Asymptotic Optimality of Consensus-Based Sequential Probability Ratio Test

Shang Li and Xiaodong Wang

***Abstract*—This work considers the sequential hypothesis testing problem in the fully distributed sensor network. In specific, each sensor can observe samples over time, exchange information with adjacent sensors, and perform testing based on its own locally available decision statistic. Under such setting, we study the sequential probability ratio test based on the statistic that is obtained by running consensus algorithm. It is shown that, under certain regularity conditions on the data distribution and network topology, this distributed sequential test procedure yields the order-2 asymptotically optimal performance at all sensors.**

***Index Terms*—Distributed sequential detection, sensor networks, asymptotic optimality, running consensus.**

I. Introduction

The sequential hypothesis testing has been widely adopted in practice to reduce the data sample size compared to its fixed-sample-size counterpart. Notably, the sequential probability ratio test (SPRT) attains the minimum expected sample sizes under both hypotheses subject to the error probability constraints [1, 2]. In the meantime, the recent decade has witnessed the surge of smart devices that can be connected through wireless links and form cooperative networks. Some examples include the vehicular Ad Hoc network as part of the intelligent transportation system, and the social network that connects people through online friendship. Many applications pertaining to these examples involve choosing between two hypotheses with stringent requirements on the decision latency, necessitating solutions that can integrate the sequential hypothesis testing into the cooperative networks.

The existing studies mainly approach this problem by assuming a fusion center for the sensor network, which may not be realistic in general. Rather, the fully distributed network topology, where sensors can only exchange information with neighbours defined by a network graph (cf. Fig. 1), is yet to be investigated. Naturally, the distributed network is prone to sub-optimal cooperative performance due to the lack of global information at each sensor. Therefore, the key challenge is to devise efficient information exchange mechanism and test procedure such that each sensor can optimize its distributed sequential test, and, if feasible, achieve the globally optimal performance. To that end, the consensus algorithm serves as a promising technique for disseminating the global information to each sensor in the network.

The distributed consensus problem has been well studied under the static/fixed-sample-size paradigm [3]. There, each sensor starts with a private sample and aims to obtain the average of all private samples in the system through inter-sensor information exchange [4, 5]. In this series of works, a new sample is not allowed to enter into the network during the process of "reaching consensus", thus they are only relevant to the fixed-sample-size inference problems.

In contrast, the distributed sequential inference problem, where the complication arises from the successively arriving samples, is much less understood. Preliminarily, some existing works tackle this challenge by assuming that the consensus is reached before new samples are taken [6–9]. The more practical and interesting scenario is that the sampling and data aggregation processes take place simultaneously. Under this setup, [10] proposed the "consensus + innovation" approach for distributed recursive parameter estimation; [11] intended to track a stochastic process using a "running consensus" algorithm. The same method was then applied to the distributed locally optimal sequential test in [12], where the alternative parameter is assumed to be close to the null one.

While most of the above works focus on reaching consensus on the value of local decision statistics, limited light has been shed upon the expected sample size, i.e., stopping time, and error probabilities of the distributed sequential test. Recently, [13] combined the "consensus + innovation" approach and the SPRT to detect the mean-shift of Gaussian samples. However, their analysis is restricted to one specific testing problem, and does not reveal any asymptotic optimality. To fill that void, in this work, we investigate the consensus-based distributed SPRT for a generic hypothesis testing problem, and provides the sufficient conditions such that it achieves the asymptotically optimal test performance at all sensors.

The remainder of the paper is organized as follows. Section II discusses the problem formulation. In Section III, we investigate the asymptotic optimality of the consensus-algorithm-based distributed SPRT, together with the corresponding regularity conditions. In Section IV, simulation based on Gaussian samples is given to corroborate the theoretical results.

II. Problem Statement

Consider a network of K sensors that sequentially take samples in parallel. Conditioned on the hypothesis, these samples are independent and identically distributed at each sensor and independent across sensors, i.e.,

$$\mathcal{H}_0 : X_t^{(k)} \sim f_0^{(k)}(x) \;\; \text{v.s.} \;\; \mathcal{H}_1 : X_t^{(k)} \sim f_1^{(k)}(x),$$

where $k=1,2,\dots,K,\ t=1,2,\dots$ The log-likelihood ratio (LLR) and the cumulative LLR are denoted respectively as

$$s_t^{(k)} \triangleq \log \frac{f_1^{(k)}(X_t^{(k)})}{f_0^{(k)}(X_t^{(k)})} \text{ and } S_t^{(k)} \triangleq \sum_{j=1}^{t} s_j^{(k)}. \qquad (1)$$

The inter-sensor communication links determine the network topology, which can be represented by an undirected graph $\mathcal{G} \triangleq \{\mathcal{N}, \mathcal{E}\}$, with $\mathcal{N}$ being the set of sensors and $\mathcal{E}$ the set of edges. In addition, let $\mathcal{N}_k$ be the set of neighbouring sensors that are directly connected to sensor k, i.e., $\mathcal{N}_k \triangleq \{j \in \mathcal{N} : \{k,j\} \in \mathcal{E}\}$. In distributed sequential test, at every time slot t and each sensor k, the following actions take place in order: 1) taking a new sample, 2) exchanging messages with neighbours, and 3) deciding to stop for decision or to wait for more data at time $t+1$. Note that the first two actions, i.e., sampling and communication, will continue even after the local test at sensor k stops so that other sensors can still benefit from the same sample diversity. To evaluate the test performance, two performance metrics are used, namely, the expected stopping times $\mathbb{E}_i \mathsf{T}^{(k)},\ i=0,1$, and the type-I and type-II error probabilities, i.e., $\mathbb{P}_0\left(D^{(k)}=1\right)$ and $\mathbb{P}_1\left(D^{(k)}=0\right)$ respectively. As such, we aim to find the distributed sequential test such that the expected stopping times at sensors are minimized subject to the error probability constraints:

$$\begin{aligned} \min_{\{\mathsf{T}^{(k)}, D^{(k)}\}} \quad & \mathbb{E}_i\left(\mathsf{T}^{(k)}\right), \quad i=0,1 \\ \text{subject to} \quad & \mathbb{P}_0\left(D^{(k)}=1\right) \le \alpha, \\ & \mathbb{P}_1\left(D^{(k)}=0\right) \le \beta, \quad k=1,2,\dots,K. \end{aligned} \qquad (2)$$

Solving (2) at the same time for $k=1,2,\dots,K$ is a formidable task except for some special cases (for example, the fully connected or fully disconnected network); therefore the asymptotically optimal solution is the next best goal to pursue. We first introduce the widely-adopted definition for the asymptotic optimality [14]:

Definition 1. *Let $\mathsf{T}^\star$ be the stopping time of the optimum sequential test that satisfies the two error probability constraints with equality. Then, as the Type-I and Type-II error probabilities $\alpha,\beta \to 0$, the sequential test that satisfies the error probability constraints with stopping time T is said to be order-2 asymptotically optimal if*

$$0 \le \mathbb{E}_i(\mathsf{T}) - \mathbb{E}_i(\mathsf{T}^\star) = O(1).$$

In the ideal case, recalling the optimality of SPRT [1], the globally optimal performance for (2) is achieved by the so-called centralized SPRT (C-SPRT) when the network is fully connected, i.e.,

$$\mathsf{T}_c \triangleq \min\left\{t : S_t \triangleq \frac{1}{K}\sum_{k=1}^{K} S_t^{(k)} \notin (-A,B)\right\},\ D_c \triangleq \mathbb{1}_{\{S_{\mathsf{T}_c} \ge B\}}, \qquad (3)$$

where $\{A,B\}$ are constants chosen such that the constraints in (2) are satisfied with equalities. The asymptotic performance for the CSPRT as the error probabilities go to zero can be readily characterized by the following result [2]:

$$\mathbb{E}_1(\mathsf{T}_c) = \frac{-\log\alpha}{\sum_{k=1}^{K}\mathcal{D}_1^{(k)}} \text{ and } \mathbb{E}_0(\mathsf{T}_c) = \frac{-\log\beta}{\sum_{k=1}^{K}\mathcal{D}_0^{(k)}}, \qquad (4)$$

where $\mathcal{D}_i^{(k)} \triangleq \mathbb{E}_i\left(\log\frac{f_i^{(k)}(X)}{f_{1-i}^{(k)}(X)}\right)$ is the Kullback-Leibler divergence (KLD) at sensor k.

In reality, the network is often a sparse one, far from being fully connected. Nevertheless, if any distributed sequential test attains the globally optimal performance in (4) only with a constant difference, it is regarded to be asymptotically optimal according to the Definition 1. In next section, we will analyze the consensus-based SPRT and provides the sufficient conditions such that it achieves the order-2 asymptotic optimality.

III. Consensus-Based Sequential Probability Ratio Test

In this section, we consider the distributed SPRT based on the communication protocol known as the consensus algorithm, in which the sensors aggregate their local decision statistics. Denoting the decision statistic at sensor k and time t as $\eta_t^{(k)}$, then during every time slot t, the consensus algorithm is carried out as follows:

1) Take a new sample, and add the LLR $s_t^{(\ell)}$ to the local decision statistic from previous time, resulting in the intermediate statistic
$$\widetilde{\eta}_t^{(k)} = \eta_{t-1}^{(k)} + s_t^{(k)}. \qquad (5)$$
2) Every sensor exchanges its local intermediate statistic $\widetilde{\eta}_t^{(k)}$ with the neighbours, and updates it as the weighted sum of the available statistics from the neighbours, i.e.,
$$\widetilde{\eta}_t^{(k)} = w_{k,k}\,\widetilde{\eta}_t^{(k)} + \sum_{\ell\in\mathcal{N}_k} w_{\ell,k}\,\widetilde{\eta}_t^{(\ell)}, \qquad (6)$$
where the weight coefficients $w_{i,j}$ will be specified later.
3) Go to Step 1) for the next sampling time slot $t+1$.

To express the consensus algorithm in a compact form, we define the following vectors:

$$\boldsymbol{\eta}_t \triangleq [\eta_t^{(1)}, \eta_t^{(2)}, \dots, \eta_t^{(K)}]^T,\ \boldsymbol{s}_t \triangleq [s_t^{(1)}, s_t^{(2)}, \dots, s_t^{(K)}]^T,$$

such that the decision statistic vector evolves over time as

$$\boldsymbol{\eta}_t = \boldsymbol{W}\left(\boldsymbol{\eta}_{t-1} + \boldsymbol{s}_t\right), \quad \text{with } \boldsymbol{\eta}_0 = \boldsymbol{0}, \qquad (7)$$

where the matrix $\boldsymbol{W} \triangleq (w_{i,j}) \in \mathbb{R}^{K\times K}$ is formed by $w_{i,j}$'s defined in (6).

As such, the consensus-algorithm-based distributed SPRT (CA-DSPRT) at sensor k can be implemented with the following stopping time and decision rule:

$$\mathsf{T}_{\text{ca}}^{(k)} \triangleq \inf\left\{t : \eta_t^{(k)} \notin (-A,B)\right\}, \quad D_{\text{ca}}^{(k)} \triangleq \mathbb{1}_{\left\{\eta^{(k)}_{\mathsf{T}_{\text{ca}}^{(k)}} \ge B\right\}} \quad (8)$$

where $\{A,B\}$ are chosen to satisfy the error probability constraints.

Note that (7) resembles the consensus algorithm in the *fixed-sample-size test* [12], where no innovation are introduced, i.e., $\boldsymbol{\eta}_t = \boldsymbol{W}\boldsymbol{\eta}_{t-1}$. In that case, under certain regularity conditions for $\boldsymbol{W}$, consensus is reached in the sense $\boldsymbol{\eta}_t \rightarrow \left[\frac{1}{K}\sum_{i=1}^K \eta_0^{(k)}, \ldots, \frac{1}{K}\sum_{i=1}^K \eta_0^{(k)}\right]^T$ as $t \rightarrow \infty$. In contrast, with the new samples constantly arriving, how such a message-passing protocol can affect the *sequential test* at each sensor is of great interest both from theory and practice point of view.

To begin with, we first impose the following two conditions on the weight matrix $\boldsymbol{W}$ and the distribution of LLR respectively.

Condition 1. *The weight matrix $\boldsymbol{W}$ satisfies*

$$\boldsymbol{W}\mathbf{1} = \mathbf{1},\ \ \mathbf{1}^T\boldsymbol{W} = \mathbf{1}^T,\ \ 0 < \sigma_2(\boldsymbol{W}) < 1,$$

where $\mathbf{1}$ is an all-one vector, and $\sigma_i(\boldsymbol{W})$ denotes the ith singular value of $\boldsymbol{W}$.

Condition 2. *The LLR for the hypothesis testing problem satisfies that $\mathbb{E}_i\left(e^{K\sqrt{K}|s_j^{(k)}|}\right)$ is bounded for $i \in \{0,1\}$, $k = 1,\ldots,K$, where $s_j^{(k)}$ is the LLR defined in* (1).

The first condition essentially regulates the network topology and weight coefficients in (7). If we further require $w_{i,j} \geq 0$, then Condition 1 is equivalent to $\boldsymbol{W}$ being doubly stochastic. The second condition regulates the tail distribution of the LLR at each sensor. Provided that the testing problem, network topology, and weight coefficients satisfies the above conditions, then we have the main theorem for this paper.

Theorem 1. *Given that Conditions 1-2 are satisfied, the asymptotic performance of the CA-DSPRT as $\alpha, \beta \rightarrow 0$ is characterized by*

$$\mathbb{E}_1\left(\mathsf{T}_{ca}^{(k)}\right) \leq \frac{-\log\alpha}{\sum_{k=1}^K \mathcal{D}_1^{(k)}} + \frac{\sigma_2(\boldsymbol{W})}{1-\sigma_2(\boldsymbol{W})}O(1),$$
$$\mathbb{E}_0\left(\mathsf{T}_{ca}^{(k)}\right) \leq \frac{-\log\beta}{\sum_{k=1}^K \mathcal{D}_0^{(k)}} + \frac{\sigma_2(\boldsymbol{W})}{1-\sigma_2(\boldsymbol{W})}O(1). \quad (9)$$

Theorem 1 can be readily proved by invoking the following two key lemmas. While we omit the proofs for these two lemmas due to the space limitation, interested readers can refer to [15] for a detailed development.

Lemma 1. *All sensors achieve the same expected stopping time in the asymptotic regime:*

$$\mathbb{E}_1\left(\mathsf{T}_{ca}^{(k)}\right) = \frac{B}{\sum_{k=1}^K \mathcal{D}_1^{(k)}/K} + O(1),$$
$$\mathbb{E}_0\left(\mathsf{T}_{ca}^{(k)}\right) = \frac{A}{\sum_{k=1}^K \mathcal{D}_0^{(k)}/K} + O(1), \quad (10)$$

for $k = 1,2,\ldots,K$, as $A, B \rightarrow \infty$.

Lemma 1 characterizes how the expected sample sizes of the CA-DSPRT vary as the decision thresholds go to infinity. The next lemma relates the error probabilities of the CA-DSPRT in the same asymptotic regime to the decision thresholds.

Lemma 2. *The error probabilities of CA-DSPRT in the asymptotic regime as $A, B \rightarrow \infty$ at each sensor is bounded above by*

$$\log\alpha = \log\mathbb{P}_0\left(D_{ca}^{(k)} = 1\right) \leq -KB + O(1),$$
$$\log\beta = \log\mathbb{P}_1\left(D_{ca}^{(k)} = 0\right) \leq -KA + O(1). \quad (11)$$

Combining (10) and (11) leads to (9). By recalling (4), and the fact that $\sigma_2(\boldsymbol{W})$ is a constant, we have

$$\mathbb{E}_i\left(\mathsf{T}_{\text{ca}}^{(k)}\right) - \mathbb{E}_i(\mathsf{T}_c) \leq \frac{\sigma_2(\boldsymbol{W})}{1-\sigma_2(\boldsymbol{W})}O(1), \quad (12)$$

implying that CA-DSPRT provides order-2 asymptotically optimal solution to (2). Note that $k = 1,2,\ldots,K$, thus this conclusion applies to all sensors.

IV. Example and Numerical Results

In this section, we examine the performance of CA-DSPRT using numerical simulations. The problem of detecting the mean-shift of Gaussian samples will be considered. Without loss of generality, the variance is assumed to be one in the hypothesis testing problem, i.e.,

$$\mathcal{H}_0 : X_t^{(k)} \sim \mathcal{N}(0,1), \ \text{ v.s. } \ \mathcal{H}_1 : X_t^{(k)} \sim \mathcal{N}(\mu,1).$$

Accordingly, the LLR at sensor k is given by

$$s_t^{(k)} = X_t^{(k)}\mu - \frac{\mu^2}{2} \sim \begin{cases} \mathcal{N}\left(-\frac{\mu^2}{2}, \mu^2\right), & \text{under } \mathcal{H}_0, \\ \mathcal{N}\left(\frac{\mu^2}{2}, \mu^2\right), & \text{under } \mathcal{H}_1, \end{cases} \quad (13)$$

with KLDs equal to $\mathcal{D}_0^{(k)} = \mathcal{D}_1^{(k)} = \frac{\mu^2}{2}$. The LLR (13) can be readily shown to satisfy the Condition 2.

There are various methods to choose $\boldsymbol{W}$ such that Condition 1 can be satisfied, one of which is assigning equal weights to the data from neighbours [13, 16]. Specifically, the message-passing protocol (6) becomes

$$\widetilde{\eta}_t^{(k)} = (1 - |\mathcal{N}_k|\delta)\ \widetilde{\eta}_t^{(k)} + \delta \sum_{\ell\in\mathcal{N}_k} \widetilde{\eta}_t^{(\ell)},$$
$$= \widetilde{\eta}_t^{(k)} + \delta \sum_{\ell\in\mathcal{N}_k} \left(\widetilde{\eta}_t^{(\ell)} - \widetilde{\eta}_t^{(k)}\right).$$

As such, the weight matrix admits

$$\boldsymbol{W} = \mathbf{I} - \delta(\underbrace{\boldsymbol{D} - \boldsymbol{A}}_{\boldsymbol{L}}), \quad (14)$$

where $\boldsymbol{A}$ is the adjacent matrix, whose entries $a_{i,j} = 1$ if and only if $\{i,j\} \in \mathcal{E}$, and $\boldsymbol{D} \triangleq \text{diag}\{|\mathcal{N}_1|, |\mathcal{N}_2|, \ldots, |\mathcal{N}_K|\}$ is the called the degree matrix. Their difference is called the Laplacian matrix $\boldsymbol{L}$ which is positive semidefinite. First, $\boldsymbol{W}\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^T\boldsymbol{W} = \mathbf{1}^T$ hold for any value of δ due to the definition of $\boldsymbol{L}$ (i.e., $\boldsymbol{L}\mathbf{1} = \mathbf{0}$ and $\mathbf{1}^T\boldsymbol{L} = \mathbf{0}^T$). Second, note that $\boldsymbol{W}$ in (14) is a symmetric matrix, whose second largest singular value $\sigma_2(\boldsymbol{W}) = \max\{1 - \delta\lambda_{n-1}(\boldsymbol{L}), \delta\lambda_1(\boldsymbol{L}) - 1\} < 1$, if and only if $0 < \delta < \frac{2}{\lambda_1(\boldsymbol{L})}$. Within this interval, we set $\delta = \frac{2}{\lambda_1(\boldsymbol{L}) + \lambda_{n-1}(\boldsymbol{L})}$ such that the constant terms in Theorem 1 are minimized, or equivalently, $\sigma_2(\boldsymbol{W})$ is minimized. Now that Condition 1-2 are both satisfied, according to Theorem

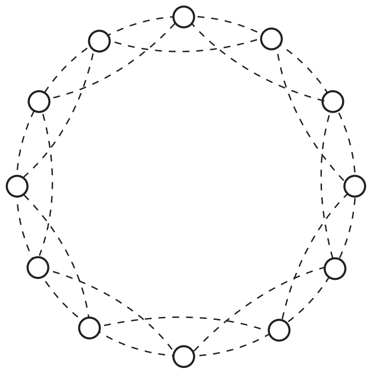

Fig. 1. The sensor network represented by a graph $\mathcal{G}(12, 2)$.

1, CA-DSPRT achieves the order-2 asymptotically optimal performance at all sensors.

In the experiment, the alternative mean is set as $\mu = 0.3$; the network as depicted in Fig. 1 will be used, where each sensor is connected to the sensors within range 2; and the resulting weight matrix (14) has $\sigma_2(\boldsymbol{W}) = 0.6511$. Since sensors in this network have the identical sample distributions, adjacent sensors and message-passing weights, they should result in the same test performance. Thus, we only plot the performance at sensor 1 for illustrative purpose. Also note that, due to the symmetry of the statistic distribution under $\mathcal{H}_0$ and $\mathcal{H}_1$, it is sufficient to plot the performance under one hypothesis, while the other follows identically.

Fig. 2 illustrates how the error probability and expected sample size change with the threshold in CA-DSPRT. Specifically, Fig. 2-(a) shows that the error probability of the CA-DSPRT (marked in blue circles) aligns parallel to the solid line as expected by Lemma 2, which also corresponds to the C-SPRT method. Fig. 2-(b) shows that the expected sample size of CA-DSPRT aligns parallel to that of the C-SPRT as the threshold increases, which agrees with Lemma 1. Moreover, for comparison, we will also plot the analytical bounds derived in [13] for the error probabilities of CA-DSPRT under the Gaussian-mean-shift setting. They are referred to as the "existing approximation" in the figures.

Combining 2-(a) and (b) gives the performance curves as shown in Fig. 3. First, CA-DSPRT only deviates from the globally optimal performance by a constant margin as $A, B \to \infty$, exhibiting the order-2 asymptotic optimality as stated in Theorem 1. Again, it is seen that the curve by [13] again substantially deviates from the true performance; therefore it does not reveal the asymptotic optimality of $\mathsf{T}_{\text{ca}}^{(k)}$.

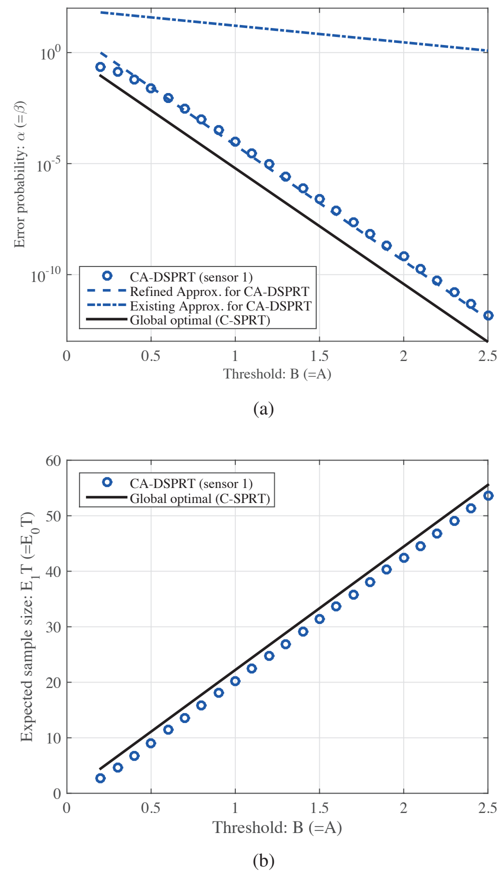

Fig. 2. The false alarm probability and expected sample size in terms of the threshold B.

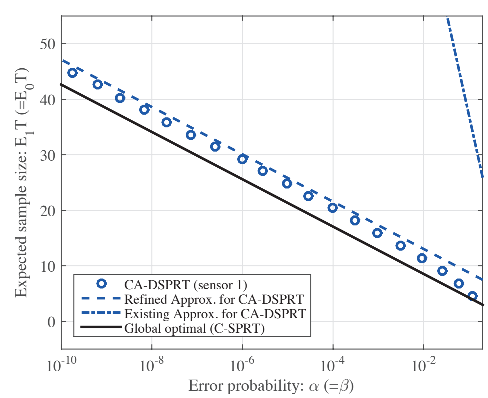

Fig. 3. The expected sample size subject to the error probability.

References

[1] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *The Annals of Mathematical Statistics*, vol. 19, no. 3, pp. 326–339, 1948.

[2] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential analysis: Hypothesis testing and changepoint detection*. Boca Raton: CRC Press, 2014.

[3] M. H. DeGroot, "Reaching a consensus," *J. Am. Statist. Assoc.*, vol. 69, no. 345, pp. 118–121, 1974.

[4] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. 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," *Proc. IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.

[6] R. Olfati-Saber, "Distributed Kalman filter with embedded consensus filters," in *Proc. 44th IEEE Conf. Decision Control Eur. Control Conf. CDC-ECC'05*, 12-15 Dec. 2005, pp. 8179–8184.

[7] U. A. Khan and J. M. F. Moura, "Distributing the Kalman filter for large-scale systems," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4919–4935, Oct. 2008.

[8] A. A. Amini and X. Nguyen, "Sequential detection of multiple change points in networks: A graphical model approach," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5824–5841, Sep. 2013.

[9] Z. Guo, S. Li, X. Wang, and W. Heng, "Distributed point-based Gaussian approximation filtering for forecasting-aided state estimation in power systems," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 2597–2608, Jul. 2016.

[10] S. Kar and J. M. F. Moura, "Consensus + innovations distributed inference over networks," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 99–109, May 2013.

[11] P. Braca, S. Marano, and V. Matta, "Enforcing consensus while monitoring the environment in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3375–3380, Jul. 2008.

[12] P. Braca, S. Marano, V. Matta, and P. Willett, "Asymptotic optimality of running consensus in testing binary hypohteses," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 814–825, Feb. 2010.

[13] A. K. Sahu and S. Kar, "Distributed sequential detection for Gaussian shift-in-mean hypothesis testing," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 89–103, Jan. 2016.

[14] G. Fellouris and G. V. Moustakides, "Decentralized sequential hypothesis testing using asynchronous communication," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 534–548, Jan. 2011.

[15] S. Li, X. Li, X. Wang, and J. Liu, "Order-2 asymptotic optimality of the fully distributed sequential hypothesis test," [Online]. Available: http://arxiv.org/abs/1606.04203.

[16] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 52, pp. 65–78, 2004.