Distributed Recursive Composite Hypothesis Testing: Imperfect Communication

Anit Kumar Sahu and Soummya Kar
Department of Electrical and Computer Engineering
Carnegie Mellon University, Pittsburgh, PA 15213
email:{anits, soummyak}@andrew.cmu.edu

***Abstract*—This paper focuses on the problem of distributed composite hypothesis testing in a noisy network of sparsely interconnected agents in which a pair of agents exchange information over an additive noise channel. The network objective is to test a simple null hypothesis against a composite alternative concerning the state of the field, modeled as a vector of (continuous) unknown parameters determining the parametric family of probability measures induced on the agents' observation spaces under the hypotheses. A recursive generalized likelihood ratio test (GLRT) type algorithm in a distributed setup of the *consensus+innovations* form is proposed, in which the agents update their parameter estimates and decision statistics by simultaneously processing the latest sensed information (*innovations*) and information obtained from neighboring agents (*consensus*). This paper characterizes the conditions and the testing algorithm design parameters which ensure that the probabilities of decision errors decay to zero asymptotically in the large sample limit.**

1. INTRODUCTION

This paper focuses on the problem of distributed composite hypothesis testing in multi-agent networks where the objective is not only to estimate the state (possibly high dimensional) of the environment but also detect as to which hypothesis is in force based on the all the measurement data at all the agents, where both the estimation and detection schemes run in parallel. In particular, we focus on distributed multi-agent networks where the communication and information exchange between the agents is over additive noise channels and limited to a pre-assigned, possibly sparse, communication graph. This problem is relevant to practical applications such as cooperative spectrum sensing and MIMO radars. Generalized Likelihood Ratio Test (GLRT) ([1]) is a classical approach and has garnered a lot of interest in centralized setups for addressing problems of composite testing. Not only is GLRT an offline[1] batch processing algorithm, but also involves an inherent waiting time as the detection procedure which uses the *optimal* underlying parameter estimate as a plug-in estimate cannot start until a reasonably accurate estimate typically the maximum likelihood estimate of the underlying parameter (state) is obtained. Moreover, in order to achieve a reasonable detection performance, more sensing is needed. In most multi-agent networked scenarios, which are typically energy constrained, the priority is to obtain reasonable inference performance by expending fewer amount of resources. Motivated by requirements such as the latter, we propose an algorithm $\mathcal{CIGLRT}$, where the inter-agent collaboration is over noisy channels and is restricted to a pre-assigned possibly sparse communication graph and the detection and estimation schemes run in a parallel fashion with a view to reduce energy and resource consumption while achieving reasonable detection performance.

We briefly review the related existing work in the literature on distributed hypothesis testing in multi-agent networks before discussing the problem setup and our algorithm. Existing distributed detection schemes in literature can be broadly classified into three classes. The first class includes architectures characterized by the presence of a fusion center where the fusion center receives the decisions or local measurements or test statistic or its quantized version from all the agents (see, for example [4], [5]) and subsequently conducts the estimation and detection schemes. The second class consists of consensus schemes (see, for example [6], [7]) with no fusion center and where, in the first phase the agents collect information for a long period of time from the environment followed by the second phase, where agents exchange information in their respective neighborhoods to arrive at a decision collectively. The third class consists of schemes which perform simultaneous assimilation of information obtained from sensing and communication (see, for example [8]–[10]).

The algorithm we present in this paper belongs to the third class, where agents make conditionally independent and temporally identically distributed (but possibly spatially heterogeneous) observations and update their parameter estimate and test statistic by simultaneous assimilation of the (noisy) information obtained from the neighboring agents (*consensus*) and the latest locally sensed information (*innovation*). This justifies the name $\mathcal{CIGLRT}$ which is a distributed GLRT type algorithm of the *consensus* + *innovations* form. *Consensus* + *innovations* generalizes stochastic approximation to distributed multi-agent networked scenarios. Technically speaking, *consensus* + *innovations* algorithms may be viewed as mixed time scale stochastic approximation procedures. By mixed time scale, we mean stochastic approximation algorithms where two potentials influence the update step through different gain sequences. It is to be noted that the above notion of mixed time scale is different from stochastic

This work was supported in part by NSF under grants CCF-1513936 and ECCS-1408222.

[1]By offline, we strictly refer to the classical implementation of the GLRT. Various recursive versions of GLRT type approaches catering to problems like sequential composite hypothesis testing and change detection (see, for example, [2], [3]), although in centralized processing scenarios have been developed.

algorithms with coupling (see [11]), where a quickly switching parameter influences the relatively slower dynamics of another state, leading to averaged dynamics. In this paper, so as to closely replicate typical practical sensing environments, we assume an agent's observations, say for agent n, is M_n dimensional, where $M_n << M$, M being the dimension of the underlying static vector parameter. In addition to the local non-observability, in order to replicate typical communication environments, we assume that the communication between agents is noisy, i.e., the agents' communicate over additive noise channels. Under a global identifiability condition and connectivity of the inter-agent communication network, we not only show the consistency of the parameter estimate sequence but also show the existence of feasible choice of thresholds and other algorithm design parameters which ensure that the probabilities of errors decay to zero asymptotically. (Fully) distributed detection schemes, in prior literature are mostly concerned with either binary simple hypothesis testing (see, for example [8], [9], [12]) or multiple simple hypothesis testing (finite classification) (see, for example [13], [14]) in contrast with the composite hypotheses with constant vector parameterization which can take values in a continuous space is studied in this paper. Recently, [15], [16] considered the problem of distributed composite hypothesis testing in ideal communication scenarios, i.e., noiseless communication. Distributed detection problems based on the *consensus+innovations* approach in the noisy communication setting has been addressed in [9]. In this paper, our analysis contributes to understanding large deviations of mixed time scale stochastic approximation processes in addition to the characterization of the decision statistic and the parameter estimate sequences that evolve in a closed loop fashion. Addressing the fully composite testing setup with a continuous range of alternatives requires novel technical machinery in the form of development of analysis of efficient distributed estimation and detection procedures that interact in a closed loop fashion in a noisy communication scenario, which we pursue in this paper. The rest of the paper is organized as follows. Spectral graph theory, preliminaries and notation are discussed next. The sensing model is described in Section 2, where we also review some preliminaries concerning the classical Generalized Likelihood Ratio Tests. Section 3 presents the proposed $\mathcal{CIGLRT}$ algorithm, while Section 4 concerns with the main results of the paper. Finally, Section 5 concludes the paper.

Spectral Graph Theory. The inter-agent communication network is a simple[2] undirected graph $G = (V, E)$, where V denotes the set of agents or vertices with cardinality $|V| = N$, and E the set of edges with $|E| = M$. If there exists an edge between agents i and j, then $(i, j) \in E$. A path between agents i and j of length m is a sequence $(i = p_0, p_1, \cdots, p_m = j)$ of vertices, such that $(p_t, p_{t+1}) \in E$, $0 \leq t \leq m-1$. A graph is connected if there exists a path between all possible agent pairs. The neighborhood of an agent n is given by $\Omega_n = \{j \in V | (n, j) \in E\}$. The degree of agent n is given by $d_n = |\Omega_n|$. The structure of the graph is represented by the symmetric $N \times N$ adjacency matrix $\mathbf{A} = [A_{ij}]$, where $A_{ij} = 1$ if $(i, j) \in E$, and 0 otherwise. The degree matrix is given by the diagonal matrix $\mathbf{D} = diag(d_1 \cdots d_N)$. The graph Laplacian matrix is defined as $\mathbf{L} = \mathbf{D} - \mathbf{A}$. The Laplacian is a positive semidefinite matrix, hence its eigenvalues can be ordered and represented as $0 = \lambda_1(\mathbf{L}) \leq \lambda_2(\mathbf{L}) \leq \cdots \lambda_N(\mathbf{L})$. Furthermore, a graph is connected if and only if $\lambda_2(\mathbf{L}) > 0$.

[2]A graph is said to be simple if it is devoid of self loops and multiple edges.

2. SENSING MODEL AND PRELIMINARIES

There are N agents deployed in the network. Every agent n at time index t makes a noisy observation $y_n(t)$, a noisy nonlinear function of θ^* which is a M-dimensional parameter, i.e., $\theta^* \in \mathbb{R}^M$, which comes from a probability distribution $\mathbb{P}_0$ under the hypothesis $\mathcal{H}_0$, whereas, under the composite alternative $\mathcal{H}_1$, the observation is sampled from a probability distribution which is a member of a parametric family $\{\mathbb{P}_{\theta^*}\}$. Formally,

$$\begin{aligned} &\mathcal{H}_1 : \mathbf{y}_n(t) = \mathbf{h}_n(\theta^*) + \gamma_n(t) \\ &\mathcal{H}_0 : \mathbf{y}_n(t) = \gamma_n(t). \end{aligned} \tag{1}$$

where $\mathbf{h}_n(.)$ is, in general, non-linear function, $\{\mathbf{y}_n(t)\}$ is a $\mathbb{R}^{M_n}$-valued observation sequence for the n-th agent, where typically $M_n \ll M$ and $\{\gamma_n(t)\}$ is a zero-mean temporally independent and identically distributed (i.i.d.) Gaussian noise sequence at the n-th agent with nonsingular covariance matrix $\mathbf{\Sigma}_n$, where $\mathbf{\Sigma}_n \in \mathbb{R}^{M_n \times M_n}$. Moreover, the noise sequences at two agents n, l with $n \neq l$ are independent. We emphasize here that the parameter θ^* is deterministic but unknown.

By taking $\mathbf{h}_n(\mathbf{0}) = \mathbf{0}$, $\forall n$ and certain other identifiability and regularity conditions outlined below, in the above formulation the null hypothesis corresponds to $\theta^* = 0$ and the composite alternative to the case $\theta^* \neq \mathbf{0}$. Note that the formulation assumes no indifference zone. However, the performance of the proposed distributed approach in terms of the error probabilities under the composite alternative will depend on the specific instance of θ^* in force. We start by making some identifiability assumptions on our sensing model before stating the algorithm.

Assumption A1. *The sensing model is globally observable, i.e., any two distinct values of θ and θ^* in the parameter space $\mathbb{R}^M$ satisfy $\sum_{n=1}^N \|\mathbf{h}_n(\theta) - \mathbf{h}_n(\theta^*)\|^2 = 0$ if and only if $\theta = \theta^*$.*

For instance, if the sensing functions are taken to be linear, i.e., the observation model in (1) reduces to $\mathbf{y}_n(t) = \mathbf{H}_n\theta^* + \gamma_n(t)$ and the condition in assumption A1 reduces to the condition that $\sum_{n=1}^N \mathbf{H}_n^\top \mathbf{H}_n$ is invertible. It is to be noted that a distributed collaborative testing is necessary in general as in the setup under consideration, the local observation models at each agent is not observable for θ^*. In order to motivate our distributed testing approach (presented in Section 3), we now review some concepts from Generalized Likelihood Ratio Testing.

Consider a generalized target detection problem, where the absence of target is modeled by a simple hypothesis $\mathcal{H}_0$, whereas the alternative $\mathcal{H}_1$ which is a composite hypothesis models the presence of a target as the underlying parameter θ^* is unknown and can possibly attain a range of values. The agents in the network collect data over time and let $\mathbf{y}(t) = \left[\mathbf{y}_1(t)^\top \cdots \mathbf{y}_N(t)^\top\right]^\top$ represent the data from all the agents at time t, where $\mathbf{y}(t) \in \mathbb{R}^{\sum_{n=1}^N M_n}$. The fusion center in a centralized setup has access to all the agents' observations i.e. $\mathbf{y}(t)$ at all times t. In a centralized setup, a classical testing approach is the generalized likelihood ratio test (GLRT). Formally, the GLRT decision rule is defined as follows:

$$\mathcal{H} = \begin{cases} \mathcal{H}_1, & \text{if } \max_\theta \sum_{t=0}^{T} \log \frac{f_\theta(\mathbf{y}(t))}{f_0(\mathbf{y}(t)} > \eta, \\ \mathcal{H}_0, & \text{otherwise}, \end{cases} \tag{2}$$

where η is a predefined threshold and with the assumption that the observations made by the agents are conditionally independent, we have, $f_0(\mathbf{y}(t)) = f_0^1(\mathbf{y}_1(t)) \cdots f_0^N(\mathbf{y}_N(t))$ and $f_\theta(\mathbf{y}(t)) = f_\theta^1(\mathbf{y}_1(t)) \cdots f_\theta^N(\mathbf{y}_N(t))$, which represent the likelihood of observing $\mathbf{y}(t)$ under $\mathcal{H}_0$ and $\mathcal{H}_1$ respectively, where $f_0^n(\mathbf{y}_n(t))$ and $f_\theta^n(\mathbf{y}_n(t))$ represent the likelihood of observing $\mathbf{y}_n(t)$ at agent n under $\mathcal{H}_0$ and $\mathcal{H}_1$ with θ as the underlying parameter respectively. Hence, the maximization in (2) can be written as

$$\max_\theta \sum_{t=0}^{T} \log \frac{f_\theta(\mathbf{y}(t))}{f_0(\mathbf{y}(t))} = \max_\theta \sum_{t=0}^{T} \sum_{n=1}^{N} \log \frac{f_\theta^n(\mathbf{y}_n(t))}{f_0^n(\mathbf{y}_n(t))}. \tag{3}$$

The key bottleneck in the implementation of the classical GLRT is the computation of the decision statistic in the maximization in (3) as it needs access to all the data collected so far. In general, as the maximizer of (3) depends on the raw data instance, it is not known apriori. Hence as far as communication complexity in the GLRT implementation is concerned, the maximization step incurs the major overhead - in fact, a direct implementation of the maximization (3) requires access to the entire raw data $\mathbf{y}$ at the fusion center.

We assume that the inter-agent communication is imperfect, i.e., noisy. To be specific, we assume that an agent pair (i, j) exchange information over a vector additive zero-mean noise channel. Formally speaking, if agent i transmits a data vector $\mathbf{z} \in \mathbb{R}^k$ to agent j, the information received at agent j is given by $\tilde{\mathbf{z}} = \mathbf{z} + \psi_{i,j}$, where the noise vector $\psi_{i,j}$ is Gaussian with zero mean and has finite variance $\mathbf{\Sigma}_{ij}$. Furthermore, we assume the transmission channel noises are independent over transmissions and across the graph links.

To mitigate the communication complexity in realizing a fusion center having access to all the data, we present a distributed algorithm in which agents collaborate locally to obtain a maximizing θ. In order to obtain reasonable decision performance with such localized communication, we propose a distributed detector of the *consensus+innovations* type, which have been introduced in [17]. In particular, each agent sequentially updates its parameter estimate and decision statistic in two parallelly running recursive schemes by assimilating the information obtained from its neighbors (*consensus potential*) and latest sensed local information (*innovation potential*).

3. $\mathcal{CIGLRT}$: ALGORITHM

In this section, we develop the algorithm $\mathcal{CIGLRT}$ which is an extension of the variant proposed in [15]. To be specific, the algorithm proposed in this paper is built around a setup where the inter-agent communication is noisy.

We skip the proofs due to space limitations. The proofs can be found in the longer manuscript ([18]).

We propose a distributed detector of the *consensus+innovations* form for the scenario outlined in (1). Before discussing the details of our algorithm, we state an assumption on the inter-agent communication graph.

Assumption A2. *The inter-agent communication graph is connected, i.e.,* $\lambda_2(\mathbf{L}) > 0$*, where* $\mathbf{L}$ *denotes the associated graph Laplacian matrix.*

Algorithm $\mathcal{CIGLRT}$

The algorithm $\mathcal{CIGLRT}$ consists of three interactive recursive processes running in parallel, namely, the parameter estimate update process, the decision statistic update process and a decision formation rule, as described below.

We state an assumption on the sensing functions, before describing the algorithm.

Assumption A3. *For each agent* n, $\forall\theta \neq \theta_1$, *the sensing functions* $\mathbf{h}_n$ *are continuously differentiable on* $\mathbb{R}^M$ *and Lipschitz continuous with constants* $k_n > 0$, *i.e.,*

$$\|\mathbf{h}_n(\theta) - \mathbf{h}_n(\theta_1)\| \leq k_n \|\theta - \theta_1\|. \tag{4}$$

Before proceeding further, we state an assumption on the sensing functions in addition to the state-dependent innovation gains which guarantees the existence of Lyapunov functions, which in turn ensures the convergence of the distributed estimation procedure.

Assumption A4. *The following aggregate strict monotonicity condition holds: there exists a constant* $c^* > 0$ *such that for each pair of* θ *and* $\acute{\theta}$ *with* $\theta \neq \acute{\theta}$ *we have that*

$$\sum_{n=1}^{N} \left(\theta - \acute{\theta}\right)^\top (\nabla\mathbf{h}_n(\theta)) \mathbf{\Sigma}_n^{-1} \left(\mathbf{h}_n(\theta) - \mathbf{h}_n\left(\acute{\theta}\right)\right)$$
$$\geq c^* \left\|\theta - \acute{\theta}\right\|^2. \tag{5}$$

Parameter Estimate Update. The algorithm $\mathcal{CIGLRT}$ generates the sequence $\{\theta_n(t)\} \in \mathbb{R}^M$ at the n-th agent according to the following recursive scheme

$$\theta_n(t+1) = \theta_n(t) - \underbrace{b\alpha_t \sum_{l\in\Omega_n} (\theta_n(t) - \theta_l(t) - \psi_{n,l}(t))}_{\text{neighborhood consensus}}$$
$$+ \underbrace{\alpha_t \nabla\mathbf{h}_n(\theta_n(t)) \mathbf{\Sigma}_n^{-1} (\mathbf{y}_n(t) - \mathbf{h}_n(\theta_n(t)))}_{\text{local innovation}}, \tag{6}$$

where Ω_n denotes the communication neighborhood of agent n, b is a positive constant, $\psi_{n,l}(t)$ is the communication noise in the link between n and l, $\nabla\mathbf{h}_n(.)$ denotes the gradient of $\mathbf{h}_n$, which is a matrix of dimension $\mathbf{M} \times \mathbf{M}_n$, with the (i, j)-th entry given by $\frac{\partial[\mathbf{h}_n(\theta_n(t))]_j}{\partial[\theta_n(t)]_i}$ and $\{\alpha_t\}$ is the innovation weight sequence (to be specified shortly). Note that, in (6), each agent $l \in \Omega_n$ intends to send it's exact estimate to agent n, but agent n receives a noisy version of estimates from agents in its neighborhood as the inter-agent communication is over noisy links. The update in (6) can be written in a compact manner as follows:

$$\theta(t+1) = \theta(t) - b\alpha_t (\mathbf{L} \otimes \mathbf{I}_M) \theta(t)$$
$$+ \alpha_t \mathbf{G}(\theta(t)) \mathbf{\Sigma}^{-1} (\mathbf{y}(t) - \mathbf{h}(\theta(t))) + b\alpha_t \mathbf{\Psi}(t), \tag{7}$$

where $\theta(t)^\top = [\theta_1(t)^\top \cdots \theta_N(t)^\top]$, $\mathbf{h}(\theta(t)) = \left[\mathbf{h}_1^\top(\theta_1(t)) \cdots \mathbf{h}_N^\top(\theta_N(t))\right]^\top$, $\mathbf{y}(t)^\top = [y_1(t)^\top \cdots y_N(t)^\top]^\top$, $\mathbf{G}(\theta(t)) = diag[\nabla\mathbf{h}_1(\theta_1(t)), \cdots, \nabla\mathbf{h}_N(\theta_N(t))]$, $\mathbf{\Sigma} = diag[\mathbf{\Sigma}_1, \cdots, \mathbf{\Sigma}_N]$ and

$\mathbf{\Psi}^\top(t) = \left[\left(\sum_{l\in\Omega_1}\psi_{1,l}(t)\right)^\top \cdots \left(\sum_{l\in\Omega_N}\psi_{N,l}(t)\right)^\top\right]^\top.$

We make the following assumption on the weight sequence $\{\alpha_t\}$.

Assumption A5. *The weight sequence $\{\alpha_t\}$ is of the form $\alpha_t = (t+1)^{-1}$ and the positive constant b is such that $b < \frac{1}{\lambda_N(\mathbf{L})}$.*

Decision Statistic Update. The algorithm $\mathcal{CIGLRT}$ generates the decision statistic sequence $\{z_n(t)\}$ at the n-th agent according to the distributed recursive scheme

$$z_n(t+1) = z_n(t) - \underbrace{b\alpha_t \sum_{l\in\Omega_n}(z_n(t) - z_l(t) - \zeta_{n,l}(t))}_{\text{consensus}} + \underbrace{\alpha_t\left(\log\frac{f_{\theta_n(t)}(y_n(t))}{f_0(y_n(t))} - z_n(t)\right)}_{\text{innovation}}, \quad (8)$$

where $\zeta_{n,l}(t)$ is the communication noise in the link between n and l, $f_\theta(.)$ and $f_0(.)$ represent the likelihoods under $\mathcal{H}_1$ and $\mathcal{H}_0$ respectively and $\log\frac{f_{\theta_n(t)}(y_n(t))}{f_0(y_n(t))} = \mathbf{h}_n^\top(\theta_n(t))\mathbf{\Sigma}_n^{-1}y_n(t) - \frac{\mathbf{h}_n^\top(\theta_n(t))\mathbf{\Sigma}_n^{-1}\mathbf{h}_n(\theta_n(t))}{2}$. Note that, in (8), each agent $l\in\Omega_n$ intends to send it's exact test statistic to agent n, but agent n receives a noisy version of test statistics from agents in its neighborhood as the inter-agent communication is over noisy links.

Decision Rule.
The following decision rule is adopted at all times t at all agents n :

$$\mathcal{H}_n(t) = \begin{cases}\mathcal{H}_0 & z_n(t) \le \eta\\ \mathcal{H}_1 & z_n(t) > \eta,\end{cases} \quad (9)$$

where $\mathcal{H}_n(t)$ represents the local decision at agent n at time t. Under the aegis of such a decision rule, the associated probabilities of errors are as follows:

$$\mathbb{P}_{M,\theta^*}(t) = \mathbb{P}_{1,\theta^*}(z_n(t)\le\eta),\ \mathbb{P}_{FA}(t) = \mathbb{P}_0(z_n(t)>\eta), \quad (10)$$

where $\mathbb{P}_0(.)$ and $\mathbb{P}_{1,\theta}(.)$ denote the probability of the event conditioned on the null hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$, where θ is the parametric alternative and $\mathbb{P}_{M,\theta^*}$ and $\mathbb{P}_{FA}$ refer to probability of miss and probability of false alarm respectively. Since, the sources of randomness in our formulation are the observations $\mathbf{y}_n(t)$'s made by the agents in the network, the communication noises $\mathbf{\Psi}(t)$ and $\zeta(t)$ encountered for exchange of parameter estimates and decision statistics respectively, we define the natural filtration $\{\mathcal{F}_t\}$ generated by the random observations and the communication noise, i.e., $\mathcal{F}_t = \sigma\left(\{\mathbf{\Psi}(s), \zeta(s), \{\mathbf{y}_n(s)\}_{n=1}^N\}_{s=0}^{t-1}\right)$, which is the sequence of σ-algebras induced by the observation processes, in order to model the overall available network information at all times. Finally, a stochastic process $\{\mathbf{x}(t)\}$ is said to be $\{\mathcal{F}_t\}$-adapted if the σ-algebra $\sigma(\mathbf{x}(t))$ is a subset of $\mathcal{F}_t$ at each t.

4. $\mathcal{CIGLRT}$: MAIN RESULTS

In this section, we specifically characterize the thresholds for which the probability of miss and probability of false alarm decay to zero asymptotically.

We first state the main result concerning the parameter estimate sequences $\{\theta_n(t)\}$.

Theorem 4.1. *Consider the $\mathcal{CIGLRT}$ algorithm under Assumptions A1-A5, and the sequence $\{\theta(t)\}_{t\ge0}$ generated according to* (6)*. We then have*

$$\mathbb{P}_{\theta^*}\left(\lim_{t\to\infty}(t+1)^\tau\|\theta_n(t)-\theta^*\| = 0, \forall 1\le n\le N\right) = 1, \quad (11)$$

where $\tau\in[0,1/2)$.

Proof. The proof follows exactly from the proof of Theorem 4.1 in [15]. ☐

We define the following quantities which will be crucial for stating the next theorem : let $\mathbf{\Sigma}_{c,1}^*$ and $\mathbf{\Sigma}_{c,0}^*$ be given by

$$\mathbf{\Sigma}_{c,1}^* = \mathbf{V}_L\mathbf{M}_1\mathbf{V}_L^\top,\ \mathbf{\Sigma}_{c,0}^* = \mathbf{V}_L\mathbf{M}_0\mathbf{V}_L^\top \quad (12)$$

respectively, where $\mathbf{M}_1$ and $\mathbf{M}_0$ are given by

$$[\mathbf{M}_l]_{ij} = \left[\mathbf{V}_L^\top\mathbf{\Sigma}_l^*\mathbf{V}_L\right]_{ij}(b[\mathbf{D}_L]_{ii} + b[\mathbf{D}_L]_{jj} + 1)^{-1}, l = 0,1, \quad (13)$$

where $[\mathbf{A}]_{ij}$ denotes the (i,j)-th entry of a matrix $\mathbf{A}$ and $\mathbf{\Sigma}_1^*$ and $\mathbf{\Sigma}_0^*$ are given by

$$\mathbf{\Sigma}_1^* = \mathbf{h}^*(\mathbf{1}_N\otimes\theta^*)\mathbf{\Sigma}^{-1}(\mathbf{h}^*(\mathbf{1}_N\otimes\theta^*))^\top + b^2\mathbf{\Sigma}_c, \mathbf{\Sigma}_0^* = b^2\mathbf{\Sigma}_c, \quad (14)$$

respectively, whereas $\mathbf{V}_L$ and $\mathbf{D}_L$ represent the matrix of eigenvectors and eigenvalues of $\mathbf{L}$ respectively, i.e., $\mathbf{L} = \mathbf{V}_L^\top\mathbf{D}_L\mathbf{V}_L$, and $\mathbf{\Sigma}_c$ denotes the covariance matrix of the channel noise encountered in the test statistic exchange among agents given by the process $\{\zeta(t)\}$.

Theorem 4.2. *Consider the $\mathcal{CIGLRT}$ algorithm under Assumptions A1-A5, and the sequence $\{\mathbf{z}(t)\}$ generated according to* (8)*. We then have under $\mathbb{P}_{\theta^*}$*

$$\sqrt{t+1}\left(\mathbf{z}(t) - (b\mathbf{L}+\mathbf{I})^{-1}\frac{\mathbf{h}^*(\mathbf{1}_N\otimes\theta^*)\mathbf{\Sigma}^{-1}\mathbf{h}(\mathbf{1}_N\otimes\theta^*)}{2}\right) \overset{\mathcal{D}}{\Longrightarrow} \mathcal{N}\left(0,\mathbf{\Sigma}_{c,1}^*\right) \quad (15)$$

$\forall n$, and under $\mathbb{P}_0$

$$\sqrt{t+1}\,(\mathbf{z}(t)) \overset{\mathcal{D}}{\Longrightarrow} \mathcal{N}\left(0,\mathbf{\Sigma}_{c,0}^*\right),\ \forall n, \quad (16)$$

where $\overset{\mathcal{D}}{\Longrightarrow}$ denotes convergence in distribution (weak convergence).

Theorem 4.2 asserts the asymptotic normality of the test statistic $\{z_n(t)\}, \forall n$. It is to be noted that the asymptotic mean of $\mathbf{z}(t)$ which is given by $(b\mathbf{L}+\mathbf{I})^{-1}\frac{\mathbf{h}^*(\mathbf{1}_N\otimes\theta^*)\mathbf{\Sigma}^{-1}\mathbf{h}(\mathbf{1}_N\otimes\theta^*)}{2}$ has all of its entries positive, as $(b\mathbf{L}+\mathbf{I})$ is a M-matrix (see, [19]) and hence its inverse has all of its entries non-negative, i.e., $\left[(b\mathbf{L}+\mathbf{I})^{-1}\right]_{ij} \ge 0, \forall i,j = 1,\cdots,N$. The next result concerns with the characterization of thresholds which ensures the probability of miss and probability of false alarm as defined in (10) decay to zero asymptotically.

Theorem 4.3. *Let the hypotheses of Theorem 4.2 hold. Consider the decision rule defined in* (9)*. For agent n, all θ^* which satisfy $\left[(b\mathbf{L}+\mathbf{I})^{-1}\frac{\mathbf{h}^*(\mathbf{1}_N\otimes\theta^*)\mathbf{\Sigma}^{-1}\mathbf{h}(\mathbf{1}_N\otimes\theta^*)}{2}\right]_n > \frac{2\sum_{n=1}^N M_n}{N}$,* *where* $[\mathbf{a}]_n$ *denotes the* n*-th element of a vector* $\mathbf{a}$*, we have the following choice of the thresholds*

$$\frac{2\sum_{n=1}^{N} M_n}{N} < \eta_n < \left[(b\mathbf{L}+\mathbf{I})^{-1}\frac{\mathbf{h}^*(\mathbf{1}_N\otimes\theta^*)\mathbf{\Sigma}^{-1}\mathbf{h}(\mathbf{1}_N\otimes\theta^*)}{2}\right]_n, \tag{17}$$

for which we have that $\mathbb{P}_{M,\theta^*}(t)\to 0$ *and* $\mathbb{P}_{FA}(t)\to 0$ *as* $t\to\infty$*. Specifically,* $\mathbb{P}_{FA}(t)$ *decays to zero exponentially with the following large deviations exponent*[3]

$$\limsup_{t\to\infty}\frac{1}{t}\log\left(\mathbb{P}_0\left(z_n(t)>\eta_n\right)\right)\le\max\left\{-\frac{\eta_n^2}{8b^2\left\|\mathbf{\Sigma}_c\right\|},-LE(\lambda^*)\right\}, \tag{18}$$

where $LE(\lambda)=\frac{N\eta_n\lambda}{4}+\left(\frac{\sum_{n=1}^{N}M_n}{2}\right)\log\left(1-\lambda\right)$ *and* $\lambda^*=\frac{2\sum_{n=1}^{N}M_n}{N\eta_n}$.

It is to be noted that the thresholds across agents can be chosen to be different owing to the unequal asymptotic mean at different agents and hence the large deviations upper bound across different agents may be different. We discuss how the above result can be used in practice to identify thresholds that lead to asymptotic decay of the probabilities of error. It is to be noted that as the observation parameters, i.e., M_n, N are known apriori, the threshold can be chosen to be $\frac{2\sum_{n=1}^{N}M_n}{N}+\epsilon$, where ϵ can be chosen to be arbitrarily small. Further, from the feasible range of thresholds in Theorem 4.3, a range on the θ^*s' can be obtained in terms of $\left\|\mathbf{h}\left(\mathbf{1}_N\otimes\theta^*\right)\right\|$ such that under $\mathcal{H}_1$, as long as the true value θ^* of the parameter belongs to this range, the probability of miss is guaranteed to decay to zero asymptotically. It is important to note in this context that there exists some weak signals, i.e., signals with low $\left\|\mathbf{h}\left(\mathbf{1}_N\otimes\theta^*\right)\right\|$ (but non-zero), for which there may not exist a choice of thresholds to ensure asymptotically decaying probability of miss. The signals for which Theorem 4.3 is rendered to be inconclusive in the manner described above, can be categorized in terms of θ^*. Theorem 4.3 characterizes the range of feasible thresholds exist that guarantee $\mathbb{P}_{M,\theta^*}(t)$, $\mathbb{P}_{FA}(t)\to 0$ as $t\to\infty$. The incorporation of inaccurate initial parameter estimates into the decision statistic, though sub-optimal with respect to the classical GLRT, makes the detection scheme of $\mathcal{CIGLRT}$ a recursive *online* procedure, while the classical GLRT is an *offline* batch procedure as the corresponding parameter estimate used at any time instant depends on the entire raw data obtained at all agents so far and needs to be estimated first before computing the decision statistic. In spite of the sub-optimality in the update of the corresponding decision statistic, the algorithm $\mathcal{CIGLRT}$ ensures that the probabilities of errors decay to zero in the large sample (time) limit in the scenario of imperfect communication.

5. CONCLUSION

In this paper, we have proposed a *consensus* + *innovations* type algorithm, $\mathcal{CIGLRT}$, in which every agent updates its parameter estimate and decision statistic by simultaneous processing of (noisy) neighborhood information and local newly sensed information and where the inter-agent collaboration is over additive noise channels and is restricted to a possibly sparse but connected communication graph. Under rather generic assumptions on the agent sensing model, and an (aggregate) global observability condition we established the consistency of the parameter estimate sequence and characterized the feasible choice of thresholds and other algorithm parameters which ensure that the probabilities of errors pertaining to the detection scheme decay to zero asymptotically. A natural direction for future research consists of considering models with non-Gaussian noise.

[3]By large deviations exponent, we mean the large deviations upper bound.

REFERENCES

[1] O. Zeitouni, J. Ziv, and N. Merhav, "When is the generalized likelihood ratio test optimal?" *IEEE Transactions on Information Theory*, vol. 38, no. 5, pp. 1597–1602, 1992.

[2] D. Siegmund and E. Venkatraman, "Using the generalized likelihood ratio statistic for sequential detection of a change-point," *The Annals of Statistics*, pp. 255–271, 1995.

[3] A. S. Willsky and H. L. Jones, "A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems," *Automatic Control, IEEE Transactions on*, vol. 21, no. 1, pp. 108–112, 1976.

[4] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors i. advanced topics," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 64–79, 1997.

[5] J. N. Tsitsiklis, "Decentralized detection," *Advances in Statistical Signal Processing*, vol. 2, no. 2, pp. 297–344, 1993.

[6] S. Kar and J. M. F. Moura, "Consensus based detection in sensor networks: Topology optimization under practical constraints," *Proc. 1st Intl. Wrkshp. Inform. Theory Sensor Networks*, 2007.

[7] R. Olfati-Saber, E. Franco, E. Frazzoli, and J. S. Shamma, "Belief consensus and distributed hypothesis testing in sensor networks," in *Networked Embedded Sensing and Control*. Springer, 2006, pp. 169–182.

[8] D. Bajovic, D. Jakovetic, J. Xavier, B. Sinopoli, and J. M. F. Moura, "Distributed detection via Gaussian running consensus: Large deviations asymptotic analysis," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4381–4396, 2011.

[9] S. Kar, R. Tandon, H. V. Poor, and S. Cui, "Distributed detection in noisy sensor networks," in *IEEE International Symposium on Information Theory Proceedings (ISIT), 2011*. IEEE, 2011, pp. 2856–2860.

[10] 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, 2011.

[11] V. S. Borkar, "Stochastic approximation with two time scales," *Systems & Control Letters*, vol. 29, no. 5, pp. 291–294, 1997.

[12] D. Jakovetic, J. M. F. Moura, and J. Xavier, "Distributed detection over noisy networks: Large deviations analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4306–4320, 2012.

[13] A. Lalitha, A. Sarwate, and T. Javidi, "Social learning and distributed hypothesis testing," in *IEEE International Symposium on Information Theory (ISIT), 2014*. IEEE, 2014, pp. 551–555.

[14] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, "Non-Bayesian social learning," *Games and Economic Behavior*, vol. 76, no. 1, pp. 210–225, 2012.

[15] A. K. Sahu and S. Kar, "Recursive distributed detection for composite hypothesis testing : Algorithms and Asymptotics," *arXiv preprint arXiv:1601.04779*, 2016.

[16] ——, "Distributed generalized likelihood ratio tests : Fundamental limits and Tradeoffs," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2016. ICASSP'16*. IEEE, 2016.

[17] 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.

[18] A. K. Sahu and S. Kar, "Distributed recursive composite hypothesis testing : Imperfect communication," https://dl.dropboxusercontent.com/u/33858933/isitextended.pdf, 2016, [Online].

[19] G. Poole and T. Boullion, "A survey on M-matrices," *SIAM review*, vol. 16, no. 4, pp. 419–427, 1974.