

# Distributed Detection over Time Varying Networks: Large Deviations Analysis

Dragana Bajović, Dušan Jakovetić, João Xavier, Bruno Sinopoli and José M. F. Moura

**Abstract**—We apply large deviations theory to study asymptotic performance of *running consensus distributed detection* in sensor networks. Running consensus is a stochastic approximation type algorithm, recently proposed. At each time step  $k$ , the state at each sensor is updated by a local averaging of the sensor’s own state and the states of its neighbors (consensus) and by accounting for the new observations (innovation). We assume Gaussian, spatially correlated observations. We allow the underlying network be time varying, provided that the graph that collects the union of links that are online at least once over a finite time window is connected. This paper shows through large deviations that, under stated assumptions on the network connectivity and sensors’ observations, the running consensus detection asymptotically approaches in performance the optimal centralized detection. That is, the Bayes probability of detection error (with the running consensus detector) decays exponentially to zero as  $k \rightarrow \infty$  at the Chernoff information rate—the best achievable rate of the asymptotically optimal centralized detector.

## I. INTRODUCTION

We apply large deviations to study the asymptotic performance of distributed detection in sensor networks. Each node in the network senses the environment and cooperates locally with its neighbors to decide between the two hypothesis,  $H_1$  and  $H_0$ . The nodes are connected by a generic, time varying network, and there is no fusion center. Specifically, we consider distributed detection

This work is partially supported by: the Carnegie Mellon—Portugal Program under a grant from the Fundação para a Ciência e Tecnologia (FCT) from Portugal; by FCT grants SIPM PTDC/EEA-ACR/73749/2006; and by ISR/IST plurianual funding (POSC program, FEDER). Work of José M. F. Moura is partially supported by NSF under grants CCF-1011903 and CCF-1018509, and by AFOSR grant FA95501010291. Dragana Bajović and Dušan Jakovetić hold fellowships from FCT.

D. Bajović and Dušan Jakovetić are with the Institute for Systems and Robotics (ISR), Instituto Superior Técnico (IST), Lisbon, Portugal, and with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA dbajovic@andrew.cmu.edu, djakovet@andrew.cmu.edu

J. Xavier is with the Institute for Systems and Robotics (ISR), Instituto Superior Técnico (IST), Lisbon, Portugal jxavier@isr.ist.utl.pt

B. Sinopoli and José M. F. Moura are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA brunos@ece.cmu.edu, moura@ece.cmu.edu

via running consensus<sup>1</sup> that has been recently proposed in [2]. With running consensus, at each time  $k$ ,  $N$  nodes update their decision variables by: 1) incorporating new observation (innovation step); and 2) mixing their decision variables locally with the neighbors (consensus step).

We allow the underlying communication graph be (deterministically) time varying; but we assume that the graph that collects all communication links that are online (at least once) within a finite time window  $B$  is connected. We assume Gaussian, spatially correlated, time-uncorrelated sensors’ observations. Under stated assumptions on the network connectivity and the sensors’ observations, we show that the running consensus distributed detector is asymptotically optimal, as the number of observations  $k$  goes to infinity. That is, the running consensus distributed detector asymptotically approaches the performance of the optimal centralized detector. We apply large deviations to study the asymptotic performance of both the (asymptotically) optimal centralized detector, which collects observations from all nodes  $i$  at each time  $k$ , and the running consensus detector. For both detectors, the Bayes probability of error decays as  $e^{-kC}$ , where  $C$  is the Chernoff distance between the distributions of the  $N \times 1$  observation vectors under the two hypothesis, i.e., the Chernoff information.

We now briefly review the existing work on distributed detection. Distributed detection has been extensively studied. Prior work studies parallel fusion architectures (see, e.g., [3], [4], [5], [6], [7], [8]) where all nodes communicate with a fusion node. Also, consensus-based detection schemes have been studied (with no fusion node) in, for example, [9], [10], [11], where nodes in the network: 1) collect measurements; and 2) *subsequently* run the consensus algorithm to fuse their detection rules. The running consensus distributed detection has been proposed in [12]. Running consensus is different from classical consensus detection, as it incorporates new observations at each time step  $k$ , in real time; thus,

<sup>1</sup>The running consensus algorithm is a type of recursive stochastic approximation algorithm, see, e.g., [1]. Reference [1] studies more general stochastic approximation type algorithms in the context of distributed estimation. We use the algorithm in form given in [2] and will refer to it as running consensus.

unlike classical consensus, no delay is introduced from collecting observations to reaching consensus.

We now comment on the differences between this paper and reference [12], which also studies asymptotic optimality of distributed detection via running consensus. Reference [12] considers the Neyman-Pearson framework, while we adopt the Bayesian framework. Reference [12] considers that, as the number of observations  $k$  grows, the distribution means under the two hypothesis become closer and closer, at the rate of  $1/\sqrt{k}$ ; consequently, as  $k \rightarrow \infty$ , there is an asymptotic, non zero, probability of miss, and asymptotic, non zero, probability of false alarm. In contrast, we assume that the distributions do not change with  $k$  (do not approach each other,) and the Bayes probability of error decays to zero; we then examine the rate of decay of the Bayes error probability. Further, reference [12] assumes that the observations at different sensors are independent identically distributed, with generic distribution, while we assume Gaussian; however, we allow for spatial correlation among observations—a well-suited assumption, e.g., for densely deployed wireless sensor networks (WSNs). Reference [12] studies the case where the underlying network is randomly varying; we consider deterministically time varying network.

**Paper organization.** Section II reviews the large deviations results and the Chernoff lemma in hypothesis testing. Section III explains data and network models that we assume. Section IV introduces the (asymptotically) optimal centralized detection, as if there was a fusion node and its detection performance. Section V shows that the distributed running consensus detector asymptotically approaches in performance the optimal centralized detector. Finally, section VI summarizes the paper.

## II. BACKGROUND

In this section, we briefly review standard large deviations analysis for binary hypothesis testing and standard asymptotic results (in particular, Chernoff lemma) in binary hypothesis testing. We will later use these results throughout the paper.

### A. Binary hypothesis testing problem: Log-likelihood ratio test

Consider the sequence of independent identically distributed (i.i.d.)  $d$ -dimensional random vectors (observations)  $y(k)$ ,  $k = 1, 2, \dots$ , and the binary hypothesis testing problem of deciding whether the probability measure (law) generating  $y(k)$  is  $\nu_0$  (under hypothesis  $H_0$ ) or  $\nu_1$  (under  $H_1$ ). Assume that  $\nu_1$  and  $\nu_0$  are mutually absolutely continuous, distinguishable measures. Based on the observations  $y(1), \dots, y(k)$ , formally, a decision test  $T$  is a sequence of maps  $T_k : \mathbb{R}^{kd} \rightarrow \{0, 1\}$ ,  $k = 1, 2, \dots$ ,

with the interpretation that  $T_k(y(1), \dots, y(k)) = l$  means that  $H_l$  is decided,  $l = 0, 1$ . Specifically, consider the log-likelihood ratio (LLR) test to decide between  $H_0$  and  $H_1$ , where  $T_k$  is given as follows:

$$\mathcal{D}(k) := \frac{1}{k} \sum_{j=1}^k \log \frac{d\nu_1}{d\nu_0}(y(j)) \quad (1)$$

$$T_k = \mathcal{I}_{\{\mathcal{D}(k) > \gamma_k\}}. \quad (2)$$

Here  $L(k) := \log \frac{d\nu_1}{d\nu_0}(y(k))$  is the LLR (given by the Radon-Nikodym derivative of  $\nu_1$  with respect to  $\nu_0$  evaluated at  $y(k)$ ),  $\gamma_k$  is a chosen threshold, and  $\mathcal{I}_A$  is the indicator of event  $A$ . The LLR test with threshold  $\gamma_k = 0$ ,  $\forall k$ , is asymptotically optimal in the sense of Bayes probability of error decay rate, as will be explained in next subsection (II-B).

### B. Log-likelihood ratio test: Large deviations

This subsection studies large deviations for the LLR decision test with decision variables  $\mathcal{D}(k)$  given in eqn. (1). The large deviations analysis will be very useful in estimating the exponential rate at which the Bayes probability of error decays and in showing the asymptotic optimality of the distributed running consensus detector. We first give the definition of the large deviations principle [13].

#### Definition 1 (Large deviations principle (LDP))

Consider a sequence of real valued random variables  $\{\Theta(k)\}_{k=1}^\infty := \{\Theta(k)\}$  and denote by  $\theta_k$  the probability measure of  $\Theta(k)$ . We say that the sequence of measures  $\{\theta_k\}$  satisfies the LDP with a rate function  $\mathcal{J} : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$  if the following holds:

- 1) For any closed, measurable set  $F \subset \mathbb{R}$ :

$$\limsup_{k \rightarrow \infty} \frac{1}{k} \log \theta_k(F) \leq - \inf_{t \in F} \mathcal{J}(t)$$

- 2) For any open, measurable set  $G \subset \mathbb{R}$ :

$$\liminf_{k \rightarrow \infty} \frac{1}{k} \log \theta_k(G) \geq - \inf_{t \in G} \mathcal{J}(t).$$

It can be shown that the sequence of LLR's  $\{L(k)\}$ , conditioned on  $H_l$ ,  $l = 0, 1$ , is i.i.d. Denote by  $\mu_k^{(l)}$  the probability measure of  $\mathcal{D}(k)$  under hypothesis  $H_l$ . Using Cramér's theorem ([13]), it can be shown that the sequence of measures  $\{\mu_k^{(l)}\}$ ,  $l = 0, 1$ , satisfies the LDP with good<sup>2</sup> rate function:

$$\Lambda_{(l)}^*(t) = \sup_{\lambda \in \mathbb{R}} (\lambda t - \Lambda_{(l)}(\lambda)), \quad (3)$$

where  $\Lambda_{(l)}(\cdot)$  is the log-moment generating function of  $L(k)$  under hypothesis  $H_l$ :

$$\Lambda_{(l)}(\lambda) = \log \mathbb{E} \left[ e^{\lambda L(k)} | H_l \right]. \quad (4)$$

<sup>2</sup>Goodness of rate function is compactness of its sublevel sets.

That is, the rate function  $\Lambda_{(l)}^*(t)$  is the Fenchel-Legendre (F-L) ([13]) transform of the log-moment generating function of  $L(k)$  under  $H_l$ . It can be shown that  $\Lambda_{(1)}^*(t) = \Lambda_{(0)}^*(t) - t$ . We summarize this result in the following theorem, e.g., [13]:

*Theorem 2* The sequence of measures  $\{\mu_k^{(l)}\}$  of  $\mathcal{D}(k)$  under  $H_l$  satisfies the LDP with good rate function given by eqn. (3).

### C. Asymptotic Bayes detection performance: Chernoff lemma

We adopt the Bayes minimum probability of error detection. Denote by  $P^e(k)$  the Bayes probability of error after  $k$  samples are processed:

$$P^e(k) = P(H_0)\alpha(k) + P(H_1)\beta(k), \quad (5)$$

where  $P(H_l)$  are the prior probabilities,  $\alpha(k) := P(\mathcal{D}(k) > \gamma_k | H_0)$  and  $\beta(k) := P(\mathcal{D}(k) \leq \gamma_k | H_1)$  are, respectively, the probability of false alarm and the probability of miss, and  $\gamma_k$  is the test threshold.

We will be interested in the rate at which the Bayes probability of error decays to zero as the number of observations  $k$  goes to infinity. Also, as auxiliary results, we will need the rates at which  $\alpha(k)$  and  $\beta(k)$  go to zero as  $k \rightarrow \infty$ . That is, we will be interested in the following quantities:

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log P^e(k) \quad (6)$$

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log \alpha(k) \quad (7)$$

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log \beta(k). \quad (8)$$

Theorem 4 ([13]) states that, among all possible decision tests, the LLR test with zero threshold minimizes (6). This result is a corollary of the Theorem 3 ([13]), that asserts that, for a LLR test with fixed threshold  $\gamma_k = \gamma$ ,  $\alpha(k)$  and  $\beta(k)$  indeed (simultaneously) decay to zero exponentially; also, Theorem 3 expresses the exponential rate of decay in terms of the rate functions defined in eqns. (7) and (8). Before stating the Theorem, define  $\bar{L}_{(l)} := \mathbb{E}(L(k) | H_l)$ ,  $l = 0, 1$ .

*Theorem 3* The LLR test with constant threshold  $\gamma_k = \gamma$ ,  $\gamma \in (\bar{L}_{(0)}, \bar{L}_{(1)})$  satisfies:

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log \alpha(k) = -\Lambda_{(0)}^*(\gamma) < 0 \quad (9)$$

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log \beta(k) = \gamma - \Lambda_{(0)}^*(\gamma) < 0. \quad (10)$$

*Theorem 4 (Chernoff lemma)* If  $P(H_0) \in (0, 1)$ , then:

$$\inf_T \liminf_{k \rightarrow \infty} \left\{ \frac{1}{k} \log P^e(k) \right\} = -\Lambda_{(0)}^*(0), \quad (11)$$

where the infimum over all possible tests  $T$  is attained for the LLR test with  $\gamma_k = 0$ ,  $\forall k$ .

The quantity  $\Lambda_{(0)}^*(0) = \Lambda_{(1)}^*(0)$  is called the Chernoff distance between the distributions of  $y(k)$  under  $H_0$  and  $H_1$ , or Chernoff information, [13].

**Asymptotically optimal test.** We introduce the following definition of the asymptotically optimal test.

*Definition 5* The decision test  $T$  is asymptotically optimal if it attains the infimum in eqn. (11).

We will show that, for the distributed Gaussian hypothesis testing over random networks, the running consensus is asymptotically optimal in the sense of Definition 5.

## III. DISTRIBUTED DETECTION MODEL: DATA AND NETWORK MODELS

This section describes: 1) the data model (subsection III-A), i.e., the observation model at each sensor in the network; and 2) the model of the network through which the sensors cooperate with the running consensus distributed detection algorithm (subsection III-B). The distributed detection algorithm is detailed in Section V.

### A. Data model

We consider Gaussian binary hypothesis testing in spatially correlated noise. The sensors operate (in terms of sensing and communication) synchronously, at discrete time steps  $k$ . At time  $k$ , sensor  $i$  measures (scalar)  $y_i(k)$ . Collect the sensor measurements in a vector  $y(k) = (y_1(k), y_2(k), \dots, y_N(k))^T$ , where  $N$  is the total number of sensors. Nature can be in one of two possible states:  $H_1$ -event occurring (e.g., target present); and  $H_0$ -event not occurring (e.g., target absent.) We assume the following distribution model for the vector  $y(k)$ :

$$\text{under } H_l : y(k) = m_l + \zeta(k), \quad l = 0, 1, \quad (12)$$

where  $m_l$  is the (constant) signal under hypothesis  $H_l$ , and  $\zeta(k)$  is zero mean Gaussian additive noise. We assume that  $\{\zeta(k)\}$  is an independent identically distributed (i.i.d.) sequence of  $N \times 1$  random vectors with distribution  $\zeta(k) \sim \mathcal{N}(0, S)$ , where  $S$  is a (positive definite) covariance matrix. Thus, with our model, the noise is temporally independent, but can be spatially correlated. Spatial correlation should be taken into account due to, for example, dense deployment of wireless sensor networks, while it is still reasonable to assume that the observations are independent along time. (Conditioned to  $H_l$ ,  $\{y(k)\}$  are i.i.d. with the distribution  $\mathcal{N}(m_l, S)$ .)

### B. Network model and data mixing model

We consider distributed detection via running consensus where each node at a time  $k$ : 1) measures  $y_i(k)$ ; 2) exchanges its current decision variable (denote it by

$x_i(k)$  with its neighbors; and 3) performs a weighted average of its own decision variable and the neighbors' decision variables. The network connectivity is assumed time varying. The weighted averaging, at each time  $k$ , as with the standard consensus algorithm, is described by the  $N \times N$  weight matrix  $W(k)$ . We assume  $W(k)$  is a symmetric, stochastic matrix (it has nonnegative entries and the rows sum to 1.) The weight matrix  $W(k)$  respects the sparsity pattern of the network, i.e.,  $W_{ij}(k) = 0$ , if the link  $\{i, j\}$  is down at time  $k$ . We define also the undirected graph  $G(k) = (\mathcal{V}, \mathcal{E}(k))$ , where  $\mathcal{V}$  is the set of nodes with cardinality  $|\mathcal{V}| = N$ , and  $\mathcal{E}(k)$  is the set of undirected edges that are online at time  $k$ . Formally,  $\mathcal{E}(k) = \{\{i, j\} : i < j, W_{ij}(k) > 0\}$ . Define also  $J := (1/N)\mathbf{1}\mathbf{1}^T$ , where  $\mathbf{1}$  is  $N \times 1$  vector with unit entries. We now summarize the assumptions on the matrices  $\{W(k)\}$  and the graphs  $G(k)$ :

*Assumption 6* For the sequence of matrices  $\{W(k)\} = \{W(k)\}_{k=1}^\infty$ , we assume the following:

- 1)  $W(k)$  is symmetric and stochastic,  $\forall k$ .
- 2) There exists a scalar  $W_{\min} \in (0, 1)$ , such that *i*)  $W_{ii}(k) \geq W_{\min}$ ,  $\forall i, \forall k$ ; and *ii*)  $\forall k, W_{ij}(k) \geq W_{\min}$ , if  $i \neq j$  and  $\{i, j\} \in \mathcal{E}(k)$ .
- 3) There exists an integer  $1 \leq B < +\infty$ , such that,  $\forall k$ , the graph  $(\mathcal{V}, \cup_{l=k+1}^{k+B} \mathcal{E}(l))$  is connected.

Assumption 6-3) says that nodes should communicate sufficiently often (within finite time windows,) such that the network provides sufficiently fast information flow.

#### IV. CENTRALIZED DETECTION: BAYES OPTIMAL TEST

We first consider the centralized detection scenario, as if there was a fusion node that collects and processes all sensor observations. The decision variable  $\mathcal{D}(k)$  and the LLR decision test are given by eqns. (1) and (2), where now, under the data assumptions in subsection III-A:

$$L(k) = (m_1 - m_0)^T S^{-1} \left( y(k) - \frac{m_1 + m_0}{2} \right) \quad (13)$$

Conditioned on either hypothesis  $H_1$  and  $H_0$ ,  $L(k) \sim \mathcal{N} \left( m_L^{(l)}, \sigma_L^2 \right)$ , where

$$m_L^{(l)} = \frac{(-1)^{l+1}}{2} (m_1 - m_0)^T S^{-1} (m_1 - m_0) \quad (14)$$

$$\sigma_L^2 = (m_1 - m_0)^T S^{-1} (m_1 - m_0). \quad (15)$$

Define the vector  $v \in \mathbb{R}^N$  as

$$v := S^{-1} (m_1 - m_0). \quad (16)$$

Then, the LLR  $L(k)$  can be written as follows:

$$L(k) = \sum_{i=1}^N v_i \left( y_i(k) - \frac{[m_1]_i + [m_0]_i}{2} \right) = \sum_{i=1}^N \eta_i(k), \quad (17)$$

where  $[m_l]_i$  denotes the  $i$ -th entry of vector  $m_l$ ,  $l = 0, 1$ . Thus, the LLR at time  $k$  is separable, i.e., the LLR is the sum of the terms  $\eta_i(k)$  that depend affinely on the individual observations  $y_i(k)$ . We will exploit this fact in subsection V-A to derive the distributed, running consensus, detection algorithm.

Applying Theorem 2 to the sequence  $\{\mathcal{D}(k)\}$  (under hypothesis  $H_l$ ,  $l = 0, 1$ ), we have that the sequence of measures of  $\mathcal{D}(k)$  satisfies the LDP with good rate function  $I_{(l)} : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$ , which, by evaluating the log-moment generating function of  $L(k)$  in (13) and its F-L transform, can be shown to be:

$$I_{(l)}(t) = \frac{(t - m_L^{(l)})^2}{2\sigma_L^2}, l = 0, 1. \quad (18)$$

We state this result as a Corollary 7.

*Corollary 7* The sequence  $\{\mathcal{D}(k)\}$ , under  $H_l$ ,  $l = 0, 1$ , satisfies the LDP with good rate function  $I_{(l)}(\cdot)$ , given by eqn. (18).

We remark that Theorem 4 also applies to the detection problem explained in subsection III-A. Denote by  $P_{\text{cen}}^e(k)$  the Bayes probability of error for the centralized detector (defined in section IV,) after  $k$  samples are processed. Due to the continuity of the rate functions in (18), it can be shown that:  $\liminf_{k \rightarrow \infty} \frac{1}{k} \log P_{\text{cen}}^e(k) = \limsup_{k \rightarrow \infty} \frac{1}{k} \log P_{\text{cen}}^e(k) = \lim_{k \rightarrow \infty} \frac{1}{k} \log P_{\text{cen}}^e(k)$ . Thus, Theorem 4 in this case simplifies to the following corollary:

*Corollary 8 (Chernoff lemma for the optimal centralized detector)* The LLR test with  $\gamma_k = 0$ ,  $\forall k$ , is asymptotically optimal in the sense of definition 5. Moreover, for the LLR test with  $\gamma_k = 0$ ,  $\forall k$ , we have:

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{1}{k} \log P_{\text{cen}}^e(k) &= -I_{(0)}(0) \\ &= -\frac{1}{8} (m_1 - m_0)^T S^{-1} (m_1 - m_0). \end{aligned} \quad (19)$$

**Remark.** The LLR test with zero threshold is optimal also in the finite time  $k$  regime, for all  $k$ , in the sense that it minimizes the Bayes probability of error, when the prior probabilities are  $P(H_0) = P(H_1) = 0.5$ . When the prior probabilities are not equal, the LLR test is also optimal, but the threshold  $\gamma_k$  will be different than zero.

## V. DISTRIBUTED DETECTION ALGORITHM

### A. Distributed detection via running consensus

We now present a distributed detection algorithm via running consensus. With this detection algorithm, no fusion node is required, and the underlying network is generic, time varying. The running consensus is proposed in [2], and it is a stochastic approximation type of algorithm (see [1]). Reference [2] studies the case when the observations of different sensors at a fixed time  $k$  are i.i.d. We extend the running consensus detection algorithm to the case of spatially correlated Gaussian observations.

With the running consensus distributed detector, each node  $i$  makes local decisions based on its local decision variable  $x_i(k)$ : If  $x_i(k) > 0$ , then  $H_1$  is accepted; if  $x_i(k) \leq 0$ , then  $H_0$  is accepted. At each time step  $k$ , the local decision variable at node  $i$  is improved two-fold: 1) by exchanging information with its immediate neighbors in the network; 2) by incorporating into the decision process the new local observation  $y_i(k)$ . Recall the definition of  $\eta_i(k)$  in eqn. (17). Specifically, the update of the local decision variable at node  $i$  is given by the following equation:

$$\begin{aligned} x_i(k+1) &= \frac{k}{k+1}W_{ii}(k)x_i(k) + \frac{k}{k+1} \sum_{j \in \Omega_i(k)} W_{ij}(k)x_j(k) + \frac{1}{k+1}N\eta_i(k+1), \\ & \quad k = 1, 2, \dots \\ x_i(1) &= N\eta_i(1) \end{aligned} \quad (20)$$

Here  $\Omega_i(k)$  is the (time varying) neighborhood of node  $i$  at time  $k$ , and  $W_{ij}(k)$  are the (time varying) averaging weights, defined together with the  $N \times N$  (time varying) matrices  $W(k) = [W_{ij}(k)]$  in subsection III-B. Let  $x(k) = (x_1(k), x_2(k), \dots, x_N(k))^T$  and  $\eta(k) = (\eta_1(k), \dots, \eta_N(k))^T$ . The algorithm in matrix form is given by:

$$\begin{aligned} x(k+1) &= \frac{k}{k+1}W(k)x(k) + \frac{1}{k+1}N\eta(k+1), \\ & \quad k = 1, 2, \dots \end{aligned} \quad (21)$$

$$x(1) = N\eta(1)$$

Recall the definition of the  $N \times 1$  vector  $v$  in (16). The sequence of  $N \times 1$  random vectors  $\{\eta(k)\}$ , conditioned to  $H_l$ , is i.i.d. Vector  $\eta(k)$  (under hypothesis  $H_l$ ,  $l = 0, 1$ ) is Gaussian with mean  $m_\eta^{(l)}$  and covariance  $S^\eta$ :

$$\begin{aligned} m_\eta^{(l)} &= (-1)^{(l+1)} \text{Diag}(v) \frac{1}{2}(m_1 - m_0) \\ S^\eta &= \text{Diag}(v) S \text{Diag}(v). \end{aligned} \quad (22)$$

$$(23)$$

Here  $\text{Diag}(v)$  is a diagonal matrix with the diagonal entries equal to the entries of  $v$ .

### B. Asymptotic optimality of the distributed detection algorithm

In this subsection, we present our main result, which states that the distributed detection via running consensus asymptotically achieves the performance of the optimal centralized detector, in the sense that it approaches the exponential error decay rate of the (asymptotically) optimal centralized detector.

Denote the probability measure of  $x_i(k)$  under hypothesis  $H_l$  with  $\chi_{i,k}^{(l)}$ . First, we show that the sequence of measures  $\{\chi_{i,k}^{(l)}\}$ , for all nodes  $i$ , satisfies the LDP with good rate function; the rate function for all nodes  $i$  is the same, and it is the same as the rate function of the optimal centralized detector in eqn. (18).

We prove that the sequence of measures for  $\{x_i(k)\}$  (under  $H_l$ ,  $l = 0, 1$ ) satisfies the LDP using the Gärtner-Ellis Theorem from large deviations theory, see [13]. We now state Theorem 9.

*Theorem 9* Let assumption 6 hold. The sequence of measures  $\{\chi_{i,k}^{(l)}\}$ , for all nodes  $i$ , satisfies the large deviations principle with good rate function. The rate function is the same as for the optimal centralized detector and is given by  $I_{(l)}(\cdot)$  in eqn. (18).

Before proving Theorem 9, define  $\Phi(k, j)$ , for  $k > j \geq 1$ , as follows:

$$\Phi(k, j) := W(k-1)W(k-2)\dots W(j), \quad (24)$$

and remark that the algorithm in eqn. (21) can be written as:

$$x(k) = \frac{N}{k} \sum_{j=1}^{k-1} \Phi(k, j)\eta(j) + \frac{N}{k}\eta(k), \quad k = 2, 3, \dots \quad (25)$$

Next, recall that  $J = (1/N)11^T$ , introduce notation:

$$\tilde{\Phi}(k, j) := \tilde{W}(k-1)\tilde{W}(k-2)\dots\tilde{W}(j), \quad k > j \geq 1, \quad (26)$$

and remark that

$$\tilde{\Phi}(k, j) = \Phi(k, j) - J.$$

To prove Theorem 9, we borrow the following result (Lemma 10) on the matrices  $\tilde{\Phi}(k, j)$  from reference [14] (Lemma 3.2). First, denote by  $[\tilde{\Phi}(k, j)]_{il}$  the entry in  $i$ -th row and  $l$ -th column of matrix  $\tilde{\Phi}(k, j)$ .

*Lemma 10* Let Assumption 6 hold. Then, for the matrices  $\tilde{\Phi}(k, j)$ , defined by eqn. (26), there holds:

$$\max_{i,l=1,\dots,N} |[\tilde{\Phi}(k, j)]_{il}| \leq \theta \beta^{k-j}, \quad (27)$$

where  $\theta = (1 - \frac{W_{\min}}{4N^2})^{-2}$ , and  $\beta = (1 - \frac{W_{\min}}{4N^2})^{1/B} < 1$ .

Lemma 10 says that, under Assumption 6, the size of the matrix  $\tilde{\Phi}(k, j)$  decays geometrically (in  $k - j$ ) to zero. This fact will be important in showing Theorem 9.

*Proof of Theorem 9:* Define, for  $\mu \in \mathbb{R}$ , the quantity:

$$\begin{aligned} \Lambda_k^{(l)}(\mu) &:= \log \mathbb{E}_l [\exp(\mu x_i(k))] \quad (28) \\ &= \log \mathbb{E}_l [\exp(\lambda^\top x(k))], \quad (29) \end{aligned}$$

where  $\lambda = \mu e_i$ ,  $\lambda \in \mathbb{R}^N$ , and  $\mathbb{E}_l[a] := \mathbb{E}[a|H_l]$ ,  $l = 0, 1$ . Here  $e_i$  denotes the  $i$ -th column of  $N \times N$  identity matrix. We drop the dependence on  $i$  in the definition of  $\Lambda_k^{(l)}(\mu)$  for notation simplicity. Recall the expressions for  $m_L^{(l)}$  and  $\sigma_L^2$  in eqns. (14) and (15). We will show, for all  $\mu \in \mathbb{R}$ , the following equality:

$$\lim_{k \rightarrow \infty} \frac{1}{k} \Lambda_k^{(l)}(k\mu) = \frac{1}{2} \sigma_L^2 \mu^2 + m_L^{(l)} \mu, \quad (30)$$

Consider the function  $\mu \mapsto \frac{1}{2} \sigma_L^2 \mu^2 + m_L^{(l)} \mu$ ; this function is essentially smooth, continuous, and its domain is  $\mathbb{R}$ ; hence, by the Gärtner-Ellis theorem ([13], Theorem 2.3.6),  $\{\chi_{i,k}^{(l)}\}$  (the sequence of measures of  $x_i(k)$  under  $H_l$ ) satisfies the LDP. The corresponding rate function equals the F-L transform of the function  $\mu \mapsto \frac{1}{2} \sigma_L^2 \mu^2 + m_L^{(l)} \mu$ ; and it is easy to show that the F-L transform of  $\mu \mapsto \frac{1}{2} \sigma_L^2 \mu^2 + m_L^{(l)} \mu$  equals the rate function  $I^{(l)}(\cdot)$  given by eqn. (18). Thus, proving Theorem 9 reduces to showing (30). We thus proceed with showing (30). Namely, we have:

$$\begin{aligned} &\frac{1}{k} \Lambda_k^{(l)}(k\mu) = \\ &\frac{1}{k} \log \mathbb{E}_l \left[ \exp \left( N \lambda^\top \sum_{j=1}^{k-1} \Phi(k, j) \eta(j) + N \lambda^\top \eta(k) \right) \right] \\ &= \frac{1}{k} \log \mathbb{E}_l \left[ \exp \left( N \lambda^\top \sum_{j=1}^{k-1} \Phi(k, j) \eta(j) \right) \right] \\ &+ \frac{1}{k} \log \mathbb{E}_l [\exp(N \lambda^\top \eta(k))], \end{aligned}$$

where the last equality holds because  $\eta(k)$  is independent from  $\eta(j)$ ,  $j = 1, \dots, k-1$ . We will be interested in computing the limit  $\lim_{k \rightarrow \infty} \frac{1}{k} \Lambda_k^{(l)}(k\mu)$ , for all  $\mu \in \mathbb{R}$ ; with this respect, remark that

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log \mathbb{E}_l [\exp(N \lambda^\top \eta(k))] = 0,$$

for all  $\lambda \in \mathbb{R}^N$ , because  $\eta(k)$  is a Gaussian random vector and hence it has finite log-moment generating function at any point  $N\lambda$ .

Thus, we have that  $\lim_{k \rightarrow \infty} \frac{1}{k} \Lambda_k^{(l)}(k\mu) =$

$\lim_{k \rightarrow \infty} \frac{1}{k} \mathcal{L}_k^{(l)}(k\mu)$ , where

$$\mathcal{L}_k^{(l)}(k\mu) = \log \mathbb{E}_l \left[ \exp \left( N \lambda^\top \sum_{j=1}^{k-1} \Phi(k, j) \eta(j) \right) \right],$$

and we proceed with the computation of  $\mathcal{L}_k^{(l)}(k\mu)$ . The random variables  $N \lambda^\top \Phi(k, j) \eta(j)$ ,  $j = 1, \dots, k-1$ , are independent; moreover, they are Gaussian random variables, as linear transformation of the Gaussian variables  $\eta(j)$ . Recall that  $m_\eta^{(l)}$  and  $S^\eta$  denote the mean and the covariance of  $\eta(k)$  under hypothesis  $H_l$ . Using the independence of  $\eta(j)$  and  $\eta(s)$ ,  $s \neq j$ , and using the expression for the moment generating function of  $\eta(j)$ , we obtain successively:

$$\begin{aligned} \frac{1}{k} \mathcal{L}_k^{(l)}(k\mu) &= \frac{1}{k} \log \mathbb{E}_l \left[ \exp \left( N \lambda^\top \sum_{j=1}^{k-1} \Phi(k, j) \eta(j) \right) \right] \\ &= \frac{1}{k} \log \mathbb{E}_l [\Pi_{j=1}^{k-1} \exp(N \lambda^\top \Phi(k, j) \eta(j))] \quad (31) \\ &= \frac{1}{k} \log \Pi_{j=1}^{k-1} \exp \left( N \lambda^\top \Phi(k, j) m_\eta^{(l)} \right) \\ &\quad \exp \left( \frac{1}{2} N^2 \lambda^\top \Phi(k, j) S^\eta \Phi(k, j)^\top \lambda \right) \\ &= \frac{1}{k} \log \Pi_{j=1}^{k-1} \exp \left( N \lambda^\top \left( \tilde{\Phi}(k, j) + J \right) m_\eta^{(l)} \right) \\ &\quad \exp \left( \frac{1}{2} N^2 \lambda^\top \left( J + \tilde{\Phi}(k, j) \right) S^\eta \left( J + \tilde{\Phi}(k, j)^\top \right) \lambda \right). \end{aligned}$$

Denote further:

$$\begin{aligned} \delta(k) &= \frac{N}{k} \lambda^\top \sum_{j=1}^{k-1} \tilde{\Phi}(k, j) m_\eta^{(l)} \quad (32) \\ &+ \frac{N^2}{2k} \lambda^\top \sum_{j=1}^{k-1} \tilde{\Phi}(k, j) S^\eta \tilde{\Phi}(k, j)^\top \lambda \\ &+ \frac{N^2}{2k} \lambda^\top J S^\eta \sum_{j=1}^{k-1} \tilde{\Phi}(k, j)^\top \lambda \\ &+ \frac{N^2}{2k} \lambda^\top \sum_{j=1}^{k-1} \tilde{\Phi}(k, j) S^\eta J \lambda \\ \bar{\Lambda}_{(l)}(\mu, k) &= \frac{(k-1) N \lambda^\top J m_\eta^{(l)}}{k} \\ &+ \frac{N^2 (k-1) \lambda^\top J S^\eta J \lambda}{2k}, \end{aligned}$$

where dependence on  $H_l$  is dropped in the definition of  $\delta(k)$ . Then, it is easy to see that  $\frac{1}{k} \mathcal{L}_k^{(l)}(k\mu) = \bar{\Lambda}_{(l)}(\mu, k) + \delta(k)$ . Also, we have:  $\lim_{k \rightarrow \infty} \bar{\Lambda}_{(l)}(\mu, k) = N \lambda^\top J m_\eta^{(l)} + \frac{N^2}{2} \lambda^\top J S^\eta J \lambda =: \bar{\Lambda}_{(l)}(\mu)$ .

Recall the expressions for  $v$ ,  $m_L^{(l)}$ ,  $\sigma_L^2$ ,  $m_\eta^{(l)}$ , and  $S^\eta$

in eqns. (16), (14), (15), (22), (23). We proceed with the computation of  $\bar{\Lambda}_{(l)}(\mu)$ :

$$\begin{aligned}\bar{\Lambda}_{(l)}(\mu) &= (-1)^{l+1} \frac{N}{2} \mu (m_1 - m_0)^\top \text{Diag}(v) J e_i \\ &+ \frac{N^2}{2} \mu^2 e_i^\top J S^\eta J e_i \\ &= (-1)^{l+1} \frac{1}{2} \mu \mathbf{1}^\top \text{Diag}(v) (m_1 - m_0) \\ &+ \frac{1}{2} \mu^2 \mathbf{1}^\top \text{Diag}(v) S \text{Diag}(v) \mathbf{1} \\ &= (-1)^{l+1} \frac{1}{2} \mu (m_1 - m_0)^\top S^{-1} (m_1 - m_0) \\ &+ \frac{1}{2} \mu^2 (m_1 - m_0)^\top S^{-1} (m_1 - m_0) \\ &= m_L^{(l)} \mu + \frac{1}{2} \sigma_L^2 \mu^2.\end{aligned}$$

We proceed by showing that  $\delta(k) \rightarrow 0$  as  $k \rightarrow \infty$ , which implies the equality in eqn (30). Define the quantities  $\bar{S}$ ,  $\bar{m}$ , and  $\bar{b}$ , by:

$$\bar{S} := \max_{i,l=1,\dots,N} |[S^\eta]_{il}| \quad (33)$$

$$\bar{m} := \max_{i=1,\dots,N} |[m_\eta^{(l)}]_i| \quad (34)$$

$$\bar{b} := \max_{l=1,\dots,N} |[S^\eta J e_i]_l|. \quad (35)$$

Then, it can be shown that  $|\delta(k)|$  is bounded as follows:

$$\begin{aligned}|\delta(k)| &\leq \frac{N^2}{k} |\mu| \bar{m} \sum_{j=1}^{k-1} \max_{i,l=1,\dots,N} |[\tilde{\Phi}(k,j)]_{il}| \quad (36) \\ &+ \frac{N^4}{2k} \mu^2 \bar{S} \sum_{j=1}^{k-1} \left( \max_{i,l=1,\dots,N} |[\tilde{\Phi}(k,j)]_{il}| \right)^2 \\ &+ \frac{N^3}{k} \mu^2 \bar{b} \sum_{j=1}^{k-1} \max_{i,l=1,\dots,N} |[\tilde{\Phi}(k,j)]_{il}|.\end{aligned}$$

Applying Lemma 10 to (36), and using the fact that  $\beta^{k-j} < \beta^{k-j-1}$ ,  $k > j$ , we obtain successively:

$$|\delta(k)| \leq \frac{\theta}{k} (N^2 \bar{m} |\mu| + N^3 \mu^2 \bar{b}) \sum_{j=1}^{k-1} \beta^{k-j-1} \quad (37)$$

$$+ \frac{\theta^2}{k} \frac{N^4}{2} \mu^2 \bar{S} \sum_{j=1}^{k-1} \beta^{2(k-j-1)} \quad (38)$$

$$\leq \frac{\theta}{k} (N^2 \bar{m} |\mu| + N^3 \mu^2 \bar{b}) \frac{1}{1-\beta} \quad (39)$$

$$+ \frac{\theta^2}{k} \frac{N^4}{2} \mu^2 \bar{S} \frac{1}{1-\beta^2}. \quad (40)$$

Letting  $k \rightarrow +\infty$ , we get that  $|\delta(k)| \rightarrow 0$ , and hence,  $\delta(k) \rightarrow 0$ , which establishes eqn. (30). ■

We are now ready to state the main result on asymptotic optimality of the distributed detector (in the sense of Definition 5.)

*Corollary 11 (Chernoff lemma for the distributed detector: Asymptotic optimality)* The local decision test  $T_{k,i} := \mathcal{I}_{\{x_i(k) > 0\}}$ ,  $k = 1, 2, \dots$ , at each node  $i$ , is asymptotically optimal in the sense of Definition 5. The corresponding exponential decay rate of the Bayes probability of error, at each node  $i$ , is given by:

$$\begin{aligned}\lim_{k \rightarrow \infty} \frac{1}{k} \log P_{i,\text{dis}}^e(k) &= -I_{(0)}(0) \quad (41) \\ &= -\frac{1}{8} (m_1 - m_0) S^{-1} (m_1 - m_0).\end{aligned}$$

*Proof:* Denote by  $\alpha_{i,\text{dis}}(k)$  and  $\beta_{i,\text{dis}}(k)$ , respectively, the probability of false alarm and the probability of miss for the distributed detector at sensor  $i$ , i.e.,

$$\alpha_{i,\text{dis}}(k) = P(x_i(k) > 0 | H_0) = \chi_{i,k}^{(0)}((0, +\infty))$$

$$\beta_{i,\text{dis}}(k) = P(x_i(k) \leq 0 | H_1) = \chi_{i,k}^{(1)}((-\infty, 0]).$$

Consider now only  $\alpha_{i,\text{dis}}(k)$  but the same applies to  $\beta_{i,\text{dis}}(k)$ . By Theorem 9, the sequence of measures  $\chi_{i,k}^{(0)}$  satisfies the LDP with good rate function  $I_{(0)}(\cdot)$  given in eqn. (18). Thus, we have the following bounds:

$$\limsup_{k \rightarrow \infty} \frac{1}{k} \log \chi_{i,k}^{(0)}([0, +\infty)) \leq - \inf_{t \in [0, +\infty)} I_{(0)}(t) \quad (42)$$

$$\liminf_{k \rightarrow \infty} \frac{1}{k} \log \chi_{i,k}^{(0)}((0, +\infty)) \geq - \inf_{t \in (0, +\infty)} I_{(0)}(t) \quad (43)$$

Due to the continuity of the function  $I_{(0)}(\cdot)$  (see eqn. (18)), the infima on the righthand sides in eqns. (42) and (43) are equal; it is easy to see that they are equal to  $-I_{(0)}(0)$ . Thus, we have:

$$\begin{aligned}-I_{(0)}(0) &\leq \liminf_{k \rightarrow \infty} \frac{1}{k} \log \chi_{i,k}^{(0)}((0, +\infty)) \\ &\leq \limsup_{k \rightarrow \infty} \frac{1}{k} \log \chi_{i,k}^{(0)}((0, +\infty)) \\ &\leq \limsup_{k \rightarrow \infty} \frac{1}{k} \log \chi_{i,k}^{(0)}([0, +\infty)) \\ &\leq -I_{(0)}(0).\end{aligned}$$

From the last set of inequalities we conclude that:

$$\lim_{k \rightarrow +\infty} \frac{1}{k} \log \alpha_{i,\text{dis}}(k) = -I_{(0)}(0). \quad (44)$$

Similarly, it can be shown that:

$$\lim_{k \rightarrow +\infty} \frac{1}{k} \log \beta_{i,\text{dis}}(k) = -I_{(1)}(0) \quad (45)$$

$$= -I_{(0)}(0). \quad (46)$$

Now, consider

$$P_{i,\text{dis}}^e(k) = \alpha_{i,\text{dis}}(k) P(H_0) + \beta_{i,\text{dis}}(k) P(H_1), \quad (47)$$

for which the following inequalities hold:

$$P_{i,\text{dis}}^e(k) \leq \alpha_{i,\text{dis}}(k) + \beta_{i,\text{dis}}(k) \quad (48)$$

$$P_{i,\text{dis}}^e(k) \geq \alpha_{i,\text{dis}}(k) P(H_0).$$

By eqns. (48), we obtain:

$$\begin{aligned} & \limsup_{k \rightarrow \infty} \frac{1}{k} \log P_{i, \text{dis}}^e(k) = \\ & \max \left\{ \limsup_{k \rightarrow +\infty} \frac{1}{k} \log \alpha_{i, \text{dis}}(k), \limsup_{k \rightarrow +\infty} \frac{1}{k} \log \beta_{i, \text{dis}}(k) \right\} \\ & = -I_{(0)}(0) \\ & \liminf_{k \rightarrow \infty} \frac{1}{k} \log P_{i, \text{dis}}^e(k) \geq \liminf_{k \rightarrow \infty} \frac{1}{k} \log \alpha_{i, \text{dis}}(k) \\ & = -I_{(0)}(0), \end{aligned}$$

and the claim of Corollary follows. ■

**Remarks on Corollary 11.** Corollary 11 says that, for large  $k$  (i.e., in the asymptotic regime,) the Bayes probability of error at each node  $i$  behaves as:  $P_{i, \text{dis}}^e(k) \sim e^{-kI_{(0)}(0)}$ . That is,  $P_{i, \text{dis}}^e(k)$ , for large  $k$ , decays exponentially at the best possible rate, equal to the rate  $I_{(0)}(0)$  of the (asymptotically) optimal centralized detector. This rate does not depend on the network connectivity, provided that the graph that collects all the links that are online (at least once) within finite time window (of length  $B$ ) is connected (see Assumption 6.) Intuitively, an arbitrary time varying network, whose nodes communicate sufficiently often (within finite length time windows,) provides sufficient information flow to achieve asymptotic optimality.

We now comment on the non asymptotic finite time regime. To this end, remark that  $P_{i, \text{dis}}^e(k)$  can be expressed as:  $P_{i, \text{dis}}^e(k) = F_i(k)e^{-kI_{(0)}(0)}$ , where  $\lim_{k \rightarrow \infty} \frac{1}{k} \log F_i(k) = 0$  (and thus,  $F_i(k)$  has no effect when  $k$  grows large.) The sequence  $\{F_i(k)\}$  plays a role in a finite time regime; it clearly depends on the network connectivity and can be, in general, different for different sensors. Analysis (by simulation) of the finite time regime is, due to the lack of space, omitted, and is pursued elsewhere. We briefly comment here that our numerical experience suggests that, in the finite time regime, the sequence  $F_i(k)$  does not have a very large effect. The best distributed sensor–detector, among all  $N$  sensors, is typically close in performance to the optimal centralized detector, in the finite time regime also.

## VI. SUMMARY

We applied large deviations theory to analyze the performance of the running consensus distributed detection algorithm. We considered spatially correlated Gaussian noise and time varying networks. With running consensus, the state at each node is updated at each time step by: 1) exchanging information with the immediate neighbors in the network; and 2) incorporating into the decision process new local observations. We allowed the underlying network be time varying, provided that the graph that collects all the links that are at least once online within a finite time window is connected. We

showed that, under spatially correlated Gaussian noise and stated network connectivity assumptions, the running consensus asymptotically approaches the optimal centralized detector. That is, the Bayes probability of detection error at each sensor decays exponentially at the best achievable rate, the Chernoff information rate.

## REFERENCES

- [1] S. Kar and J. M. F. Moura, “Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication,” August 2008, submitted for publication, 51 pages. [Online]. Available: arXiv:0809.0009v1 [cs.MA]
- [2] 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.
- [3] R. Viswanathan and P. R. Varshney, “Decentralized detection with multiple sensors: Part I—fundamentals,” *Proc. IEEE*, vol. 85, pp. 54–63, January 1997.
- [4] J. F. Chamberland and V. Veeravalli, “Decentralized detection in sensor networks,” *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 407–416, February 2003.
- [5] —, “Asymptotic results for decentralized detection in power constrained wireless sensor networks,” *IEEE Journal on selected areas in communications*, vol. 22, no. 6, pp. 1007–1015, August 2004.
- [6] J. N. Tsitsiklis, “Decentralized detection,” *Adv. Statist. Signal Processing*, vol. 2, pp. 297–344, 1993.
- [7] R. S. Blum, S. A. Kassam, and H. V. Poor, “Decentralized detection with multiple sensors: Part II—advanced topics,” *Proc. IEEE*, vol. 85, pp. 64–79, January 1997.
- [8] S. A. Aldosari and J. M. F. Moura, “Detection in sensor networks: the saddlepoint approximation,” *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 327–340, January 2007.
- [9] S. Kar, S. A. Aldosari, and J. M. F. Moura, “Topology for distributed inference on graphs,” *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2609–2613, June 2008.
- [10] S. Kar and J. M. F. Moura, “Consensus based detection in sensor networks: topology design under practical constraints,” in *Proc. Workshop Inf. Theory in Sensor Networks*, Santa Fe, NM, June 2007.
- [11] V. Saligrama, M. Alanyali, and O. Savas, “Distributed detection in sensor networks with packet losses and finite capacity links,” *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4118–4132, November 2006.
- [12] P. Braca, S. Marano, V. Matta, and P. Willet, “Asymptotic optimality of running consensus in testing binary hypothesis,” *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 814–825, February 2010.
- [13] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*. Boston, MA: Jones and Barlett, 1993.
- [14] S. S. Ram, A. Nedic, and V. V. Veeravalli, “Distributed stochastic subgradient projection algorithms for convex optimization,” *Journal of Optimization Theory and Applications*, to appear, 2010.