terms with $\Pr\{\mathbf{h} = \mathbf{h}_k^+\}$ are left in (63) as $\text{SNR} \rightarrow \infty$ and we have $E(\text{Var}(\hat{\theta} | \mathbf{h})) \rightarrow E(\text{Var}(\hat{\theta}_{\text{Ideal}} | \mathbf{h}))$ almost surely by the definition of $\hat{\theta}_{\text{Ideal}}$.

The second term on the right-hand side of (62) can be expressed as

$$\begin{aligned} \text{Var}(E(\hat{\theta} | \mathbf{h})) &= E((E(\hat{\theta} | \mathbf{h}))^2) - E^2(E(\hat{\theta} | \mathbf{h})) \\ &= \sum_{i=1}^n E^2(\hat{\theta}^{i+} | \mathbf{h} = \mathbf{h}_i^+) \Pr\{\mathbf{h} = \mathbf{h}_i^+\} \\ &\quad + \sum_{i=1}^n E^2(\hat{\theta}^{i-} | \mathbf{h} = \mathbf{h}_i^-) \Pr\{\mathbf{h} = \mathbf{h}_i^-\} \\ &\quad - \left[\sum_{i=1}^n E(\hat{\theta}^{i+} | \mathbf{h} = \mathbf{h}_i^+) \Pr\{\mathbf{h} = \mathbf{h}_i^+\} \right. \\ &\quad \left. + \sum_{i=1}^n E(\hat{\theta}^{i-} | \mathbf{h} = \mathbf{h}_i^-) \Pr\{\mathbf{h} = \mathbf{h}_i^-\} \right]^2 \end{aligned} \quad (66)$$

where

$$E(\hat{\theta}^{i+} | \mathbf{h} = \mathbf{h}_i^+) = \frac{\sum_{j=i}^n \mu_{(j)}}{n - i + 1}, \quad (67)$$

$$E(\hat{\theta}^{i-} | \mathbf{h} = \mathbf{h}_i^-) = \frac{\sum_{j=1}^{n-i+1} \mu_{(j)}}{n - i + 1}. \quad (68)$$

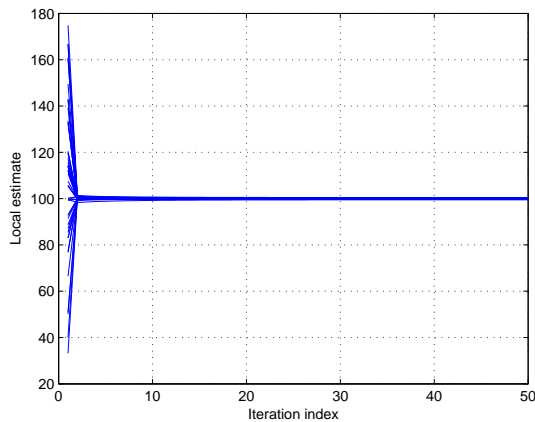


Fig. 1. The convergence of the MDE algorithm, $\theta = 100$.

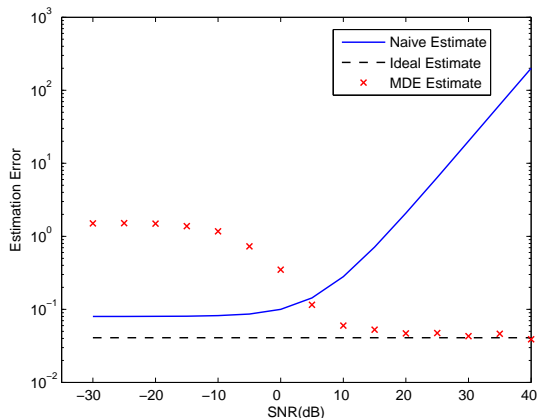


Fig. 2. The performance comparison among the MDE algorithm, the naive averaging algorithm, and the ideal estimate.

with $\mu_{(j)}$ as the mean of $Y_{(j)}$, which converges to $F^{-1}(\frac{j}{n})$ when n goes to infinity by (55).

According to (52), $\Pr\{\mathbf{h} = \mathbf{h}_k^{(\cdot)}\}$ is exponentially decreasing over SNR when $\mathbf{h} \neq \mathbf{h}_k^+$, due to the Gaussian assumption. By combining (67) and (68) with (55), we have that the linear rate of $E(\hat{\theta}^{i(\cdot)} | \mathbf{h} = \mathbf{h}_i^{(\cdot)})$ changing over SNR is lower than the exponential rate of $\Pr\{\mathbf{h} = \mathbf{h}_i^{(\cdot)}\}$ decreasing over SNR when $\mathbf{h} \neq \mathbf{h}_k^+$. Thus, only the terms with $\Pr\{\mathbf{h} = \mathbf{h}_k^+\}$ are left in (66) as $\text{SNR} \rightarrow \infty$ and we have $\text{Var}(E(\hat{\theta} | \mathbf{h})) \rightarrow \text{Var}(E(\hat{\theta}_{\text{Ideal}} | \mathbf{h}))$ almost surely by the definition of $\hat{\theta}_{\text{Ideal}}$.

Combining the results in the above two parts, we have $\text{Var}(\hat{\theta}) \rightarrow \text{Var}(\hat{\theta}_{\text{Ideal}})$ almost surely. ■

IX. SIMULATION RESULTS

In this section, we present simulation results that demonstrate the estimation performance of the proposed MDE algorithm. In our network setting, 50 nodes are uniformly distributed over a unit square where two nodes are connected by an edge if their distance is less than 0.3, which is the predefined transmission range. In addition, h_i 's are independently generated with $p_1 = 0.5$, w_i 's are independent white Gaussian noises with zero mean and unit variance, and the other parameter values are specified in the description of each figure.

In Fig. 1, we demonstrate the convergence (Theorem 1) of the proposed algorithm. Realizations of the local esti-

mates at the 50 nodes over 50 rounds of iterations, i.e., $\hat{\theta}_i(t), i \in [1, \dots, 50], t \in [1, \dots, 50]$, are plotted. The target θ is 100, which implies $\text{SNR} = 40\text{dB}$. In the figure, about half of the nodes start around the value 100 and the rest start around 0, indicating that the former ones correspond to valid observations and the latter ones are the nodes with invalid observations. We observe that the local estimates of both types of nodes converge as the number of iteration increases.

In Fig. 2, we compare the performance of the proposed MDE algorithm with the naive averaging algorithm (2) and the ideal algorithm (4) discussed in Section II. In the figure, the estimation error of these three estimates are plotted with SNR ranging from -30 dB to 40 dB. For each SNR, we generate 500 runs of the MDE algorithm, with the limiting consensus value of the local estimate for each realization being taken to be the estimate in the first node at the end of the 3000-th iteration. The estimation error plotted in the figure is the average squared deviation of the limiting consensus value from the true value of θ over these 500 realizations, i.e., $(\sum (\hat{\theta}_1(3000) - \theta)^2) / 500$. The topology of the communication graph (given by the random node placement) and the observation values across the nodes are independently generated for each realization. We make several observations from this figure. First, the numerical result of the naive averaging algorithm (2) matches the theoretical results as derived in (3), i.e., the estimation error variance grows exponentially over SNR; second, the numerical result of the ideal algorithm (4) matches the theoretical results as derived in (5), where the estimation error is the lowest among the three algorithms; and third, although the estimation error of MDE is higher than that of the naive averaging in the lower SNR regime ($\text{SNR} < 5\text{dB}$), it performs much better in the mid and high SNR regimes ($\text{SNR} > 20\text{dB}$), where it approaches the performance of the ideal estimator.

In the following we provide some intuitive explanation of the observed simulation behavior: 1) In the low SNR regime, the target value is relatively small as compared with the Gaussian noise, which leads to a high detection error in (7) and (8). Some invalid observations are wrongly detected as valid ones and negatively incorporated into the estimate update process, whereas, some valid observations are discarded as invalid ones. Thus, the estimate is largely distorted from the ideal estimate, which leads to the poor estimation performance; 2) in the high SNR regime, the detection error in (7) and (8) is very small and almost every observation is correctly detected as valid or invalid. Therefore, the MDE estimate is quite close to the ideal estimate and the MSE of the MDE algorithm approaches the lower bound (i.e., that achieved by the ideal algorithm).

X. CONCLUSIONS

We studied an algorithm named MDE, for distributed estimation of a scalar target signal with imperfect sensing mode information (due to node defects) in a sensor network. For the proposed algorithm, an online learning step assesses the validity of the local observations at each iteration, and then refines the ongoing estimation update process in an iterative fashion. We analytically established the convergence of the MDE algorithm. From the asymptotic results of the

performance analysis, we have shown that in the high SNR regime, as the number of nodes goes to infinity, the MDE estimation error converges to that of an ideal estimator with perfect information about the node sensing modes.

APPENDIX A
VARIANCE OF IDEAL ESTIMATOR IN (5)

We have the first entry of the conditional variance calculated as

$$\begin{aligned} \mathbb{E}(\text{Var}(\hat{\theta}_{\text{Ideal}}^n | \mathbf{h})) &= \sum_{\mathbf{h}} \text{Var}(\hat{\theta}_{\text{Ideal}}^n | \mathbf{h}) p(\mathbf{h}) \\ &= \sum_{k=0}^n \text{Var} \left(\frac{\sum_{i:h_i=1} y_i}{k} \right) \Pr \left\{ \sum h_i = k \right\} \\ &= \sum_{k=0}^n \text{Var} \left(\theta + \frac{\sum_{i:h_i=1} w_i}{k} \right) \Pr \left\{ \sum h_i = k \right\} \\ &= \sigma^2 \cdot \sum_{k=0}^n \frac{1}{k} \binom{n}{k} p_1^k p_0^{n-k}. \end{aligned}$$

We have the second entry calculated as

$$\text{Var}(\mathbb{E}(\hat{\theta}_{\text{Ideal}}^n | \mathbf{h})) = \text{Var}_{\mathbf{h}}(\theta) = 0,$$

where \mathbf{h} is given, and $\mathbb{E}(\hat{\theta}_{\text{Ideal}}^n | \mathbf{h}) = \mathbb{E} \left(\frac{\sum h_i y_i}{\sum h_i} | \mathbf{h} \right) = \theta$, which is a constant independent with \mathbf{h} . Thus we have derived the variance of ideal estimator shown in (5).

APPENDIX B

LEMMAS USED IN SECTION VI TO PROVE THEOREM 1

Lemma 10: If the moving status of $\max\{\hat{\theta}_{\text{current}}^+(t), 0\}$ is in Case 1 with $t = t_1$, then $\max\{\hat{\theta}_{\text{current}}^+(t), 0\}$ either converges without leaving Case 1 for all $t > t_1$, or the moving status of $\max\{\hat{\theta}_{\text{current}}^+(t), 0\}$ changes to Case 2 after $\tilde{t}_1 > t_1$.

Proof: If there is a time $\tilde{t}_1, \tilde{t}_1 > t_1$, such that the moving status of $\max\{\hat{\theta}_{\text{current}}^+(t), 0\}$ changes to Case 2, we have the desired result. Otherwise, for all $t > t_1$, we have that the moving status of $\max\{\hat{\theta}_{\text{current}}^+(t), 0\}$ stays in Case 1. In order to show the convergence, we only need to show that $\hat{\theta}_{\text{current}}^+(t)$ is a monotonic and bounded sequence. In this proof, we first prove that $\hat{\theta}_{\text{goal}}^+(t)$ converges when $t > t_1$. After that, we show the monotonicity of $\hat{\theta}_{\text{current}}^+(t)$. At last, we prove that $\hat{\theta}_{\text{current}}^+(t)$ is a bounded sequence for $t > t_1$.

Here we first prove that $\hat{\theta}_{\text{goal}}^+(t)$ converges when $t > t_1$. Since the moving status of $\max\{\hat{\theta}_{\text{current}}^+(t), 0\}$ stays in Case 1 for all $t > t_1$, and $\mathcal{G}^+(t)$ cannot jump to a different interval without touching any boundary by the smooth moving condition, we have that $\mathcal{G}^+(t)$ belongs to \mathcal{I}_j for all $t > t_1$. By the definition of $\mathcal{G}^+(t)$, we have $\theta_i^+(t) \in \mathcal{G}^+(t), \forall i$. Since $\mathcal{G}^+(t)$ belongs to the same \mathcal{I}_j for all $t > t_1$, the inclusion relationship of $\mathcal{G}^+(t)$ and \mathcal{D}_i 's do not change for all $t > t_1$. In other words, the detection results of $\hat{h}_i^+(t)$'s stay the same for all $t > t_1$. Specifically, if we replace $\hat{\theta}_i^+(t)$ with an arbitrary $x_j, \forall x_j \in \mathcal{I}_j$, the detection result of $\hat{h}_i^+(t)$ does not change in the detection step (step 2) for any i , i.e.,

$$x_j^2 - 2y_i x_j \begin{matrix} \hat{h}_i^+(t)=0 \\ \geq \\ \hat{h}_i^+(t)=1 \end{matrix} \geq 2\sigma^2 \ln \frac{p_1}{p_0}, \forall i, t > t_1. \quad (69)$$

Therefore, we have that $\hat{h}_i(t)$'s converge. Thus, we conclude that $\hat{\theta}_{\text{goal}}^+(t)$ converges by the definition. Meanwhile, in the proof, $\hat{h}_i^+(t)$'s and $\hat{\theta}_{\text{goal}}^+(t)$ are only related to the index of \mathcal{I}_j covering $\mathcal{G}^+(t)$. Thus, we define that $\hat{h}_i^+(t) = \hat{h}_i^+[j]$ and $\hat{\theta}_{\text{goal}}^+(t) = \hat{\theta}_{\text{goal}}^+[j]$ by using j , the index of \mathcal{I}_j .

Next, we show the monotonicity of $\hat{\theta}_{\text{current}}^+(t)$. To this end, we want to prove that

$$\begin{aligned} &(\hat{\theta}_{\text{goal}}^+[j] - \hat{\theta}_{\text{current}}^+(t+1)) \\ &(\hat{\theta}_{\text{current}}^+(t+1) - \hat{\theta}_{\text{current}}^+(t)) > 0, t > t_1. \end{aligned} \quad (70)$$

By taking average on both sides of (9), we have

$$\begin{aligned} \bar{u}^+(t) &= \bar{u}^+(t-1) + \alpha(t) (\{y_i \hat{h}_i^+(t)\}_{\text{avg}} - \bar{u}^+(t-1)) \\ &= (1 - \alpha(t)) \bar{u}^+(t-1) + \alpha(t) \{y_i \hat{h}_i^+(t)\}_{\text{avg}}. \end{aligned} \quad (71)$$

Similarly, by taking average on both sides of (10), we have

$$\bar{v}^+(t) = (1 - \alpha(t)) \bar{v}^+(t-1) + \alpha(t) \{\hat{h}_i^+(t)\}_{\text{avg}}, \quad (72)$$

which is a positive sequence.

Thus, we have

$$\begin{aligned} &(\hat{\theta}_{\text{goal}}^+[j] - \hat{\theta}_{\text{current}}^+(t+1)) (\hat{\theta}_{\text{current}}^+(t+1) - \hat{\theta}_{\text{current}}^+(t)) \\ &= \left(\frac{\{y_i \hat{h}_i^+[j]\}_{\text{avg}}}{\{\hat{h}_i^+[j]\}_{\text{avg}} + \delta} - \frac{\bar{u}^+(t)}{\bar{v}^+(t) + \delta} \right) \\ &\quad \left(\frac{\bar{u}^+(t)}{\bar{v}^+(t) + \delta} - \frac{\bar{u}^+(t-1)}{\bar{v}^+(t-1) + \delta} \right) \\ &= \frac{\alpha(t)(1 - \alpha(t)) \Upsilon^2}{(\bar{v}^+(t) + \delta)((1 - \alpha(t)) \bar{v}^+(t-1) + \alpha(t) \{\hat{h}_i^+[j]\}_{\text{avg}} + \delta)^2} \\ &\quad \frac{1}{(\{\hat{h}_i^+[j]\}_{\text{avg}} + \delta)} \end{aligned} \quad (73)$$

where $\Upsilon = \{y_i \hat{h}_i^+[j]\}_{\text{avg}} (\bar{v}^+(t-1) + \delta) - \bar{u}^+(t-1) (\{\hat{h}_i^+[j]\}_{\text{avg}} + \delta)$. Note that all of the elements multiplied together in (73) are positive.

At last, we prove that $\hat{\theta}_{\text{current}}^+(t)$ is a bounded sequence for $t > t_1$. Since both $\bar{u}^+(t)$ and $\bar{v}^+(t)$ are bounded by Lemma 5 and both $\bar{v}^+(t)$ and δ are positive, we conclude that $\hat{\theta}_{\text{current}}^+(t) = \frac{\bar{u}^+(t)}{\bar{v}^+(t) + \delta}$ is a bounded sequence. ■

Lemma 11: If the moving status of $\max\{\hat{\theta}_{\text{current}}^+(t), 0\}$ is in Case 2 when $t = t_2$, $\max\{\hat{\theta}_{\text{current}}^+(t), 0\}$ either converges without leaving Case 2 for all $t > t_2$, or $\exists \tilde{t}_2, \tilde{t}_2 > t_2$, such that the moving status of $\max\{\hat{\theta}_{\text{current}}^+(t), 0\}$ changes to Case 1 from \tilde{t}_2 .

Proof: If there is a time $\tilde{t}_2, \tilde{t}_2 > t_2$, such that the moving status of $\max\{\hat{\theta}_{\text{current}}^+(t), 0\}$ changes to Case 1, we have the desired result. Otherwise, for all $t > t_2$, we have that the moving status of $\max\{\hat{\theta}_{\text{current}}^+(t), 0\}$ stays in Case 2. Since $\max\{\hat{\theta}_{\text{current}}^+(t), 0\} \in \mathcal{G}^+(t), b_k \in \mathcal{G}^+(t)$, and $|\mathcal{G}^+(t)| = 2\varepsilon$, we have $|\hat{\theta}_{\text{current}}^+(t) - b_k| \leq 2\varepsilon$. Together with the fact that ε can be arbitrarily small, we conclude with convergence automatically. ■

Lemma 12: For the moving status of $\max\{\hat{\theta}_{\text{current}}^+(t), 0\}$, the number of switching times between Case 1 and Case 2 is finite.

Proof: First, we prove that after coming back to Case 1 from Case 2, the monotonicity of $\hat{\theta}_{\text{current}}^+(t)$ stays the same as

the one in the previous Case 1 (i.e., Case 1 before going into Case 2). Then, we prove that the sequence of $\{\hat{\theta}_{current}^+(t_s)\}$, which is a subsequence of $\{\hat{\theta}_{current}^+(t)\}$, is also monotonic, where we only consider $\{t_s\}$ at which the moving status is in Case 1. Together with the fact that the number of b_k 's is finite, lastly we conclude that the number of switching times between Case 1 and Case 2 is finite.

By following the above logic flow, we first prove that after coming back to Case 1 from Case 2, the monotonicity of $\hat{\theta}_{current}^+(t)$ stays the same as the one in the previous Case 1 (i.e., Case 1 before going into Case 2). To this end, since we have proven that $\hat{\theta}_{current}^+(t)$ changes monotonically in Case 1 by Lemma 10, it is sufficient to show that

$$(\hat{\theta}_{goal}^+(\acute{t}) - \hat{\theta}_{current}^+(\acute{t}))(\hat{\theta}_{goal}^+(\grave{t}) - \hat{\theta}_{current}^+(\grave{t})) \geq 0, \quad (74)$$

where \acute{t} is the time before going into Case 2 and \grave{t} is the time after coming out from Case 2. Assume that b_k under concern is one of the boundaries of node j , i.e., $b_k \in \{r_j^-, r_j^+\}$. Without loss of generality, we assume that b_k is r_j^- and comes into the gathering region from the right side. Therefore, we have $\hat{\theta}_{goal}^+(\acute{t}) = \hat{\theta}_{goal}^+[k]$, $\hat{\theta}_{goal}^+(\grave{t}) = \hat{\theta}_{goal}^+[k+1]$, and $\hat{\theta}_{goal}^+(\acute{t}) > \hat{\theta}_{current}^+(\acute{t})$. In order to prove (74), we only need to show that $\hat{\theta}_{goal}^+[k+1] > \hat{\theta}_{current}^+(\acute{t})$. For $\hat{\theta}_{goal}^+(\acute{t}-1)$, there are two possible values, i.e., $\hat{\theta}_{goal}^+[k]$ and $\hat{\theta}_{goal}^+[k+1]$. Specifically, if $\hat{\theta}_j^+(\acute{t})$ is on the right of b_k , we have $\hat{\theta}_{goal}^+(\acute{t}-1) = \hat{\theta}_{goal}^+[k+1]$; otherwise, we have $\hat{\theta}_{goal}^+(\acute{t}-1) = \hat{\theta}_{goal}^+[k]$. Next, we prove $\hat{\theta}_{goal}^+[k+1] > \hat{\theta}_{current}^+(\acute{t})$ for both cases.

1) When $\hat{\theta}_{goal}^+(\acute{t}-1) = \hat{\theta}_{goal}^+[k+1]$: By a similar derivation of (70), we have

$$\begin{aligned} & (\hat{\theta}_{goal}^+(\acute{t}-1) - \hat{\theta}_{current}^+(\acute{t})) \\ & (\hat{\theta}_{current}^+(\acute{t}) - \hat{\theta}_{current}^+(\acute{t}-1)) > 0. \end{aligned} \quad (75)$$

Since the second term on the left-hand side of (75) is positive by assumption, we have the first term on the left-hand side of (75) is also positive. Thus, we have $\hat{\theta}_{goal}^+[k+1] > \hat{\theta}_{current}^+(\acute{t})$ as desired.

2) When $\hat{\theta}_{goal}^+(\acute{t}-1) = \hat{\theta}_{goal}^+[k]$: By definition, $\hat{\theta}_{goal}^+[k+1]$ can be expressed as

$$\begin{aligned} \hat{\theta}_{goal}^+[k+1] &= \frac{\sum_i \hat{h}_i^+[k+1]y_i}{\sum_i \hat{h}_i^+[k+1] + n\delta} \\ &= \frac{\sum_i \hat{h}_i^+[k]y_i + y_j}{\sum_i \hat{h}_i^+[k] + 1 + n\delta} \\ &= \frac{\sum_i \hat{h}_i^+[k] + n\delta}{\sum_i \hat{h}_i^+[k] + 1 + n\delta} \frac{\sum_i \hat{h}_i^+[k]y_i}{\sum_i \hat{h}_i^+[k] + n\delta} \\ &\quad + \frac{y_j}{\sum_i \hat{h}_i^+[k] + 1 + n\delta} \\ &= \frac{\sum_i \hat{h}_i^+[k] + n\delta}{\sum_i \hat{h}_i^+[k] + 1 + n\delta} \hat{\theta}_{goal}^+[k] + \frac{y_j}{\sum_i \hat{h}_i^+[k] + 1 + n\delta}, \end{aligned} \quad (76)$$

where the sum of the weights on $\hat{\theta}_{goal}^+[k]$ and y_j equals to 1, i.e., $\frac{\sum_i \hat{h}_i^+[k] + n\delta}{\sum_i \hat{h}_i^+[k] + 1 + n\delta} + \frac{1}{\sum_i \hat{h}_i^+[k] + 1 + n\delta} = 1$. In order to

prove $\hat{\theta}_{goal}^+[k+1] > \hat{\theta}_{current}^+(\acute{t})$, we only need to show $\hat{\theta}_{goal}^+[k] > \hat{\theta}_{current}^+(\acute{t})$ and $y_j > \hat{\theta}_{current}^+(\acute{t})$. The first part can be proved by the result in (75) and $\hat{\theta}_{goal}^+(\acute{t}-1) = \hat{\theta}_{goal}^+[k]$. The second part is due to the smooth moving condition defined in Section VI-B, which implies that $y_j > b_k + 3\varepsilon$ and $\hat{\theta}_{current}^+(\acute{t}) \leq b_k + 2\varepsilon$.

Then, we prove that the sequence of $\{\hat{\theta}_{current}^+(t_s)\}$, which is a subsequence of $\{\hat{\theta}_{current}^+(t)\}$, is also monotonic, where we only consider $\{t_s\}$ at which the moving status is in Case 1. Specifically, we need to prove

$$(\hat{\theta}_{goal}^+(\acute{t}) - \hat{\theta}_{current}^+(\acute{t}))(\hat{\theta}_{current}^+(\acute{t}) - \hat{\theta}_{current}^+(\acute{t}-1)) \geq 0. \quad (77)$$

Assume that b_k is the boundary under concern in this visit of Case 2. Without loss of generality, we assume that b_k comes into the gathering region from the right side. Therefore, we have $\hat{\theta}_{goal}^+(\acute{t}) = \hat{\theta}_{goal}^+[k]$, $\hat{\theta}_{goal}^+(\grave{t}) = \hat{\theta}_{goal}^+[k+1]$, and $\hat{\theta}_{goal}^+(\acute{t}) > \hat{\theta}_{current}^+(\acute{t})$. In order to prove (77), we only need to show that $\hat{\theta}_{current}^+(\acute{t}) > \hat{\theta}_{current}^+(\acute{t}-1)$. Since the moving status of $\hat{\theta}_{current}^+(\acute{t}+1)$ and $\hat{\theta}_{current}^+(\acute{t}-1)$ is in Case 2, we have that both $\hat{\theta}_{current}^+(\acute{t}+1)$ and $\hat{\theta}_{current}^+(\acute{t}-1)$ are in the region of $[b_k - 2\varepsilon, b_k + 2\varepsilon]$. Together with the assumption that b_k comes into the gathering region from the right side, we have that $\hat{\theta}_{current}^+(\acute{t}) \leq b_k - 2\varepsilon$ and $\hat{\theta}_{current}^+(\acute{t}) \geq b_k + 2\varepsilon$. Hence, we have $\hat{\theta}_{current}^+(\acute{t}) > \hat{\theta}_{current}^+(\acute{t}-1)$ as desired.

So far, we have proved that the overall monotonicity of $\hat{\theta}_{current}^+(t)$ stays the same as when we only consider the iteration in Case 1, which means that the b_j 's for each visit of Case 2 are different. Together with the fact that the number of b_j 's is finite, we have that the number of switching from Case 1 to Case 2 is finite as desired. ■

Lemma 13: If the moving status of $\max\{\hat{\theta}_{current}^+(t), 0\}$ is in Case 1 when $t > t_1$ and $\max\{\hat{\theta}_{current}^+(t), 0\}$ converges without leaving Case 1 for all $t > t_1$, the limiting value is given by

$$\max \left\{ \frac{\sum_{i=1}^n \hat{h}_i^+[j]y_i}{\sum_{i=1}^n \hat{h}_i^+[j] + \delta n}, 0 \right\}, \quad (78)$$

where j is the index of \mathcal{I}_j , $\mathcal{G}^+(t_1) \subseteq \mathcal{I}_j$.

Proof: Since $\hat{\theta}_{current}^+(t) = \frac{\bar{u}^+(t)}{\bar{v}^+(t) + \delta}$, we only need to show that $\lim_{t \rightarrow \infty} \bar{u}^+(t) = \{\hat{h}_i^+[j]y_i\}_{avg}$ and $\lim_{t \rightarrow \infty} \bar{v}^+(t) = \{\hat{h}_i^+[j]\}_{avg}$. Here we only prove the part of $\bar{u}^+(t)$, while the proof for the part of $\bar{v}^+(t)$ is similar. By (23), we have

$$\begin{aligned} \bar{u}^+(t+1) &= \\ & (1 - \alpha(t+1))\bar{u}^+(t) + \alpha(t+1)\{y_i \hat{h}_i^+(t+1)\}_{avg}. \end{aligned} \quad (79)$$

Since the moving status of $\hat{\theta}_{current}^+(t)$ converges in Case 1 for all $t > t_1$, we have $\hat{h}_i^+(t) = \hat{h}_i^+[j]$, $\forall t > t_1$, with a similar derivation as (69). Thus, $\{\hat{h}_i^+(t)y_i\}_{avg}$ is a deterministic value when $t > t_1$ and equals $\{\hat{h}_i^+[j]y_i\}_{avg}$, $\forall t > t_1$.

Thus, we rewrite (79) as

$$\bar{u}^+(t+1) - K = [1 - \alpha(t+1)][\bar{u}^+(t) - K], \quad t > t_1, \quad (80)$$

where $K = \{\hat{h}_i^+[j]y_i\}_{avg}$. The limiting value of $\bar{u}^+(t) - K$ can be expressed as

$$\begin{aligned} \lim_{t \rightarrow \infty} |\bar{u}^+(t) - K| &= \left| \prod_{t=t_1}^{\infty} (1 - \alpha(t)) [\bar{u}^+(t_1) - K] \right| \\ &\leq \exp^{-\sum_{t=t_1}^{\infty} \alpha(t)} |\bar{u}^+(t_1) - K|. \end{aligned} \quad (81)$$

Since $\sum_t \alpha(t) = \infty$ and $\alpha(t) \in (0, 1)$, we have that the right-hand side of (81) equals 0. Then we conclude that $u(t)$ converges to K . ■

Lemma 14: If the moving status of $\max\{\hat{\theta}_{current}(t), 0\}$ is in Case 2 and $\max\{\hat{\theta}_{current}(t), 0\}$ converges without leaving Case 2 for all $t > t_2$ (according to the definition of Case 2, for certain node c , h_c^+ changes in Case 2), the limiting value is given by either

$$\max \left\{ \frac{\sum_{i=1}^n \hat{h}_i^+[j]y_i}{\sum_{i=1}^n \hat{h}_i^+[j] + \delta n}, 0 \right\}, \quad (82)$$

when $h_c^+[j] = 1$ and $h_c^+[j+1] = 0$, or

$$\max \left\{ \frac{\sum_{i=1}^n \hat{h}_i^+[j+1]y_i}{\sum_{i=1}^n \hat{h}_i^+[j+1] + \delta n}, 0 \right\}, \quad (83)$$

when $h_c^+[j] = 0$ and $h_c^+[j+1] = 1$, where j is the index of \mathcal{I}_j , $\mathcal{G}^+(t_2) \subseteq \mathcal{I}_j$.

Proof: Since the region of $\hat{\theta}_{current}^+(t_2)$ in Case 2 is $[b_k - 2\varepsilon, b_k + 2\varepsilon]$, $b_k \in \mathcal{G}^+(t_2)$, $\hat{\theta}_{current}^+(t)$ automatically converges if the moving status never changes to Case 1 for all $t > t_2$, as ε could be arbitrarily small. Thus, we only need to derive the limiting value in this proof.

There are only two possible limiting values implied by Lemma 13, i.e., $\hat{\theta}_{goal}^+[j]$ and $\hat{\theta}_{goal}^+[j+1]$. Without loss of generality, we assume that the limiting value is $\hat{\theta}_{goal}^+[j]$ given by (82), i.e., $\hat{\theta}_{goal}^+(t) = \hat{\theta}_{goal}^+[j]$, $\forall t > t'_2$, where t'_2 is a certain value greater than t_2 . If $\hat{\theta}_{goal}^+[j]$ stays in $[b_k - 2\varepsilon, b_k + 2\varepsilon]$, we come to the desired result.

Next, we prove that $\hat{\theta}_{goal}^+[j]$ falls in $[b_k - 2\varepsilon, b_k + 2\varepsilon]$ by contradiction. Without loss of generality, we assume $\hat{\theta}_{goal}^+[j] > b_k + 2\varepsilon$, and $\hat{\theta}_{current}^+(t)$ moves into $[b_k - 2\varepsilon, b_k + 2\varepsilon]$ from the left. By incorporating (23) into the definition of $\hat{\theta}_{current}^+(t)$, we have

$$\begin{aligned} \hat{\theta}_{current}^+(t'_2 + 1) &= \frac{\bar{u}^+(t'_2 + 1)}{\bar{v}^+(t'_2 + 1) + \delta} = \\ &= \frac{(1 - \alpha(t'_2 + 1))\bar{u}^+(t'_2) + \alpha(t'_2 + 1)\{y_i \hat{h}^+(t'_2 + 1)\}_{avg}}{(1 - \alpha(t'_2 + 1))[\bar{v}^+(t'_2) + \delta] + \alpha(t'_2 + 1)[\{\hat{h}^+(t'_2 + 1)\}_{avg} + \delta]}. \end{aligned}$$

Thus, the limiting value of $\hat{\theta}_{current}^+(t)$ can be expressed as

$$\begin{aligned} &\lim_{t \rightarrow \infty} \hat{\theta}_{current}^+(t) \\ &= \frac{\sum_{t=t'_2}^{\infty} \alpha(t)(1 - \alpha(t))^{t-t'_2} \{y_i \hat{h}^+(t)\}_{avg}}{\sum_{t=t'_2}^{\infty} \alpha(t)(1 - \alpha(t))^{t-t'_2} [\{\hat{h}^+(t)\}_{avg} + \delta]} \\ &> \frac{\sum_{t=t'_2}^{\infty} \alpha(t)(1 - \alpha(t))^{t-t'_2} [\{\hat{h}^+(t)\}_{avg} + \delta] (b_k + 2\varepsilon)}{\sum_{t=t'_2}^{\infty} \alpha(t)(1 - \alpha(t))^{t-t'_2} [\{\hat{h}^+(t)\}_{avg} + \delta]} \\ &= b_k + 2\varepsilon, \end{aligned} \quad (84)$$

which is in contradiction to the fact that $\hat{\theta}_{current}^+(t)$ is in $[b_k - 2\varepsilon, b_k + 2\varepsilon]$ for all $t > t_2$. Thus, we conclude that $\hat{\theta}_{goal}^+[j]$ stays in $[b_k - 2\varepsilon, b_k + 2\varepsilon]$. ■

REFERENCES

- [1] Q. Zhou, S. Kar, L. Huie, H. Poor, and S. Cui, "Robust distributed least-squares estimation in sensor networks with node failures," in *IEEE Global Telecommunications Conference (GLOBECOM)*, pp. 1-6, Houston, USA, Dec. 2011.
- [2] Q. Zhou, S. Kar, L. Huie, and S. Cui, "Distributed estimation in sensor networks with imperfect model information: An adaptive learning-based approach," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3109-3112, Kyoto, Mar. 2012.
- [3] J. N. Tsitsiklis, "Problems in Decentralized Decision Making and Computation," *Ph. D. Dissertation*, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 1984.
- [4] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and Cooperation in Networked Multi-Agent Systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215-233, Jan. 2007.
- [5] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip Algorithms for Distributed Signal Processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847-1864, Nov. 2010.
- [6] D. Li, S. Kar, J. M. F. Moura, H. V. Poor, and S. Cui, "Distributed Kalman Filtering over Massive Data Sets: Analysis Through Large Deviations of Random Riccati Equations," *IEEE Transactions on Information Theory*, vol. 61, no. 3, pp. 1351-1372, Mar. 2015.
- [7] D. Li, S. Kar, F. E. Alsaadi, A. M. Dobaie, and S. Cui, "Distributed Kalman Filtering with Quantized Sensing State," *IEEE Transactions on Signal Processing*, vol. 63, no. 19, pp. 5180-5193, Oct. 2015.
- [8] V. Borkar and P. Varaiya, "Asymptotic Agreement in Distributed Estimation," *IEEE Transactions on Automatic Control*, vol. 27, no. 3, pp. 650-655, June 1982.
- [9] M. H. deGroot, "Reaching a Consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118-121, Mar. 1974.
- [10] J. B. Leger and M. Kieffer, "Guaranteed Robust Distributed Estimation in a Network of Sensors," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 3378-3381, Dallas, TX, Mar. 2010.
- [11] S. Ramanan and J. M. Walsh, "Distributed Estimation of Channel Gains in Wireless Sensor Networks," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3097-3107, June 2010.
- [12] A. Bertrand and M. Moonen, "Distributed Adaptive Estimation of Node-Specific Signals in Wireless Sensor Networks With a Tree Topology," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2196-2210, May 2011.
- [13] G. Mateos and G. B. Giannakis, "Distributed Recursive Least-Squares: Stability and Performance Analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3740-3754, July 2012.
- [14] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the World with Wireless Sensor Networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2033-2036, Salt Lake City, UT, May 2001.
- [15] M. G. Rabbat and R. D. Nowak, "Quantized Incremental Algorithms for Distributed Optimization," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 798-808, Apr. 2007.
- [16] L. Li, J. A. Chambers, C. G. Lopes, and A. H. Sayed, "Distributed Estimation Over an Adaptive Incremental Network Based on the Affine Projection Algorithm," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 151-164, Jan. 2010.
- [17] L. Varagnolo, G. Pillonetto, and L. Schenato, "Distributed Statistical Estimation of the Number of Nodes in Sensor Networks," in the *49th IEEE Conference on Decision and Control*, pp. 1498-1503, Atlanta, GA, Dec. 2010.
- [18] C. Baquero, P. S. Almeida, R. Menezes, and P. Jesus, "Extrema Propagation: Fast Distributed Estimation of Sums and Network Sizes," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 4, pp. 668 - 675, Apr. 2012.
- [19] L. Xiao, and S. Boyd, "Distributed Average Consensus with Time-Varying Metropolis Weights," *Unpublished Manuscript*, http://www.stanford.edu/~boyd/papers/avg_metropolis.html, June 2006.
- [20] L. Xiao, and S. Boyd, "Fast Linear Iterations for Distributed Averaging," in the *42nd IEEE Conference on Decision and Control*, vol. 5, pp. 4997-5002, Maui, HI, Dec. 2003.
- [21] A. G. Dimakis, A. Sarwate, and M. J. Wainwright, "Geographic Gossip: Efficient Averaging for Sensor Networks," *IEEE Transactions on Signal Processing*, vol. 53, pp. 1205-1216, Mar. 2008.
- [22] R. Olfati-Saber and P. Jalalkamali, "Coupled Distributed Estimation and Control for Mobile Sensor Networks," *IEEE Transactions on Automatic Control*, vol. 57, no. 10, pp. 2609-2614, Oct. 2012.

- [23] W. Li and Y. Jia, "Consensus-Based Distributed Multiple Model UKF for Jump Markov Nonlinear Systems," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 227-233, Jan. 2012.
- [24] S. Kar and J. M. F. Moura, "Consensus + innovations distributed inference over networks: Cooperation and sensing in networked systems," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 99-109, May 2013.
- [25] S. Kar, J. M. F. Moura, and H. V. Poor, "QD-learning: A collaborative distributed strategy for multi-agent reinforcement learning through consensus + innovations," *IEEE Transactions on Signal Processing*, vol. 61, no. 7, pp. 1848-1862, Apr. 2013.
- [26] S. Kar and J. M. F. Moura, "Asymptotically Efficient Distributed Estimation With Exponential Family Statistics," *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4811-4831, Aug. 2014.
- [27] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3122-3136, Jul. 2008.
- [28] J. Chen and A. H. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4289-4305, Aug. 2012.
- [29] A. H. Sayed, S. Tu, J. Chen, X. Zhao, and Z. J. Towfic, "Diffusion strategies for adaptation and learning over networks: an examination of distributed strategies and network behavior," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 155-171, May 2013.
- [30] P. Braca, S. Marano, and V. Matta, "Enforcing consensus while monitoring the environment in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3375-3380, July 2008.
- [31] P. Braca, S. Marano, V. Matta, and P. Willett, "Asymptotic optimality of running consensus in testing binary hypotheses," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 814-825, Feb. 2010.
- [32] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS-based distributed detection over adaptive networks," in *IEEE 43rd Asilomar Conference on Signals, Systems and Computers*, pp. 1711-1715, Pacific Grove, CA, Nov. 2009.
- [33] F. S. Cattivelli and A. H. Sayed, "Distributed detection over adaptive networks based on diffusion estimation schemes," in *IEEE 10th Workshop on Signal Processing Advances in Wireless Communications*, pp. 6165, Perugia, June 2009.
- [34] F. S. Cattivelli and A. H. Sayed, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 1917-1932, May 2011.
- [35] D. Bajovic, D. Jakovetic, J. Xavier, B. Sinopoli, and J. M. Moura, "Distributed detection via gaussian running consensus: Large deviations asymptotic analysis," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4381-4396, Sep. 2011.
- [36] D. Jakovetic, J. M. Moura, and J. Xavier, "Distributed detection over noisy networks: Large deviations analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4306-4320, Aug. 2012.
- [37] S. Kar, R. Tandon, H. V. Poor, and S. Cui, "Distributed detection in noisy sensor networks," in *2011 IEEE International Symposium on Information Theory*, pp. 2856-2860, St. Petersburg, July 2011.
- [38] S. Shahrampour, M. A. Rahimian, and A. Jadbabaie, "Switching to Learn," arXiv preprint arXiv:1503.03517, 2015.
- [39] A. Nedić, A. Olshevsky, and C. Uribe, "Fast Convergence Rates for Distributed Non-Bayesian Learning," arXiv preprint arXiv:1508.05161, 2015.
- [40] A. K. Sahu and S. Kar, "Distributed sequential detection for Gaussian binary hypothesis testing," *IEEE Transactions on Signal Processing*, to appear. <http://arxiv.org/pdf/1411.7716v2.pdf>
- [41] V. Srivastava and N. E. Leonard, "Collective decision-making in ideal networks: The speed-accuracy tradeoff," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 1, pp. 121-132, Mar. 2014.
- [42] A. Wiesel and A. O. Hero, "Distributed Covariance Estimation in Gaussian Graphical Models," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 211-220, Jan. 2012.
- [43] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed Parameter Estimation in Sensor Networks: Nonlinear Observation Models and Imperfect Communication," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3575-3605, June 2012.
- [44] S. Kar and J. M. F. Moura, "Convergence Rate Analysis of Distributed Gossip (Linear Parameter) Estimation: Fundamental Limits and Trade-offs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 674-690, Aug. 2011.
- [45] H. A. David and H. N. Nagaraja, *Order Statistics*, New Jersey: John Wiley & Sons, Inc., 2003.