

# Distributed Detection Over Noisy Networks: Large Deviations Analysis

Dušan Jakovetić, *Student Member, IEEE*, José M. F. Moura, *Fellow, IEEE*, and João Xavier, *Member, IEEE*

**Abstract**—We study the large deviations performance of consensus+innovations distributed detection over noisy networks, where agents at a time step  $k$  cooperate with their immediate neighbors (consensus) and assimilate their new observations (innovation.) We show that, under noisy communication, all agents can still achieve an exponential error rate, even when certain (or most) agents cannot detect the event of interest in isolation. The key to achieving this is the appropriate design of the time-varying weight sequence  $\{\alpha_k = b_0/(a+k)\}$  by which agents weigh their neighbors' messages. We find a communication payoff threshold on the communication noise power, i.e., the critical noise power below which cooperation among neighbors improves detection performance and above which the noise in the communication among agents overwhelms the distributed detector performance. Numerical examples illustrate several tradeoffs among network parameters and between the time (or number of measurements) needed for a reliable decision and the transmission power invested by the agents.

**Index Terms**—Chernoff information, distributed detection, large deviations, noisy communication.

## I. INTRODUCTION

WE describe briefly the problem that the paper considers, detailed in Section II. We study *distributed* simple binary hypothesis testing:  $N$  agents cooperate, through a sparse, connected communications graph  $G = (V, E)$  ( $V$  is the set of agents, the nodes of the graph  $G$ , and  $E$  is the set of interagents channels, the edges of  $G$ ) to decide at each time  $k$ ,  $k = 1, 2, \dots$ , between two possible states of nature,  $H_0$  and  $H_1$ . This problem arises in many applications, including classical surveillance, but

Manuscript received August 02, 2011; revised January 25, 2012; accepted April 09, 2012. Date of publication May 02, 2012; date of current version July 10, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Roberto Lopez-Valcarlos. The work of D. Jakovetić and J. Xavier 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 CMU-PT/SIA/0026/2009; and by ISR/IST plurianual funding (POSC program, FEDER). The work of D. Jakovetić and J. M. F. Moura is partially supported by the NSF under Grants CCF-1011903 and CCF-1018509, and by AFOSR Grant FA95501010291. D. Jakovetić holds a fellowship from FCT.

D. Jakovetić is with the Institute for Systems and Robotics (ISR), Instituto Superior Técnico (IST), Technical University of Lisbon, 1600-011 Lisbon, Portugal, and with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213-3890 USA (e-mail: djakovet@andrew.cmu.edu).

J. M. F. Moura is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213-3890 USA (e-mail: moura@ece.cmu.edu).

J. Xavier is with the Institute for Systems and Robotics (ISR), Instituto Superior Técnico (IST), Technical University of Lisbon, 1600-011 Lisbon, Portugal (e-mail: jxavier@isr.ist.utl.pt).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2012.2197395

now in a distributed setting like in netted, multisite, or MIMO radars, e.g., [1], where a system of spatially separated networked multistatic radar stations cooperate at every time  $k = 1, 2, \dots$ , to detect the presence or absence of a target, or cognitive radio networks where distributed agents detect a primary user, e.g., [2]. We adopt the following *distributed* sequential detector. At each time  $k$ , agent  $i$  executes three tasks: 1) makes an observation  $y_i(k)$ ; 2) updates its local detection statistic  $x_i(k)$  by a *distributed* algorithm:

$$x_i(k+1) = \underbrace{W_{ii}^1(k)x_i(k) + \sum_{j \in O_i} W_{ij}^1(k)x_j(k)}_{\text{consensus}} + \underbrace{W_i^2(k)\eta_i(k+1)}_{\text{innovations}} \quad (1)$$

where  $W_{ij}^1(k)$  and  $W_i^2(k)$ ,  $1 \leq i, j \leq N$ , are weights;  $\eta_i(k)$  is the local instantaneous log-likelihood ratio of agent  $i$  at time  $k$  computed from its own  $y_i(k)$ ; and  $O_i$  is the set of neighbors of agent  $i$  as determined by the edge set  $E$  of the graph  $G$ ; and 3) makes a decision by thresholding its detection statistic:

$$x_i(k) \underset{H_0}{\overset{H_1}{\geq}} \gamma, \quad k = 1, 2, \dots, i = 1, \dots, N. \quad (2)$$

Equation (1) updates the test statistic with a two-step structure: the first, given by the first two terms on the right-hand side (RHS) of (1), is like *consensus* and reflects the cooperation among agents—it averages the local statistic  $x_i(k)$  of  $i$  with the local statistics  $x_j(k)$  received from the neighbors  $j \in O_i$ ; and the second, which we refer to as an *innovations* step, assimilates the measurement  $y_i(k)$  through the instantaneous local log-likelihood  $\eta_i(k)$ . Hence, we refer to the local updating (1) at each agent  $i$  as a *consensus+innovations* distributed algorithm and to the set of  $N$  detectors (1) and (2),  $i = 1, \dots, N$ , as the consensus+innovations *distributed* detector, or distributed detector for short. We are fundamentally concerned with how ‘good’ can we make the distributed detector, i.e., what performance guarantees can we provide, when we carefully design the weight sequences  $W_{ij}^1(k)$  and  $W_i^2(k)$  in (1). To be more specific, we benchmark the error detection performance of the distributed detector with respect to the error performance of the Neyman–Pearson centralized sequential detector, which, under appropriate assumptions, is

$$x(k) \underset{H_0}{\overset{H_1}{\geq}} \gamma, \quad k = 1, 2, \dots \quad (3)$$

where the centralized log-likelihood ratio  $x(k)$  is given by

$$x(k) = \frac{1}{k} \sum_{i=1}^N \sum_{j=1}^k \eta_i(j) \quad (4)$$

and  $\eta_i(k)$  is a renormalization of the local instantaneous log-likelihood ratio computed by agent  $i$  at time  $k$  from its instantaneous observation  $y_i(k)$ . Our goal is to determine the conditions under which and then show that, by carefully designing the weights  $W_{ij}^1(k)$  and  $W_i^2(k)$  in (1), we can similarly guarantee exponential rate decay at *every* agent  $i$  by a distributed detector (1) and (2), i.e., the error probability of the distributed detector at each and every agent  $i$ , decays *asymptotically* exponentially fast.

We consider these design and performance guarantee questions under a fairly general setting that takes into consideration limitations that may not necessarily arise in a centralized setting but are natural in many distributed applications. Because of limited power, not only are 1) the observations  $y_i(k)$  of agent  $i$  noisy, affected by *sensing* noise, but also 2) the communications among agent  $i$  and its neighboring agents (when they cooperate) are noisy, impacted by *communications* noise. This paper considers Gaussian sensing and communication noises and establishes exponential error rate of decay for distributed detection. Note that, with noisy communications, the updating of the local statistic at agent  $i$  does not follow (1) but is more complex as we will see in Section II. We extend elsewhere these results to (non-Gaussian) quantized inter-agent communication, and to Gaussian temporally correlated sensing and communication noises. In [3], we consider non-Gaussian sensing noises.

*Brief Comment on the Literature:* There is a vast literature on decentralized and distributed inference. While we consider a distributed architecture, i.e., with no fusion center, [4]–[8] consider decentralized parallel fusion architectures, where all agents communicate with a fusion center. References [9]–[13] have a distributed architecture (no fusion center) but are of the consensus type—each sensor makes a single observation and then the sensors fuse their local decisions by the consensus algorithm, or by belief propagation like in [9]. Reference [14] and the algorithm in [15, Sec. IV] are essentially of the consensus type, since they run consensus till convergence between each round of measurements. The algorithm in [15, Sec. V] assumes a complete architecture, or, if not, it uses a multihop protocol, so that each sensor has access to the observations of *all* the sensors at *each* and *every* time step. These references stand in contrast with the class of algorithms we consider: We use a *consensus+innovations* algorithm, i.e., a distributed algorithm (no fusion center) that interleaves *consensus* with *innovations* (processing of the observations) at the *same* time step, rather than running consensus to convergence in *between* successive observations.

We now contrast our work with [16]–[18], [2], and [19]–[26] that, like ours, are distributed, include communication among neighbors, and process the new observations at every time step as they are measured. We first comment that the main features that distinguish our paper from these works are 1) we consider *single scale* distributed detectors; 2) the communications among agents is corrupted by additive noise; and 3) we are primarily concerned with showing exponential error rate (with appropriate choice of the weights  $W_{ij}^1(k)$ ,  $W_i^2(k)$  in (1)). References [2], [17], and [18] look at distributed LMS and RLS adaptive algorithms. They assume noiseless communications among agents (no additive noise), and they do not study the decay rate of the

error probability.<sup>1</sup> Reference [20] addresses the problem of distributed change detection (a tracking type of problem) allowing for random averaging matrices and spatio-temporally correlated data, but this work does not consider noise in the communication among agents, nor is it concerned with establishing the exponential error rate of the algorithm therein. References [21]–[23] consider link failures but no additive noise in the intra-agents' communication. Also, [21] considers the limiting behavior of their distributed detector when the difference between the means under the two hypotheses goes to zero, a very different problem from ours. Reference [24] considers deterministically time varying networks and no communications noise. Reference [25] is concerned with estimation and considers a very general model that includes agent failures, link failures, and various degrees of either quantized or noisy communications. Because this reference studies estimation and not detection, it is not concerned with exponential decay rates of the error probability as we are here; rather, it shows consistency, asymptotic efficiency, and normality of the estimates through stochastic approximation and Lyapounov function arguments and through bounding pathwise behavior, rather than through large deviations arguments as we apply here to our detection analysis. A nonlinear estimator in [25] is *mixed scale*, while the class of detectors we study in this paper is *single scale*. The corresponding mixed scale algorithms for detection are presented and studied in [26], which, to the best of our knowledge, and within the consensus+innovations *detection* literature, is, like us, the only reference to consider additive noise in the communications among agents (also, with no link failures.) Our results contrast with [26], for the distributed sequential detector that we design, we establish that the error probability *at each agent* decays exponentially fast; we demonstrate this under broad conditions, including unequal local agents' sensing signal-to-noise ratios and when certain or most agents are locally not detectable; in contrast, in these settings, the algorithm in [26] does not achieve exponential decay rate. We can show that the multiscale algorithm in [26] does not achieve exponential error rate: the maximal probability of error across agents is  $\Omega(e^{-c k^\tau})$ ,  $\tau \in (0.5, 1)$  (see paragraph with *Notation* heading for the meaning of the symbol  $\Omega$ ). Due to lack of space, we omit the proof of this statement here.

Finally, we relate our paper with our prior work [22], [23]. These papers focus on how link failures impinge on detection performance, while here we show that *additive communication noise* in the links impacts in a qualitatively different way the error performance; with link failures, more communication among agents can only improve the error performance, since when communication does happen agents receive their neighbors detection statistics unencumbered by noise. With additive communication noise, however, a clear tradeoff arises between communication noise and amount of information flow (or how often agents communicate); this leads to a phase change behavior: Only when the communication noise power is below a threshold does increased or more frequent cooperation improve performance—in that the distributed error performance

<sup>1</sup>Coupling [2] with the results in [16] and [19], which considers diffusion estimators with additive communication noise, the probability of error of the LMS detector in [2] does not go to zero as the number of observations grows to infinity, let alone achieve exponential decay rate, in contrast with the performance of our distributed detector.

of the worst (noisiest) agent is better than the isolated (no cooperation) performance of the best agent. In [22] and [23], we model certain averaging matrices as independent identically distributed (i.i.d.) so that their distribution is time invariant; here, because of time-decaying weights, the corresponding weight matrices are time varying, forcing us to develop new analysis to show asymptotic stability of certain time varying systems. The methods presented here are of independent interest and may find application in many other contexts.

*Paper organization:* Section II describes the problem model and presents our distributed detector. Section III establishes the asymptotic performance of our distributed detector. Section IV interprets the results and presents simulations. Section V concludes the paper. Appendices A and B prove all our results.

*Notation:* We adopt the following: lower and upper bold-face letters represent vectors and matrices;  $A_{ij}$  or  $[A]_{ij}$  are the  $(i, j)$ th entry of a matrix  $A$ ;  $a_i$  or  $[a]_i$  are the  $i$ th entry of a vector  $a$ ;  $A^\top$  and  $A^{-1}$  are the transpose and inverse of  $A$ ;  $I$ ,  $\mathbf{1}$ , and  $\mathbf{e}_i$  are the identity matrix, the column vector with unit entries, and the  $i$ th column of  $I$ ;  $\mathbf{J} := (1/N)\mathbf{1}\mathbf{1}^\top$  is the  $N \times N$  ideal averaging matrix;  $\|\cdot\|_l$  is the vector (respectively, matrix)  $l$ -norm;  $\|\cdot\| = \|\cdot\|_2$  is the Euclidean (respectively, spectral) norm;  $\lambda_i(\cdot)$  and  $\text{tr}(\cdot)$  are the  $i$ th smallest eigenvalue, and the trace of a matrix;  $\otimes$  the Kronecker product of matrices;  $\text{Diag}(\mathbf{a})$  is the diagonal matrix with the diagonal equal to the vector  $\mathbf{a}$ ;  $\mathbf{a} = \text{Vec}(\mathbf{A})$  is the vector that stacks columns of  $\mathbf{A}$ , and the “inverse” operation is  $\mathbf{A} = \text{Vec}^{-1}(\mathbf{a})$ ;  $|\mathcal{A}|$  is the cardinality of  $\mathcal{A}$ ;  $\mathbb{E}[\cdot]$ ,  $\text{Var}(\cdot)$ ,  $\text{Cov}(\cdot)$ , and  $\mathbb{P}(\cdot)$  are the expected value, the variance, the covariance, and probability operators, (for zero mean random vectors  $\mathbf{a}$  and  $\mathbf{b}$ , the autocovariance is  $\text{Cov}(\mathbf{a}) = \mathbb{E}[\mathbf{a}\mathbf{a}^\top]$  and the cross-covariance is  $\text{Cov}(\mathbf{a}, \mathbf{b}) = \mathbb{E}[\mathbf{a}\mathbf{b}^\top]$ );  $\mathcal{I}_{\mathcal{A}}$  is the indicator function of the event  $\mathcal{A}$ ;  $\delta_{k_1 k_2}$  is the Kronecker symbol equal to 1 if  $k_1 = k_2$  and zero otherwise;  $\mathcal{Q}(\cdot)$  is the  $Q$ -function, i.e., the function that calculates the right tail probability of the standard normal distribution and given, with two bounds, by [27]

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \frac{t}{1+t^2} e^{-t^2/2} &\leq \mathcal{Q}(t) \\ &= \frac{1}{\sqrt{2\pi}} \int_t^{+\infty} e^{-\frac{u^2}{2}} du \leq \frac{1}{\sqrt{2\pi}} \frac{1}{t} e^{-t^2/2}, \quad t > 0. \end{aligned} \quad (5)$$

We also make use of the standard  $\Omega$  and  $O$  notations:  $f(k) = \Omega(g(k))$  stands for existence of a  $K > 0$  such that  $f(k) \geq cg(k)$ , for some  $c > 0$ , for all  $k \geq K$ ; and  $f(k) = O(g(k))$  means existence of  $K > 0$  such that  $f(k) \leq cg(k)$ , for some  $c > 0$ , for all  $k \geq K$ .

## II. PROBLEM FORMULATION

We formulate the distributed detection simple binary hypothesis testing problem that we study. We consider  $N$  agents that cooperate through a communication graph  $G = (V, E)$  to decide among two hypothesis  $H_0$  or  $H_1$ . Each agent  $i$  keeps its own local detection statistic  $x_i(k)$ . At time  $k$ , it performs four tasks: 1) makes a measurement  $y_i(k)$ ; 2) communicates to the neighbors its own detection statistic  $x_i(k)$ ; 3) updates  $x_i(k)$ ; and 4) performs the decision (with zero threshold in (2))

$$x_i(k) \underset{H_0}{\overset{H_1}{\geq}} 0. \quad (6)$$

We take in (6) the threshold to be zero because the best Bayes exponential error rate is achieved with a Neyman–Pearson detector with zero threshold [28, p. 93]. Henceforth, we consider the three detectors, isolated, centralized, and distributed with zero threshold.

### A. Sensing and Communication Model

*Communication Graph*  $G = (V, E)$ : The set  $V$  assembles the agents as nodes of the graph and the set  $E$  collects their communication channels as edges of the graph: if agents  $i$  and  $j$  communicate, then  $(i, j) \in E$ . The neighborhood of agent  $i$  is the set  $O_i$  of agents connected to  $i$  by an edge in the graph and the degree of  $i$  is  $d_i = |O_i|$ . The graph is sparse and can be represented by its adjacency matrix  $\mathbf{A}$ , where entry  $A_{ij}$  is zero unless  $(i, j) \in E$ , in which case  $A_{ij} = 1$ . ( $A_{ii} = 0$ , for all  $i$ .) We associate with the graph its Laplacian  $\mathbf{L} = \mathbf{D} - \mathbf{A}$ , where  $\mathbf{D}$  is the degree matrix, a diagonal matrix whose diagonal entries  $D_{ii} = d_i = \sum_{j=1}^N A_{ij}$ . The Laplacian is positive semi-definite, its eigenvalues  $\lambda_i(\mathbf{L})$ ,  $i = 1 \dots N$ , are ordered from smallest to largest, with  $\lambda_1(\mathbf{L}) = 0$ . We make the following assumption.

*Assumption 1 (Communication Graph):* The graph  $G$  is undirected, simple, and connected, i.e.,  $\lambda_2(\mathbf{L}) > 0$ .

*Sensing and Communication Noises:* We now consider the measurements made by the agents. Depending on the underlying hypothesis, agent  $i$  at time  $k$  makes the observation

$$H_1 : y_i(k) = m_i + \zeta_i(k) \quad (7)$$

$$H_0 : y_i(k) = \zeta_i(k). \quad (8)$$

The prior probabilities are  $0 < P(H_1), P(H_0) < 1$ . Here  $m_i$  is a constant known signal and  $\{\zeta_i(k)\}$  is the sensing noise. We describe the interagent communications. If  $(i, j) \in E$ , their communication channel is the standard Shannon communication model of an ideal (low pass) filter plus additive noise. For example, agent  $i$  receives from agent  $j$

$$z_{ij}(k) = x_j(k) + \nu_{ij}(k) \quad (9)$$

where  $\nu_{ij}(k)$  is the communication noise. A similar equation models the signal  $z_{ji}(k)$  received by node  $j$  from node  $i$  at time  $k$ ; note that  $\nu_{ij}(k) \neq \nu_{ji}(k)$ . In the consensus+innovations updating algorithm we analyze, we will see that the communication noise that affects the updating of the detection statistic  $x_i(k)$  of agent  $i$  at time  $k$  is actually

$$v_i(k) = \sum_{j \in O_i} \nu_{ij}(k), \quad i = 1, \dots, N. \quad (10)$$

For more compact statements, we introduce vector notation and then state our noise assumptions. Let

$$\begin{aligned} \mathbf{y}(k) &= (y_1(k), \dots, y_N(k))^\top \\ \mathbf{m} &= (m_1, \dots, m_N)^\top \end{aligned} \quad (11)$$

$$\begin{aligned} \boldsymbol{\zeta}(k) &= (\zeta_1(k), \dots, \zeta_N(k))^\top \\ \mathbf{v}(k) &= (v_1(k), \dots, v_N(k))^\top. \end{aligned} \quad (12)$$

*Assumption 2 (Sensing Noise):* The sensing noise  $\{\zeta(k)\}$  is a zero mean i.i.d. Gaussian sequence (possibly spatially correlated) with  $\text{Cov}(\zeta(k_1), \zeta(k_2)) = \mathbf{S}_\zeta \delta_{k_1 k_2}$ , with a positive definite  $\mathbf{S}_\zeta$ .

**Assumption 3 (Communication Noise):** The communication noise  $\{\nu_{ij}(k)\}$  in (9) is spatially and temporally zero mean Gaussian i.i.d. with  $\text{Cov}(\nu_{i_1 j_1}(k)\nu_{i_2 j_2}(s)) = \sigma_\nu^2 \delta_{(i_1, j_1), (i_2, j_2)} \delta_{ks}$ ; and  $v_i(k)$  and  $\zeta_j(s)$  are mutually independent over all  $i, j, k, s$ .

It is for simplicity that we let  $\nu_{ij}(k)$ 's have the same variance  $\sigma_\nu^2 := \text{Var}(\nu_{ij}(k))$  for all  $i, j, k$ . Still,  $\text{Cov}(v_i(k)) = d_i \sigma_v^2$  and  $\text{Cov}(\mathbf{v}(k)) = \sigma_v^2 \mathbf{D}$ . For a regular network,  $\text{Cov}(v_i(k)) = d \sigma_v^2$  and  $\text{Cov}(\mathbf{v}(k)) = d \sigma_v^2 \mathbf{I}$ , where  $d$  is the degree of every node  $i \in V$ .

### B. Detector: Single Agent, Centralized, and Distributed

Each agent performs the simple binary hypothesis test (6). We detail how the distributed consensus+innovations sequential detector updates the detection statistic at each agent  $i$  at each time  $k$  when Assumptions 1–3 hold; but, first, we motivate briefly its structure by considering how a single agent in isolation and an agent with access to all the data (a fusion center) update their detection statistics. Due to the Gauss noise assumptions, the test statistics are linear in all three cases.

**Single Agent: Isolated Detector:** If agent  $i$  decides in isolation, i.e., without cooperation with other agents, its detection statistic  $x_{i,\text{iso}}(k+1)$  at time  $k+1$  is

$$\begin{aligned} x_{i,\text{iso}}(k+1) &= \frac{1}{k+1} \sum_{t=1}^{k+1} \eta_i(t) \\ &= \frac{k}{k+1} x_{i,\text{iso}}(k) + \frac{1}{k+1} \eta_i(k+1) \end{aligned} \quad (13)$$

where the instantaneous log-likelihood ratio (under a diagonal  $\mathbf{S}_\zeta$ ) is  $\eta_i(k) = \frac{m_i}{[\mathbf{S}_\zeta]_{ii}} (y_i(k) - \frac{m_i}{2})$ . Equation (13) provides a sequentially (recursive) implementation of the test statistic or detection statistic  $x_{i,\text{iso}}(k)$  of agent  $i$ .

**Centralized Detector:** For the distributed detection problem we study, the centralized detector is not only motivational but also serves to benchmark its performance. It is ideal because it assumes that the observations of all agents are always available (with no communication noise) at a fusion center. The centralized detection statistic  $x_{\text{cen}}(k+1)$  is

$$\begin{aligned} x_{\text{cen}}(k+1) &= \frac{1}{k+1} \sum_{t=1}^{k+1} \frac{1}{N} \sum_{i=1}^N \eta_i(t) \\ &= \frac{k}{k+1} x_{\text{cen}}(k) + \frac{1}{k+1} \frac{1}{N} \sum_{i=1}^N \eta_i(k+1) \end{aligned} \quad (14)$$

where

$$\eta_i(t) = [\mathbf{S}_\zeta^{-1} \mathbf{m}]_i \left( y_i(t) - \frac{m_i}{2} \right). \quad (16)$$

To emphasize the structure in (14), we reverse the order of the summations. When the noise is spatially uncorrelated,  $\eta_i(t) = \frac{m_i}{[\mathbf{S}_\zeta]_{ii}} (y_i(t) - \frac{m_i}{2})$ , and the normalized by  $\frac{1}{k+1}$  sum in the time index  $k$  in (14) is the detection statistic  $x_{i,\text{iso}}(k)$  given by (13). We get

$$\begin{aligned} x_{\text{cen}}(k+1) &= \frac{1}{N} \sum_{i=1}^N x_{i,\text{iso}}(k+1) = \frac{k}{k+1} \frac{1}{N} \sum_{i=1}^N x_{i,\text{iso}}(k) \\ &\quad + \frac{1}{k+1} \frac{1}{N} \sum_{i=1}^N \eta_i(k+1). \end{aligned} \quad (17)$$

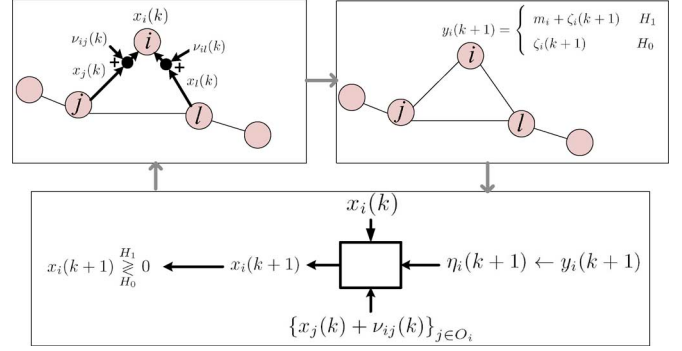


Fig. 1. System diagram and illustration of the operation of the consensus+innovations distributed detector in (19).

Equation (17) makes explicit that  $x_{\text{cen}}(k)$  in the RHS of (15) is equivalent to an average of local quantities.

**Distributed Sequential Detector: Consensus+Innovations:** We are guided by the structure of (15) or (17). Because the graph is sparse, the (spatial) averages on the RHS of (17) (or  $x_{\text{cen}}(k)$  in the RHS of (15)) cannot be performed at each time  $k$  at  $i$  since not all  $\{x_{i,\text{iso}}(k)\}$  and  $\{\eta_i(k)\}$  are available at  $i$ . We propose replacing the spatial averaging in the first term of the RHS in (17) (or  $x_{\text{cen}}(k)$  in the RHS of (15)) by consensus, and substituting the spatial averaging of the innovations over all agents (second term in the RHS in (17) or in (15)) by the instantaneous innovations  $\eta_i(k)$  of agent  $i$ . With consensus, each agent weight averages its state (detection statistic) with the states received from its neighbors. We assume that every agent uses at time  $k$  the same weight  $\alpha_k$ —the so called equal weights consensus. The updating for the distributed test statistic at agent  $i$  is then

$$\begin{aligned} x_i(k+1) &= \frac{k}{k+1} \left[ x_i(k) + \underbrace{\alpha_k \sum_{j \in O_i} [(x_j(k) - x_i(k)) + \nu_{ij}(k)]}_{\text{consensus}} \right] \\ &\quad + \frac{1}{k+1} \underbrace{\eta_i(k+1)}_{\text{"innovations"}}, \quad i = 1 \dots N \end{aligned} \quad (18)$$

$$\begin{aligned} &= \frac{k}{k+1} (1 - d_i \alpha_k) x_i(k) + \frac{k}{k+1} \alpha_k \sum_{j \in O_i} (x_j(k) + \nu_{ij}(k)) \\ &\quad + \frac{1}{k+1} \eta_i(k+1), \quad x_i(1) = \eta_i(1) \end{aligned} \quad (19)$$

where for the correlated case the innovations  $\eta_i(k+1)$ , given by (16), are still computed from the local measurement available to agent  $i$ . Fig. 1 gives a system diagram and illustrates the operation of the distributed detector (18). In (18), because of the noisy communications among agents, the communication noise  $\nu_{ij}(k)$  appears explicitly inside the summation. Equation (19) follows from (18) by recognizing the neighborhood  $O_i$  of agent  $i$  and its degree  $d_i$ . Equation (18), our approximation to the centralized update in (15), shows explicitly the “consensus” and the “innovations” terms, occurring in the same time step  $k$ . We refer to the distributed sequential detector with this updating as the consensus+innovations distributed sequential detector.

We choose and analyze in this paper what we will call the Dirichlet family of weights [29] (see also footnote 2)

$$\alpha_k = \frac{b_0 \geq 0}{a + k^\tau}, \quad b_0, a \geq b_0 \lambda_N(\mathbf{L}), \tau \geq 0. \quad (20)$$

The parameters  $a$  and  $b_0$  provide additional degrees of freedom that we will use to optimize the performance of the distributed detector. We will see that exponential error rate results when  $\tau = 1$ .

*Matrix Format:* Before establishing the exponential error rate for the distributed detector, we write the updating (19) more compactly, in matrix form. We complete the vector notation introduced in (11)–(12) by defining the decision vector  $\mathbf{x}(k)$ , the innovations vector  $\boldsymbol{\eta}(k)$ , and the weight matrix  $\mathbf{W}(k)$ :

$$\begin{aligned} \mathbf{x}(k) &= (x_1(k), \dots, x_N(k))^\top \\ \boldsymbol{\eta}(k) &= (\eta_1(k), \dots, \eta_N(k))^\top \\ \mathbf{W}(k) &= \mathbf{I} - \alpha_k \mathbf{L} \end{aligned} \quad (21)$$

where  $\mathbf{L}$  is the graph Laplacian matrix. From (16), the innovations vector and its statistics are

$$\boldsymbol{\eta}(k) = \text{Diag}(\mathbf{S}_\zeta^{-1} \mathbf{m}) \left( \mathbf{y}(k) - \frac{1}{2} \mathbf{m} \right) \quad (22)$$

$$\begin{aligned} \mathbb{E}[\boldsymbol{\eta}(k) | H_1] &= -\mathbb{E}[\boldsymbol{\eta}(k) | H_0] := \mathbf{m}_\eta \\ &= \frac{1}{2} \text{Diag}(\mathbf{S}_\zeta^{-1} \mathbf{m}) \mathbf{m} \end{aligned} \quad (23)$$

$$\text{Cov}(\boldsymbol{\eta}(k)) := \mathbf{S}_\eta = \text{Diag}(\mathbf{S}_\zeta^{-1} \mathbf{m}) \mathbf{S}_\zeta \text{Diag}(\mathbf{S}_\zeta^{-1} \mathbf{m}). \quad (24)$$

Recall the communication noises  $v_i(k)$  and communication noise vector  $\mathbf{v}(k)$  in (10) and (12), respectively. Then, in vector form, the distributed updating (19) is

$$\begin{aligned} \mathbf{x}(k+1) &= \frac{k}{k+1} \mathbf{W}(k) \mathbf{x}(k) + \frac{k}{k+1} \alpha_k \mathbf{v}(k) \\ &+ \frac{1}{k+1} \boldsymbol{\eta}(k+1), \quad k = 1, 2, \dots, \mathbf{x}(1) = \boldsymbol{\eta}(1). \end{aligned} \quad (25)$$

*Solution to (25):* Define the matrices  $\Phi(k, j)$ ,  $k \geq j \geq 1$ , by  $\Phi(k, j) := \mathbf{W}(k-1) \mathbf{W}(k-2) \dots \mathbf{W}(j)$ , if  $1 \leq j < k$ , and  $\Phi(k, k) := \mathbf{I}$ . Then, after algebraic manipulations, the solution to the distributed detector (25) is

$$\begin{aligned} \mathbf{x}(k) &= \frac{1}{k} \sum_{j=1}^k \Phi(k, j) \boldsymbol{\eta}(j) + \frac{1}{k} \sum_{j=1}^{k-1} (j \alpha_j) \Phi(k, j+1) \mathbf{v}(j), \\ &k = 1, 2, 3, \dots \end{aligned} \quad (26)$$

The matrices  $\Phi(k, j)$  are doubly stochastic for all  $k, j$ , with unit norm. Hence, all the innovations  $\boldsymbol{\eta}(j)$ ,  $j = 1, 2, \dots$ , are scaled by the same weight order  $\sim 1/k$ —as with the centralized detector (14).

### III. PERFORMANCE ANALYSIS OF THE DISTRIBUTED DETECTOR

#### A. Error Performance: Large Deviations

We are interested in the large deviations error performance of the distributed detector (6) with updating (19), i.e., in the exponential decay rate at each agent  $i$ :

$$\lim_{k \rightarrow \infty} -\frac{1}{k} \log P_{\text{dis},i}^e(k) \quad (27)$$

of the (average) error probability  $P_{\text{dis},i}^e(k) = \mathbb{P}(x_i(k) > 0 | H_0)P(H_0) + \mathbb{P}(x_i(k) \leq 0 | H_1)P(H_1)$ . We are also interested in studying how the sensing noise, the communications noise, and the connectivity of the network affect this large deviations error performance.

*Remark:* If the limit above is zero, the error probability either decays to zero slower than exponentially or remains non zero.

Under the Gauss Assumption 2, Chernoff Lemma, [28], gives the exponential decay rate of the error probability of the isolated and centralized detectors as

$$\begin{aligned} \lim_{k \rightarrow \infty} -\frac{1}{k} \log P_{\text{iso}}^e(k) &= \frac{1}{8} \frac{|m_i|^2}{([\mathbf{S}_\zeta]_{ii})^2} = \frac{1}{8} \gamma_{s,i}, \\ \lim_{k \rightarrow \infty} -\frac{1}{k} \log P_{\text{cen}}^e(k) &= \frac{1}{8} \mathbf{m}^\top \mathbf{S}_\zeta^{-1} \mathbf{m} = \frac{1}{8} \gamma_s. \end{aligned} \quad (28)$$

We refer to  $\gamma_{s,i}$  and  $\gamma_s$  as the isolated and centralized detectors sensing signal-to-noise (SNR) ratios.

*Global and Local Detectability:* Equation (28) shows that the error probability of the centralized detector decays exponentially fast to zero if and only if  $\gamma_s > 0$ . This requires that  $\mathbf{m} \neq \mathbf{0}$ , which is a trivial observation since if  $\mathbf{m} \equiv \mathbf{0}$  then the statistics under the two hypotheses were undistinguishable and the test is degenerate. A similar comment holds for an individual isolated agent  $i$ , where now the condition is that the corresponding  $\gamma_{s,i} > 0$ . For future reference, we formalize these comments.

*Definition 1 (Global and Local Detectability):* The hypothesis testing problem (7)–(8) is *globally detectable* if the error probability of the centralized detector decays exponentially fast, i.e.,  $\gamma_s > 0$ . An agent  $i$  is *locally detectable* if the probability of error of its isolated detector decays exponentially fast, i.e.,  $\gamma_{s,i} > 0$ .

In the sequel, we will have occasion to consider spatially uncorrelated noises and identical agents.

*Definition 2 (Identical Agents):* The sensing noise is spatially uncorrelated ( $\mathbf{S}_\zeta$  is diagonal) and  $\gamma_{s,i} = \gamma_{s,j} > 0$ , for all  $i \neq j$ .

*Distributed Detector:* It is not trivial to obtain the results similar to (28) for the distributed detector or to conclude that its error probability achieves exponential decay rate at *all* agents because agents may be locally undetectable, the information available at each agent is limited, and the communication among agents is noisy. We find here a generic expression for this probability of error, while in the next Section III-B, we actually compute this in terms of the noise and network parameters.

Because of the linearity of its updating and the Gauss assumptions on the noises, the decision vector  $\mathbf{x}(k)$  is Gauss. Also, for the simple binary hypothesis test (7)–(8), the distribution of  $\mathbf{x}(k)$  under  $H_0$  is the same as the distribution of  $-\mathbf{x}(k)$  under  $H_1$ . So, the statistics of  $\mathbf{x}(k)$  under either hypothesis are found from the first and second moments of the decision vector under, say, hypothesis  $H_1$ . We let  $\boldsymbol{\mu}(k) = \mathbb{E}[\mathbf{x}(k) | H_1] = -\mathbb{E}[\mathbf{x}(k) | H_0]$  and  $\boldsymbol{\Sigma}(k) = \text{Cov}(\mathbf{x}(k) | H_1) = \text{Cov}(\mathbf{x}(k) | H_0)$ . From these comments and definitions, it follows that  $\mathbb{P}(x_i(k) \leq 0 | H_1) = \mathbb{P}(-x_i(k) \geq 0 | H_1) = \mathbb{P}(x_i(k) > 0 | H_0)$ . Letting  $\mu_i(k)$  and  $\sigma_i^2(k)$  be the mean and variance of  $x_i(k)$ , it is straightforward to get

$$P_{\text{dis},i}^e(k) = \mathbb{P}(x_i(k) > 0 | H_0) = \mathcal{Q} \left( \frac{\mu_i(k)}{\sigma_i(k)} \right).$$

Taking the log, dividing by  $k$ , applying the inequalities for  $Q(t)$  in (5), and letting  $k \rightarrow \infty$ , we get

$$\lim_{k \rightarrow \infty} -\frac{1}{k} \log P_{\text{dis},i}^e(k) = \frac{1}{2} \frac{[\boldsymbol{\mu}_\infty]_i^{2,+}}{[\boldsymbol{\Sigma}_\infty]_{ii}} \quad (29)$$

where  $z^{2,+} = (\max\{0, z\})^2$ , and  $[\cdot]_i$  and  $[\cdot]_{ii}$  are the  $i$ th-entry and the  $i$ th-diagonal entry, respectively, of

$$\boldsymbol{\mu}_\infty := \lim_{k \rightarrow \infty} \boldsymbol{\mu}(k), \quad \boldsymbol{\Sigma}_\infty := \lim_{k \rightarrow \infty} k \boldsymbol{\Sigma}(k). \quad (30)$$

Section III-B finds  $\boldsymbol{\mu}_\infty$  and  $\boldsymbol{\Sigma}_\infty$ , and then it finds the rate (29) for the distributed detector (Theorem 3 and Corollary 4).

### B. Exponential Decay of the Error Probability of the Distributed Detector

*Theorem 3 (Asymptotics on the Moments of  $x_i(k)$ ):* Consider the consensus+innovations distributed detector under Assumptions 1–3. Let  $\tau = 1$  in the weights  $\alpha_k$  in (20). Further, assume that the test is globally detectable. Then, we have

$$\boldsymbol{\mu}_\infty = \lim_{k \rightarrow \infty} \boldsymbol{\mu}(k) = (\mathbf{I} + b_0 \mathbf{L})^{-1} \mathbf{m}_\eta \quad (31)$$

$$\lim_{k \rightarrow \infty} \mu_i(k) \geq \frac{1}{2N} \gamma_s \left( 1 - \frac{1}{1 + b_0 \lambda_2(\mathbf{L})} \frac{2N \|\mathbf{m}_\eta\|}{\gamma_s} \right), \quad \forall i \quad (32)$$

$$\boldsymbol{\Sigma}_\infty = \lim_{k \rightarrow \infty} k \boldsymbol{\Sigma}(k) = \text{Vec}^{-1} \left( (\mathbf{I} + b_0 (\mathbf{L} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{L}))^{-1} \right. \\ \left. \times (\text{Vec}(\mathbf{S}_\eta) + b_0^2 \text{Vec}(\mathbf{S}_v)) \right) \quad (33)$$

$$\limsup_{k \rightarrow \infty} k \sigma_i^2(k) \leq \frac{1}{N^2} \gamma_s \\ \times \left( 1 + 3 \frac{N^2}{1 + b_0 \lambda_2(\mathbf{L})} \frac{\|\mathbf{S}_\eta\|}{\gamma_s} + \frac{N^2 b_0^2 \|\mathbf{S}_v\|}{\gamma_s} \right), \quad \forall i. \quad (34)$$

*Corollary 4 (Exponential Decay of the Error Probability):* Consider the consensus+innovations distributed detector under Assumptions 1–3. Let  $\tau = 1$  in the weights  $\alpha_k$  in (20). Further assume that the test is globally detectable. Then

$$\lim_{k \rightarrow \infty} -\frac{1}{k} \log P_{\text{dis},i}^e(k) \\ \geq \frac{1}{8} \gamma_s \frac{\left\{ 1 - \frac{1}{1 + b_0 \lambda_2(\mathbf{L})} \frac{2N \|\mathbf{m}_\eta\|}{\gamma_s} \right\}^{2,+}}{1 + 3 \frac{N^2}{1 + b_0 \lambda_2(\mathbf{L})} \frac{\|\mathbf{S}_\eta\|}{\gamma_s} + \frac{N^2 b_0^2 \|\mathbf{S}_v\|}{\gamma_s}}, \quad \forall i, \quad (35)$$

where, as before,  $z^{2,+} = (\max\{0, z\})^2$ . Further, for any  $b_0 > b_L = \max \left\{ 0, \frac{\frac{2N \|\mathbf{m}_\eta\|}{\gamma_s} - 1}{\lambda_2(\mathbf{L})} \right\}$ , and  $a \geq b_0 \lambda_N(\mathbf{L})$ , the lower bound in (35) is strictly positive, i.e., each agent achieves the exponential decay rate.<sup>2</sup> If, in addition, the agents are identical:

$$\lim_{k \rightarrow \infty} -\frac{1}{k} \log P_{\text{dis},i}^e(k) \\ \geq \frac{1}{8} \gamma_s \frac{1}{\left( 1 + \frac{N}{1 + b_0 \lambda_2(\mathbf{L})} + \frac{N^2 b_0^2 \|\mathbf{S}_v\|}{\gamma_s} \right)} \quad (36)$$

$$\geq \frac{1}{8} \gamma_s \frac{1}{\left( 1 + \frac{N}{b_0 \lambda_2(\mathbf{L})} + \frac{N^2 b_0^2 \|\mathbf{S}_v\|}{\gamma_s} \right)}, \quad (37)$$

<sup>2</sup>The requirement  $a \geq b_0 \lambda_N(\mathbf{L})$  is not crucial for the detector operation; it only serves to assure that  $\|\mathbf{W}(k) - \mathbf{J}\| = 1 - \frac{b_0 \lambda_2(\mathbf{L})}{a+k}$ , for all  $k$ ; when we allow for any  $a > 0$ , the latter norm equality still holds eventually.

and hence for any  $b_0 > 0$  and  $a \geq b_0 \lambda_N(\mathbf{L})$ , each agent achieves the exponential decay rate.

*Remark:* Under global detectability and connectedness, Corollary 4 states that the error probability  $P_{\text{dis},i}^e$  at every agent  $i$  decays exponentially to zero even if agent  $i$  is (in isolation) not detectable ( $\gamma_{s,i} = 0$ ,) and even when the communication links are very noisy (high  $\|\mathbf{S}_v\|$ ).

*Remark.* The agents need only to bound  $b_L$  and  $\lambda_N(\mathbf{L})$  to be able to choose admissible  $b_0$  and  $a$  as we show now. We first discuss  $b_0$ . We devise an upper bound  $b'_L$  on  $b_L$  that requires less knowledge than what is required to compute  $b_L$ . Lower bounding the algebraic connectivity  $\lambda_2(\mathbf{L}) \geq \frac{4}{N \text{diam}(\mathcal{G})} \geq \frac{4}{N^2}$ , where  $\text{diam}(\mathcal{G})$  is the graph diameter, and recalling the norm inequalities for a  $N \times 1$  vector  $\mathbf{a}$  and a  $N \times N$  matrix  $\mathbf{A}$ :  $\|\mathbf{a}\| \leq \sqrt{N} \max_{i=1,\dots,N} |a_i|$ ,  $\|\mathbf{A}\|_1 \leq \sqrt{N} \|\mathbf{A}\|_2$ , we get

$$b_L \leq \frac{2 \|\mathbf{m}_\eta\|}{(1/N) \gamma_s \lambda_2(\mathbf{L})} \\ \leq \frac{\|(m_1 \mathbf{m}^\top [\mathbf{S}_\zeta^{-1}]_1, m_2 \mathbf{m}^\top [\mathbf{S}_\zeta^{-1}]_2, \dots, m_N \mathbf{m}^\top [\mathbf{S}_\zeta^{-1}]_N)^\top\|}{\frac{1}{N} \mathbf{m}^\top \mathbf{S}_\zeta^{-1} \mathbf{m} \frac{4}{N^2}} \\ \leq \frac{N^3 \sqrt{N} \max_{i=1,\dots,N} |m_i \mathbf{m}^\top [\mathbf{S}_\zeta^{-1}]_i|}{4 \mathbf{m}^\top \mathbf{S}_\zeta^{-1} \mathbf{m}} \\ \leq \frac{N^4 \|\mathbf{m}\|^2 \lambda_N(\mathbf{S}_\zeta^{-1})}{4 \|\mathbf{m}\|^2 \lambda_1(\mathbf{S}_\zeta^{-1})} = \frac{N^4 \lambda_N(\mathbf{S}_\zeta)}{4 \lambda_1(\mathbf{S}_\zeta)} =: b'_L.$$

Hence, each agent needs to know only an upper bound on the number of agents  $N$  and an upper bound on the condition number of  $\mathbf{S}_\zeta$ . Once an admissible  $b_0$  is set, agents can use the inequality  $\lambda_N(\mathbf{L}) \leq 2N$  to set:  $a \geq 2N b_0$ .

### C. Optimality of the Decay Rate of $\alpha_k$

We now show that the choice of  $\tau = 1$  in the weights  $\alpha_k$  in (20) gives the tightest exponential error rate in the class of weight sequences  $\beta_k = \frac{b_0}{a+k^\tau}$ , parameterized by  $\tau \geq 0$ . (The proof is in Appendix B.)

*Theorem 5:* Let Assumptions 1–3 hold and assume the agents are identical. Then for the weight sequence  $\beta_k = \frac{b_0}{a+k^\tau}$ ,  $a \geq b_0 \lambda_N(\mathbf{L})$ ,  $b_0 > 0$ , where  $\tau \geq 0$ :

$$\limsup_{k \rightarrow \infty} -\frac{1}{k} \log P_{\text{dis},i}^e(k) \\ \leq \begin{cases} 0 & \text{if } \tau < 1 \\ \frac{1}{8} \gamma_{s,i} & \text{if } \tau > 1 \\ \frac{1}{8} \gamma_s \frac{1}{1 + \frac{N-1}{1+2b_0 \lambda_N(\mathbf{L})} + \frac{b_0^2 \lambda_1(\mathbf{S}_v)}{N \gamma_{s,i}}} & \text{if } \tau = 1. \end{cases} \quad (38)$$

*Remark:* Note that, for a strong communication noise (large  $\lambda_1(\mathbf{S}_v)$ ), the bound in (38) for  $\tau = 1$  becomes smaller than  $\frac{1}{8} \gamma_{s,i}$ . The latter, according to (38), means that the noise is so strong that no choice of  $\tau \geq 0$  improves the detector's performance over the isolated detector's performance ( $\frac{1}{8} \gamma_{s,i}$ ). In other words, when the bound in (38) for  $\tau = 1$  is smaller than  $\frac{1}{8} \gamma_{s,i}$ , it does not pay off for agents to communicate. Equation (38) shows, however, that, as long as the communication noise is small enough so that there exists a choice  $\tau \geq 0$  that allows for an improvement over the isolated detector, then this choice has to be  $\tau = 1$ . We adopt this choice in the paper.

#### IV. INTERPRETATIONS OF THE RESULTS AND SIMULATION STUDIES

We present simulations and insights into distributed detection over noisy links. First, we demonstrate that distributed detection performance depends on the choice of the weights  $\alpha_k = b_0/(a+k)$ —in particular, on the choice of  $b_0$ . We find numerically the optimal  $b_0$ , and, for identical agents, we give a closed form expression for a suboptimal  $b_0$ . Second, we show that there is a threshold on the communication noise power below which it pays off for agents to communicate. This threshold occurs at a high communication noise level, so, in many applications it is advantageous to cooperate. Finally, we study how the total transmission power invested by agents during the algorithm run affects the time to decision—the time when the error probability falls below a prescribed value  $\epsilon$ . Section IV-A introduces metrics that help us assess distributed detection performance, and Section IV-B presents results.

##### A. Metrics and Simulation Setup

We benchmark the distributed detector against the centralized and isolated detectors. For intuitive comparisons, we introduce the following metrics.

*Gain of the Distributed Detector Over the Isolated Detector:* The gain  $g_{\text{dis:iso}}$  of the distributed detector over the isolated detector is the ratio of the performance of the *worst* agent with communication and the *best* agent without communication:

$$g_{\text{dis:iso}} = \frac{\min_{i=1,\dots,N} \left\{ \lim_{k \rightarrow \infty} -\frac{1}{k} \log P_{\text{dis},i}^e(k) \right\}}{\max_{i=1,\dots,N} \left\{ \frac{1}{8} \gamma_{s,i} \right\}}. \quad (39)$$

The network  $\mathcal{G} = (\mathcal{V}, E)$  achieves *communication payoff* if  $g_{\text{dis:iso}} > 1$ .

*Ratio of the Performance of the Distributed Detector and the Centralized Detector:* We define  $r_{\text{dis:cen}}$  as the ratio of the exponential decay rate of the *worst* agent and the exponential decay rate of the centralized detector:

$$r_{\text{dis:cen}} = \frac{\min_{i=1,\dots,N} \left\{ \lim_{k \rightarrow \infty} -\frac{1}{k} \log P_{\text{dis},i}^e(k) \right\}}{\frac{1}{8} \gamma_s}. \quad (40)$$

*Communication Signal-to-Noise Ratio:* To assess the quality of the interagent communication channels, i.e., the relation between the amount of the communication noise power and the transmission power that agents use, we define the communication signal to noise ratio at time  $k$  by

$$\begin{aligned} \text{CSNR}'(k) &= \frac{\sum_{i=1}^N \sum_{j \in O_i} \mathbb{E}[x_i^2(k) | H_1]}{\sum_{i=1}^N \sum_{j \in O_i} \text{Var}(\nu_{ij}^2(k))} \\ &= \frac{\sum_{i=1}^N d_i (\mu_i^2(k) + \sigma_i^2(k))}{\sum_{i=1}^N d_i \sigma_\nu^2}. \end{aligned}$$

For large  $k$ ,  $\mu_i^2(k) + \sigma_i^2(k) \approx [\boldsymbol{\mu}_\infty]_i^2$  (as the variance vanishes); hence, for simplicity, we consider

$$\text{CSNR} = \frac{\sum_{i=1}^N d_i [\boldsymbol{\mu}_\infty]_i^2}{\sum_{i=1}^N d_i \sigma_\nu^2}, \quad (41)$$

i.e., the “steady state” approximate  $\text{CSNR}'(k)$  for large  $k$ . We vary  $\sigma_\nu^2$  and investigate how this affects distributed detection performance.

*Remark.* In practice, we may vary  $\sigma_\nu^2$  (cleaner channels) or control transmission power. We illustrate this for example for amplify and forward analog communication (see [30]), where agent  $j$  receives from agent  $i$  at time  $k$ :

$$A_{ij} h_{ij} x_j(k) + \nu_{ij}(k).$$

Here  $A_{ij}$  is the amplify transmission coefficient, and  $h_{ij}$  is the channel gain. Agent  $i$  knows  $A_{ij} h_{ij}$ ; it estimates it, e.g., with a pilot signal. The average transmission power spent by agent  $i$  for transmission to agent  $j$  at time  $k$  (assuming  $h_{ij} = 1$ ) is  $A_{ij}^2 \mathbb{E}[x_i^2(k) | H_1]$ , and hence (assuming equal gains  $A_{ij} = A$  over links)  $\text{CSNR}'(k) = A^2 \frac{\sum_{i=1}^N d_i \mathbb{E}[x_i^2(k) | H_1]}{\sum_{i=1}^N \text{Var}(\nu_i^2(k))} = \frac{\sum_{i=1}^N d_i \mathbb{E}[x_i^2(k) | H_1]}{\sum_{i=1}^N d_i (\sigma_\nu^2 / A^2)}$ . Technically, we vary the parameter  $\sigma_\nu^2$  and let  $A = 1$ ; but this is equivalent to keeping  $\sigma_\nu^2$  fixed and varying  $A$ , and hence varying the transmission power.

*Normalized Time to Decision Versus Total Communication Signal-to-Noise Ratio:* For a certain small target error probability  $\epsilon$ , say  $10^{-4}$ , for a certain detector, distributed or centralized, we define the time to decision  $T_{\text{det,dis}}$  by:

$$T_{\text{det,dis}} := \inf_{k=1,2,\dots} \left\{ \max_{i=1,\dots,N} P_{i,\text{dis}}^e(k) < \epsilon \right\},$$

i.e.,  $T_{\text{det,dis}}$  is the smallest time step  $k$  at which the worst agent's error probability falls below  $\epsilon$ . We define similarly  $T_{\text{det,cen}}$  for the centralized detector. We consider

$$T := T_{\text{det,dis}} / T_{\text{det,cen}}, \quad (42)$$

the time to decision of the distributed detector normalized by the time to decision of the centralized detector. We define the total transmission signal to noise ratio for the distributed detector by

$$\text{CSNR}_{\text{tot}} := \frac{\sum_{k=1}^{T_{\text{det,dis}}} \sum_{i=1}^N d_i \mathbb{E}[x_i^2(k) | H_1]}{\sum_{k=1}^{T_{\text{det,dis}}} \sum_{i=1}^N d_i \sigma_\nu^2}, \quad (43)$$

i.e.,  $\text{CSNR}_{\text{tot}}$  is the ratio of the total transmission power invested over  $k = 1, \dots, T_{\text{det,dis}}$  normalized by the total communication noise power injected over  $k = 1, \dots, T_{\text{det,dis}}$ . Our goal is to vary the communication noise variance  $\sigma_\nu^2$  and establish the operation characteristic  $T$  versus  $\text{CSNR}_{\text{tot}}$ .

*Simulation Setup:* We consider a network with  $N = 10$  agents. We consider two sets of experiments: 1) asymmetric network: unequal agents and irregular graph; and 2) symmetric network: equal agents and regular graph. With the asymmetric network, we construct the sensing signal vector  $\mathbf{m}$  to be sparse, so that certain agents are not locally detectable. We generate each entry  $m_i$  of the signal vector  $\mathbf{m}$  randomly; with probability 0.4,  $m_i = 0$ , and with probability 0.6, it is drawn from the uniform distribution on  $[0, 1]$ . Once we generate all entries  $m_i$  and get  $\mathbf{m}$ ,  $i = 1, \dots, N$ , we normalize  $\mathbf{m}$  to have norm  $\|\mathbf{m}\|^2 = 0.1$ . For the particular realization of  $\mathbf{m}$  it turned out that  $\mathbf{m}$  has five zero entries, so that five agents are not locally detectable. We set the sensing covariance  $\mathbf{S}_\zeta = \mathbf{I}$ . With the symmetric network, we set  $\mathbf{m}$  to have equal entries and we normalize  $\mathbf{m}$  so that  $\|\mathbf{m}\|^2 = 0.1$ ; we set  $\mathbf{S}_\zeta = \mathbf{I}$ . With the asymmetric network, the network is a geometric graph. We place the nodes uniformly on a unit square and connect those pairs of nodes with interagent

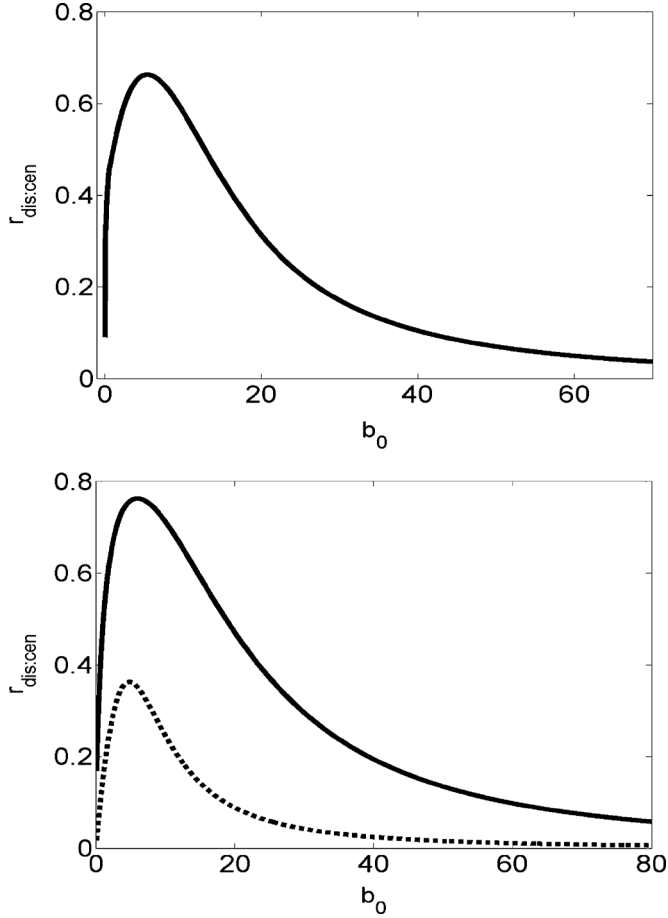


Fig. 2. Ratio  $r_{\text{dis:cen}}$  in (40) of the performance of the distributed detector and the centralized detector versus parameter  $b_0$ . Top: asymmetric network; Bottom: symmetric network. The bottom figure plots in dotted line the lower bound on  $r_{\text{dis:cen}}$  based on (37).

distance less than a given radius. The resulting network has 42 (undirected) links. With the symmetric network, we generate a regular graph with degree 4. With all experiments, we set the communication noise matrix  $\mathbf{S}_v = \sigma_v^2 \mathbf{D}$ , and  $a = b_0 \lambda_N(\mathbf{L})$ .

## B. Results

*Dependence of the Detection Performance on the Parameter  $b_0$ :* Fig. 2(top) plots the ratio  $r_{\text{dis:cen}}$  in (40) versus the parameter  $b_0$ , for the asymmetric network and CSNR = -3.22 dB. We can see that the choice of  $b_0$  significantly affects the performance of the distributed detector; hence, it is very important to select appropriately  $b_0$ .

The intuition behind the optimal choice of  $b_0$  lies in the tradeoff between the communication noise and the information flow; it can be understood from the lower bound (35). Namely, the terms

$$\frac{1}{1 + b_0 \lambda_2(\mathbf{L})} \frac{2N \|\mathbf{m}_\eta\|}{\gamma_s}, \quad 3 \frac{N^2}{1 + b_0 \lambda_2(\mathbf{L})} \frac{\|\mathbf{S}_\eta\|}{\gamma_s} \quad (44)$$

$$\frac{N^2 b_0^2 \|\mathbf{S}_v\|}{\gamma_s} \quad (45)$$

decrease the bound and so they quantify the decrease in performance of the distributed detector; the decrease comes from two effects: 1) communication noise and 2) insufficient information

flow. From (44)–(45), we can see how the parameter  $b_0$  affects in opposing ways these two effects: The terms (44) relate to the information flow, while the term (45) is due to communication noise. We see that the net effect of increasing  $b_0$  is to increase the effective algebraic connectivity ( $b_0$  multiplies  $\lambda_2(\mathbf{L})$ ), decreasing (44); on the other hand, it increases the communication noise term in (45). The optimal  $b_0$  balances these two.

Fig. 2(bottom) plots the ratio  $r_{\text{dis:cen}}$  in (40) versus the parameter  $b_0$  for the symmetric network; it also plots the lower bound on  $r_{\text{dis:cen}}$  based on (37). We have that CSNR  $\approx$  0 dB. We can see that, again, the choice of  $b_0$  significantly affects the performance; we have that the optimal  $b_0^* = 6$ , and the highest  $r_{\text{dis:cen}}^* \approx 0.76$ . We can find a suboptimal value of  $b_0$  in closed form by optimizing the bound in (37):  $b_0^* = \frac{1}{\lambda_2(\mathbf{L})^{1/3} 2^{1/3}} \left( \frac{\gamma_s}{N \|\mathbf{S}_v\|} \right)^{1/3}$ , which, for this numerical example, equals  $\approx 4.8$ . Denote by  $c_0 = 2^{1/3} + (1/2)^{2/3} \approx 1.89$ . The corresponding closed form expression for the lower bound on  $r_{\text{dis:cen}}$  is:  $1/(1 + \frac{N c_0}{\lambda_2(\mathbf{L})^{2/3} (\frac{\gamma_s}{N \|\mathbf{S}_v\|})^{1/3}})$ .

*Communication Gain and Payoff Versus CSNR:* Fig. 3(top) plots the gain  $g_{\text{dis:iso}}$  in (39) versus CSNR in (41), and the lower bound on  $g_{\text{dis:iso}}$  based on (37), for the asymmetric network. We numerically find a suboptimal parameter  $b_0$ , for each CSNR, via grid search over an interval. First, we can see that the network achieves a communication payoff for CSNR  $\geq -10.2$  dB; and the payoff threshold is CSNR = -10.2 dB. Thus, cooperation with our distributed detector enhances detection performance even under a high communication noise. For this example, the maximal possible value of  $g_{\text{dis:iso}}$  is  $\frac{\frac{1}{8} \gamma_s}{\max_{i=1, \dots, N} \frac{1}{8} \gamma_s, i} \approx 2.2$ . Clearly, when CSNR grows large,  $g_{\text{dis:iso}}$  approaches 2.2, i.e., the distributed detector approaches the centralized detector. On the other hand, when CSNR is very small, communication is too noisy and hence does not pay off. Fig. 3(center) plots the gain versus CSNR for the symmetric network and suboptimal  $b_0 = b_0^*$ ; the figure also plots the lower bounds (37) and (36) and the upper bound (38). The communication payoff threshold is CSNR = -34 dB. Based on the bound (37), we can obtain an analytic, conservative necessary condition for the communication payoff:

$$\|\mathbf{S}_v\| \leq \frac{\gamma_s}{N} \left( \frac{N-1}{N c_0} \right)^3 (\lambda_2(\mathbf{L}))^2$$

which, for this example, yields CSNR  $\approx -15$  dB. Note that, as agents are identical, the maximal possible value of  $g_{\text{dis:iso}}$  equals  $N = 10$ .

*Operation Characteristic: Normalized Time to Decision  $T$  Versus the Total Communication Signal-to-Noise Ratio CSNR<sub>tot</sub>:* Fig. 3(bottom) plots  $T$  versus CSNR<sub>tot</sub> for the symmetric network. First, we can see that, in order to have smaller  $T$  (distributed detector's time almost as small as the centralized detector's time), we need to spend higher overall transmission power during the detector operation (higher CSNR<sub>tot</sub>.) Second, there is an interesting highly nonlinear, knee-like behavior. If we invest too small amount of transmission power, of order CSNR<sub>tot</sub> < -20 dB, the decision time increases fast. However, there is a sufficient transmission power, CSNR<sub>tot</sub>  $\approx -20$  dB, at which we are reasonably close to the decision time of the centralized detector. Further



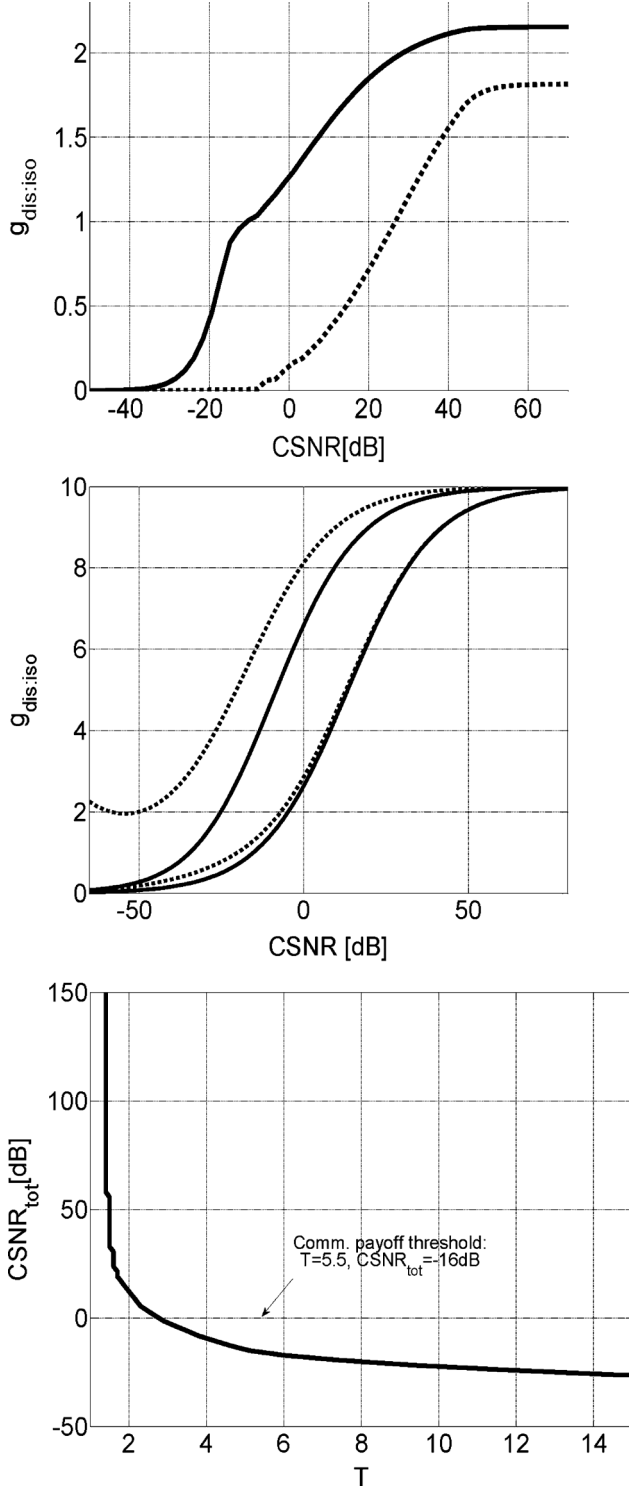


Fig. 3. Gain  $g_{\text{dis:iso}}$  in (39) of the distributed detector over the isolated detector versus the communication signal-to-noise ratio CSNR in (41). **Top:** asymmetric network, the Figure also plots the lower bound in (35); **Center:** symmetric network, the Figure also plots the lower bounds (37) and (36), and the upper bound in (38). **Bottom:** Total communication signal-to-noise ratio  $\text{CSNR}_{\text{tot}}$  in (43) versus the relative time to decision  $T$  in (42) for the symmetric network. For the isolated detector,  $T = 5.5$ .

increase of the transmission power does not shorten significantly the time to decision, as we are already close to the (minimal possible) time to decision of the centralized detector. (Note the vertical asymptote.) The operation characteristic in

Fig. 3(bottom) is valuable for the agent network design: it says, e.g., how much agents' transmission power we should invest, so that the decision time is less than, e.g., twice the centralized detector's time to decision ( $T \leq 2$ .) Finally, we note that the vertical asymptote is not at  $T = 1$  but rather at  $T = 1.4$ . That is, even when  $\sigma_v^2 = 0$ , the distributed detector does not reduce its time to decision to be equal to the centralized detector's time to decision. This is the effect of finite time (i.e., before we reach the asymptotics): the distributed detector with  $\sigma_v^2 = 0$  achieves the *asymptotic* performance of the centralized detector, but it does not achieve the *finite time* performance of the centralized detector (although it is close), simply due to the limited network connectivity.

## V. CONCLUSION

We designed a consensus+innovations distributed detector that achieves exponential error rate at all agents under noisy communication links, even when certain (or most agents) in isolation cannot perform successful detection. The key is the appropriate design of the consensus time-varying weights. We parameterized in terms of several network parameters a threshold on the communication noise power above which any agent that successfully detects the event in isolation still improves its performance through cooperation over noisy links, while below which not even the best agent can improve its detection performance by cooperation. Numerical examples demonstrate the significance on detection performance of tuning the weight sequence, show communication payoff occurring already at a high noise level—and, hence, it is typically worthwhile to cooperate, and illustrate tradeoffs between the time to decision—time to reduce the error probability below a prescribed value—and the total transmission power.

## APPENDIX

### A. Proof of Theorem 3

We first prove (31), i.e., we evaluate the limit of the mean under  $H_1$  of the detection statistic vector:  $\lim_{k \rightarrow \infty} \mu(k) = \mu_\infty = (\mathbf{I} + b_0 \mathbf{L})^{-1} \mathbf{m}_\eta$ . We summarize our strategy for this proof. First, we construct the error vector  $\epsilon(k)$  w.r.t. the assumed limit value  $(\mathbf{I} + b_0 \mathbf{L})^{-1} \mathbf{m}_\eta$ :

$$\epsilon(k) = \mu(k) - (\mathbf{I} + b_0 \mathbf{L})^{-1} \mathbf{m}_\eta.$$

Clearly, our goal is then to show  $\|\epsilon(k)\| \rightarrow 0$ . We perform this proof in three steps. First, we derive a recursive equation for  $\epsilon(k)$ . Next, we decompose  $\epsilon(k)$  into its consensus subspace component and the component orthogonal to the consensus subspace:

$$\begin{aligned} \epsilon(k) &= \mathbf{J}\epsilon(k) + (\mathbf{I} - \mathbf{J})\epsilon(k) \\ &= \frac{1}{N} \mathbf{1}(\mathbf{1}^\top \epsilon(k)) + (\mathbf{I} - \mathbf{J})\epsilon(k). \end{aligned}$$

Then, we show separately that each of these two components converges to zero:

$$\mathbf{1}^\top \epsilon(k) \rightarrow 0 \quad (\text{second step}) \quad (46)$$

$$(\mathbf{I} - \mathbf{J})\epsilon(k) \rightarrow 0 \quad (\text{third step}). \quad (47)$$

Finally, combining Steps 2 and 3 above, we conclude that  $\|\epsilon(k)\| \rightarrow 0$ .

*Proof of Step 1:* We now derive the recursive equation for  $\epsilon(k)$ . By taking the expectation on the update equation for the decision vector  $\mathbf{x}(k)$  in (25), we get that  $\boldsymbol{\mu}(k)$  evolves according to the following:

$$\boldsymbol{\mu}(k+1) = \frac{k}{k+1} \mathbf{W}(k) \boldsymbol{\mu}(k) + \frac{\mathbf{m}_\eta}{k+1}. \quad (48)$$

Consider  $\epsilon(k+1) := \boldsymbol{\mu}(k+1) - (\mathbf{I} + b_0 \mathbf{L})^{-1} \mathbf{m}_\eta$ . Subtracting  $(\mathbf{I} + b_0 \mathbf{L})^{-1} \mathbf{m}_\eta$  from both sides of (48), adding and subtracting  $\frac{k}{k+1} \mathbf{W}(k)(\mathbf{I} + b_0 \mathbf{L})^{-1} \mathbf{m}_\eta$  to the RHS of (48), we get

$$\begin{aligned} \epsilon(k+1) &= \frac{k}{k+1} \mathbf{W}(k) \boldsymbol{\mu}(k) + \frac{1}{k+1} \mathbf{m}_\eta - (\mathbf{I} + b_0 \mathbf{L})^{-1} \mathbf{m}_\eta \\ &\quad + \frac{k}{k+1} \mathbf{W}(k)(\mathbf{I} + b_0 \mathbf{L})^{-1} \mathbf{m}_\eta - \frac{k}{k+1} \mathbf{W}(k)(\mathbf{I} + b_0 \mathbf{L})^{-1} \mathbf{m}_\eta \\ &= \frac{k}{k+1} \mathbf{W}(k) \epsilon(k) + \frac{1}{k+1} \left( \mathbf{I} - (k+1)(\mathbf{I} + b_0 \mathbf{L})^{-1} \right. \\ &\quad \left. + k \mathbf{W}(k)(\mathbf{I} + b_0 \mathbf{L})^{-1} \right) \mathbf{m}_\eta \\ &= \frac{k}{k+1} \mathbf{W}(k) \epsilon(k) + \frac{1}{k+1} \Gamma(k) \mathbf{m}_\eta \end{aligned} \quad (49)$$

where

$$\begin{aligned} \Gamma(k) &= \mathbf{I} - (k+1)(\mathbf{I} + b_0 \mathbf{L})^{-1} + k \mathbf{W}(k)(\mathbf{I} + b_0 \mathbf{L})^{-1} \\ &= \mathbf{I} - (k+1)(\mathbf{I} + b_0 \mathbf{L})^{-1} + k \left( \mathbf{I} - \frac{b_0}{a+k} \mathbf{L} \right) (\mathbf{I} + b_0 \mathbf{L})^{-1}. \end{aligned}$$

*Setting Up Steps 2 and 3:* Consider the matrix  $\widetilde{\mathbf{W}}(k) := \mathbf{W}(k) - \mathbf{J}$  that represents, in a sense, the perturbation of  $\mathbf{W}(k)$  from its ideal value  $\mathbf{J}$ . For pursuing Steps 2 and 3, we need the eigenvalue decompositions of  $\widetilde{\mathbf{W}}(k)$  and of  $\Gamma(k)$ . We start with  $\widetilde{\mathbf{W}}(k)$ . Recall the graph Laplacian matrix  $\mathbf{L}$  and consider its eigenvalue decomposition:  $\mathbf{L} = \mathbf{Q} \boldsymbol{\Lambda}(\mathbf{L}) \mathbf{Q}^\top$ , with  $\mathbf{q}_i$ 's being the columns of  $\mathbf{Q}$  (the orthonormal eigenvectors); and  $0 = \lambda_1(\mathbf{L}) < \lambda_2(\mathbf{L}) \leq \dots \leq \lambda_N(\mathbf{L})$ ; and the eigenvector that corresponds to the zero eigenvalue is  $\frac{1}{\sqrt{N}} \mathbf{1}$ . Then, clearly,  $\widetilde{\mathbf{W}}(k) = \mathbf{Q}(\mathbf{I} - \alpha_k \boldsymbol{\Lambda}(\mathbf{L}) - \mathbf{e}_1 \mathbf{e}_1^\top) \mathbf{Q}^\top$ , so that we can read the eigenvalues  $\lambda_i(\widetilde{\mathbf{W}}(k))$  of  $\widetilde{\mathbf{W}}(k)$  that correspond to eigenvectors  $\mathbf{q}_i$  as:  $\lambda_i(\widetilde{\mathbf{W}}(k)) = 1 - \alpha_k \lambda_i(\mathbf{L})$ ,  $i \geq 2$ , and  $\lambda_1(\widetilde{\mathbf{W}}(k)) = 0$ . In view of the condition on the step size  $\alpha_k = b_0/(a+k)$ , namely that:  $a \geq b_0 \lambda_N(\mathbf{L})$ , and the fact that we have that  $\widetilde{\mathbf{W}}(k) \mathbf{1} = \mathbf{0}$ , we obtain, respectively,

$$\|\widetilde{\mathbf{W}}(k)\| = 1 - \alpha_k \lambda_2(\mathbf{L}) \in (0, 1) \quad (50)$$

$$\mathbf{J} \widetilde{\mathbf{W}}(k) = \frac{1}{N} \mathbf{1} \left( \mathbf{1}^\top \widetilde{\mathbf{W}}(k) \right) = \widetilde{\mathbf{W}}(k) \mathbf{J} = \mathbf{0} \quad (51)$$

$$\widetilde{\mathbf{W}}(k)(\mathbf{I} - \mathbf{J}) = \widetilde{\mathbf{W}}(k) = \mathbf{W}(k) - \mathbf{J} = (\mathbf{I} - \mathbf{J}) \mathbf{W}(k). \quad (52)$$

We use equalities (50)–(52) further ahead to prove Steps 2 and 3.

We now consider  $\Gamma(k)$  and decompose it via the eigenvalue decomposition:

$$\begin{aligned} \Gamma(k) &= \mathbf{Q} \left( \mathbf{I} - (k+1)(\mathbf{I} + b_0 \boldsymbol{\Lambda}(\mathbf{L}))^{-1} \right. \\ &\quad \left. + k \left( \mathbf{I} - \frac{b_0}{a+k} \boldsymbol{\Lambda}(\mathbf{L}) \right) \right) \mathbf{Q}^\top. \end{aligned}$$

From above, we can read the eigenvalues of  $\Gamma(k)$ ; simple calculations show that the eigenvalue  $\lambda_i(\Gamma(k))$  that corresponds to the eigenvector  $\mathbf{q}_i$  is

$$\lambda_i(\Gamma(k)) = \begin{cases} 0 & \text{if } i = 1 \\ \frac{b_0 \alpha \lambda_i(\mathbf{L})}{1 + b_0 \lambda_i(\mathbf{L})} \frac{1}{k+a} =: \frac{c_{\Gamma,i}}{k+a} & \text{otherwise.} \end{cases}$$

From the discussion on the eigenvalue decomposition of  $\Gamma(k)$ , we have the following properties that we use when proving Steps 2 and 3:

$$\|\Gamma(k)\| \leq c_\Gamma/k, \quad \text{for some } c_\Gamma > 0, \quad \mathbf{1}^\top \Gamma(k) = \mathbf{0}, \mathbf{J} \Gamma(k) = \mathbf{0}. \quad (53)$$

When proving Step 3, we invoke the following deterministic variant of a result due to Robbins and Siegmund [31, Lemma 11, Ch. 2.2].

*Lemma 6 [31]:* Let  $\{u(k)\}$ ,  $\{\rho(k)\}$ , and  $\{\kappa(k)\}$  be non-negative deterministic (scalar) sequences. Further, suppose that

$$u(k+1) \leq u(k) - \rho(k) + \kappa(k), \quad k = 1, 2, \dots$$

Suppose that  $\sum_{k=1}^\infty \kappa(k) < \infty$ ; then, 1)  $\sum_{k=1}^\infty \rho(k) < \infty$  and 2)  $\lim_{k \rightarrow \infty} u(k) = u^*$  exists.

*Proof of Step 2:* Consider the recursive equation for the error vector  $\epsilon(k)$ . Multiplying (49) from the left by  $\mathbf{1}^\top$ , using (53), and  $\mathbf{1}^\top \mathbf{W}(k) = \mathbf{1}^\top$  (doubly stochastic matrix), we get

$$\begin{aligned} \mathbf{1}^\top \epsilon(k+1) &= \frac{k}{k+1} \mathbf{1}^\top \epsilon(k) = \frac{k}{k+1} \frac{k-1}{k} \mathbf{1}^\top \epsilon(k-1) \\ &= \dots = \frac{1}{k+1} \mathbf{1}^\top \epsilon(1) \end{aligned}$$

which proves (46).

*Proof of Step 3:* Denote by  $b := b_0 \lambda_2(\mathbf{L})$ . Multiplying (49) from the left by  $(\mathbf{I} - \mathbf{J})$  and using (52) and (53), we get

$$\begin{aligned} (\mathbf{I} - \mathbf{J}) \epsilon(k+1) &= \frac{k}{k+1} (\mathbf{I} - \mathbf{J}) \mathbf{W}(k) \epsilon(k) + \frac{1}{k+1} (\mathbf{I} - \mathbf{J}) \Gamma(k) \mathbf{m}_\eta \\ &= \frac{k}{k+1} \widetilde{\mathbf{W}}(k) (\mathbf{I} - \mathbf{J}) \epsilon(k) + \frac{1}{k+1} \Gamma(k) \mathbf{m}_\eta. \end{aligned}$$

Now, by the subadditivity and submultiplicative properties of norms and (50) and (53):

$$\begin{aligned} \|(\mathbf{I} - \mathbf{J}) \epsilon(k+1)\| &\leq \frac{k}{k+1} \|\widetilde{\mathbf{W}}(k)\| \|(\mathbf{I} - \mathbf{J}) \epsilon(k)\| + \frac{1}{k+1} \|\Gamma(k)\| \|\mathbf{m}_\eta\| \\ &\leq \left( 1 - \frac{b_0 \lambda_2(\mathbf{L})}{a+k} \right) \|(\mathbf{I} - \mathbf{J}) \epsilon(k)\| + \frac{c_\Gamma \|\mathbf{m}_\eta\|}{k^2} \\ &= \left( 1 - \frac{b}{a+k} \right) \|(\mathbf{I} - \mathbf{J}) \epsilon(k)\| + \frac{c_\epsilon}{k^2} \\ &= \|(\mathbf{I} - \mathbf{J}) \epsilon(k)\| - \frac{b}{a+k} \|(\mathbf{I} - \mathbf{J}) \epsilon(k)\| + \frac{c_\epsilon}{k^2} \end{aligned} \quad (54)$$

which, invoking Lemma 6, proves (31): let  $u(k) = \|(\mathbf{I} - \mathbf{J}) \epsilon(k)\|$ ,  $\rho(k) = \frac{b}{a+k} \|(\mathbf{I} - \mathbf{J}) \epsilon(k)\|$ ,  $\kappa(k) = \frac{c_\epsilon}{k^2}$ , get  $\sum \rho(k) < \infty$ ; thus,  $\liminf_k \|(\mathbf{I} - \mathbf{J}) \epsilon(k)\| = \liminf_k u(k) = 0$ ; as  $\lim_k u(k)$  exists,  $\lim_k u(k) = 0$ .

*Proof of (32):* We now lower bound each entry  $[\mu_\infty]_i$  of the calculated limit vector  $\mu_\infty = (\mathbf{I} + b_0 \mathbf{L})^{-1} \mathbf{m}_\eta$ . To this end, we decompose  $\mu_\infty$  into an “ideal” component and a “perturbation” component. First, note

$$(\mathbf{I} + b_0 \mathbf{L})^{-1} = \mathbf{Q} (\mathbf{I} + b_0 \Lambda(\mathbf{L}))^{-1} \mathbf{Q}^\top = \mathbf{J} + \mathbf{Q} \Lambda' \mathbf{Q}^\top,$$

with  $\Lambda' = \text{Diag}(0, (1 + b_0 \lambda_2(\mathbf{L}))^{-1}, \dots, (1 + b_0 \lambda_N(\mathbf{L}))^{-1})$ . Hence,  $\mu_\infty = \mathbf{J} \mathbf{m}_\eta + \mathbf{Q} \Lambda' \mathbf{Q}^\top \mathbf{m}_\eta$ , where  $\mathbf{J} \mathbf{m}_\eta$  is the “ideal” value of the limit of the mean, as if the network was fully connected at any time  $k$ , i.e., as if  $\mathbf{W}(k) \equiv \mathbf{J}$ , for all  $k$ ; and  $\mathbf{Q} \Lambda' \mathbf{Q}^\top \mathbf{m}_\eta$  is the “perturbation value.” We further calculate this “ideal” value (we recall  $\mathbf{m}_\eta = \frac{1}{2} \text{Diag}(\mathbf{S}_\zeta^{-1} \mathbf{m})$ ):

$$\begin{aligned} \mathbf{J} \mathbf{m}_\eta &= \frac{1}{2N} \mathbf{1} \left( \mathbf{1}^\top \text{Diag}(\mathbf{S}_\zeta^{-1} \mathbf{m}) \mathbf{m} \right) \\ &= \frac{1}{2N} \mathbf{m}^\top \mathbf{S}_\zeta^{-1} \mathbf{m} \mathbf{1} = \frac{\gamma_s}{2N} \mathbf{1}. \end{aligned}$$

Hence, the entry  $[\mu_\infty]_i = \frac{\gamma_s}{2N} + [\mathbf{Q} \Lambda' \mathbf{Q}^\top \mathbf{m}_\eta]_i$ . To complete the bound, we bound the modulus of the perturbation value:  $|\mathbf{Q} \Lambda' \mathbf{Q}^\top \mathbf{m}_\eta|_i \leq \|\mathbf{Q} \Lambda' \mathbf{Q}^\top\| \|\mathbf{m}_\eta\| = \frac{1}{1+b_0 \lambda_2(\mathbf{L})} \|\mathbf{m}_\eta\|$ . Finally, the proof of (32) follows by  $[\mu_\infty]_i \geq \frac{\gamma_s}{2N} - \frac{1}{1+b_0 \lambda_2(\mathbf{L})} \|\mathbf{m}_\eta\| = \frac{\gamma_s}{2N} \{1 - \frac{N}{1+b_0 \lambda_2(\mathbf{L})} \frac{2\|\mathbf{m}_\eta\|}{\gamma_s}\}$ .

*Proof of the Exact Covariance Limit (33):* We now show that the covariance matrix of the detection statistic vector  $k\Sigma(k) := k \text{Cov}(\mathbf{x}(k)) \rightarrow \text{Vec}^{-1}\{(\mathbf{I} + b_0 \mathbf{L}_2)^{-1} (\text{Vec}(\mathbf{S}_\eta) + b_0^2 \text{Vec}(\mathbf{S}_v))\}$ , where  $\text{Vec}^{-1}$  “unvecs” or recovers the matrix from the vector and where  $\mathbf{L}_2 = \mathbf{L} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{L}$ . Interestingly, we can prove the latter by essentially mimicking the proof of the limit of the mean (31). In summary, we derive the recursive equation for  $\sigma_2(k) := k \text{Vec}(\Sigma(k))$ , study properties of  $\mathbf{L}_2$ , show that the recursion for  $\sigma_2(k)$  is essentially the same type of equation as the one for the mean vector in (48) (with some minor differences) construct the error vector w.r.t. the assumed limit and prove it converges to zero. Note that, with this proof, we work with  $N^2 \times 1$  vectors and  $N^2 \times N^2$  matrices; we do not make a specific mention of dimensions, e.g.,  $N^2 \times N^2$  identity matrix is simply denoted by  $\mathbf{I}$ .

*Recursion for  $\sigma_2(k)$ :* From the update of the decision vector (25), taking the covariance yields

$$\begin{aligned} \Sigma(k+1) &= \left( \frac{k}{k+1} \right)^2 \mathbf{W}(k) \Sigma(k) \mathbf{W}(k) \\ &\quad + \left( \frac{k}{k+1} \right)^2 \alpha_k^2 \mathbf{S}_v + \left( \frac{1}{k+1} \right)^2 \mathbf{S}_\eta. \end{aligned}$$

Multiplying the above equation by  $(k+1)$  and taking the  $\text{Vec}$ :

$$\begin{aligned} \sigma_2(k+1) &= \frac{k}{k+1} (\mathbf{W}(k) \otimes \mathbf{W}(k)) \sigma_2(k) \\ &\quad + \frac{b_0^2 k^2}{(k+1)(a+k)^2} \text{Vec}(\mathbf{S}_v) + \frac{1}{k+1} \text{Vec}(\mathbf{S}_\eta) \end{aligned}$$

where the matrix  $\mathbf{W}(k) \otimes \mathbf{W}(k) = (\mathbf{I} - \alpha_k \mathbf{L}) \otimes (\mathbf{I} - \alpha_k \mathbf{L}) = \mathbf{I} - \alpha_k \mathbf{L}_2 + \frac{b_0^2}{(a+k)^2} (\mathbf{L} \otimes \mathbf{L})$ .

*Properties of the Matrix  $\mathbf{L}_2$ :* Note that  $\mathbf{L}_2$  is a valid graph Laplacian matrix of a certain connected graph on  $N^2$  nodes: its rows sum to 0, the off-diagonal entries are either zero or one,

and the diagonal entries are positive. Also, the eigenvalues of  $\mathbf{L}_2$  equal  $\lambda_i(\mathbf{L}) + \lambda_j(\mathbf{L})$ ,  $i, j = 1, \dots, N$ , and so the second smallest eigenvalue  $\lambda_2(\mathbf{L}_2) = \lambda_2(\mathbf{L}) > 0$ . The matrix  $\mathbf{L}_2$  is symmetric. It will be useful to define the averaging matrix on our virtual graph of  $N^2$  nodes:  $\mathbf{W}_2(k) := \mathbf{I} - \alpha_k \mathbf{L}_2$ .

*Identification of  $\sigma_2(k)$  with  $\mu(k)$ :* Simple algebraic manipulations show

$$\begin{aligned} \sigma_2(k+1) &= \frac{k}{k+1} \mathbf{W}_2(k) \sigma_2(k) + \frac{1}{k+1} (b_0^2 \text{Vec}(\mathbf{S}_v) + \text{Vec}(\mathbf{S}_\eta)) \\ &\quad - \frac{b_0^2 a (2k+a)}{(k+1)(a+k)^2} \text{Vec}(\mathbf{S}_v) + \frac{k}{k+1} \alpha_k^2 (\mathbf{L} \otimes \mathbf{L}) \sigma_2(k). \end{aligned} \quad (55)$$

We note that (55) is very similar to the recursive equation for  $\mu(k)$  in (48). The difference is in the terms (56). As we will show, these do not affect the limit of  $\sigma_2(k)$ , as they are both  $O(1/k^2)$ . For the first term, this is clear; for the second, it is true because, from our *independently proved* upper bound (uniform over  $i = 1, \dots, N$ ) on  $\limsup_{k \rightarrow \infty} k \sigma_i^2(k)$  in (34), we can see that  $\sigma_2(k) \leq M < \infty$ , for all  $k \geq k_\sigma$ , and for sufficiently large  $k_\sigma$ ; thus, the second term in (56) is also  $O(1/k^2)$ . Thus, we make the following identification:  $\sigma_2(k) \equiv \mu(k)$ ;  $\mathbf{m}_\eta \equiv b_0^2 \text{Vec}(\mathbf{S}_v) + \text{Vec}(\mathbf{S}_\eta)$ ;  $\mathbf{W}(k) \equiv \mathbf{W}_2(k)$ ; and we show that  $\lim_{k \rightarrow \infty} \sigma_2(k) = (\mathbf{I} + b_0 \mathbf{L}_2)^{-1} (\text{Vec}(\mathbf{S}_\eta) + b_0^2 \text{Vec}(\mathbf{S}_v))$ . Like with the proof of (31), we construct the error w.r.t. the assumed limit,  $\epsilon_2(k) := \sigma_2(k) - (\mathbf{I} + b_0 \mathbf{L}_2)^{-1} (\text{Vec}(\mathbf{S}_\eta) + b_0^2 \text{Vec}(\mathbf{S}_v))$ , split it into the consensus subspace error component and the orthogonal subspace error component, and prove that each component converges to zero, which implies  $\|\epsilon_2(k)\| \rightarrow 0$ . The recursion for  $\epsilon_2(k)$  is

$$\begin{aligned} \epsilon_2(k+1) &= \frac{k}{k+1} \mathbf{W}_2(k) \epsilon_2(k) + \frac{1}{k+1} \Gamma_2(k) (\text{Vec}(\mathbf{S}_\eta) + b_0^2 \text{Vec}(\mathbf{S}_v)) \\ &\quad - \frac{b_0^2 a (2k+a)}{(k+1)(a+k)^2} \text{Vec}(\mathbf{S}_v) + \frac{k}{k+1} \alpha_k^2 (\mathbf{L} \otimes \mathbf{L}) \sigma_2(k), \end{aligned}$$

where  $\Gamma_2(k) \equiv \Gamma(k)$  and equals

$$\begin{aligned} \Gamma_2(k) &= \mathbf{I} - (k+1)(\mathbf{I} + b_0 \mathbf{L}_2)^{-1} + k \mathbf{W}_2(k) (\mathbf{I} + b_0 \mathbf{L}_2)^{-1} \\ &= \mathbf{I} - (k+1)(\mathbf{I} + b_0 \mathbf{L}_2)^{-1} \\ &\quad + k \left( \mathbf{I} - \frac{b_0}{a+k} \mathbf{L}_2 \right) (\mathbf{I} + b_0 \mathbf{L}_2)^{-1}. \end{aligned}$$

*Consensus Subspace Error Component:* It can be shown that the update equation is

$$\begin{aligned} &\mathbf{1}^\top \epsilon_2(k+1) \\ &= \frac{k}{k+1} \mathbf{1}^\top \epsilon_2(k) + \mathbf{1}^\top \left( -\frac{b_0^2 a (2k+a)}{(k+1)(a+k)^2} \text{Vec}(\mathbf{S}_v) \right. \\ &\quad \left. + \frac{k}{k+1} \alpha_k^2 (\mathbf{L} \otimes \mathbf{L}) \sigma_2(k) \right). \end{aligned}$$

Taking the modulus and noting that  $|\frac{b_0^2 a (2k+a)}{(k+1)(a+k)^2} \text{Vec}(\mathbf{S}_v) + \frac{k}{k+1} \alpha_k^2 (\mathbf{L} \otimes \mathbf{L}) \sigma_2(k)| \leq \frac{c_p}{k^2}$ , for some  $c_p > 0$ , and so:  $|\mathbf{1}^\top \epsilon_2(k+1)| \leq \frac{k}{k+1} |\mathbf{1}^\top \epsilon_2(k)| + \frac{c_p}{k^2}$ , and so  $|\mathbf{1}^\top \epsilon_2(k+1)| \rightarrow 0$  applying Lemma 6.

*Orthogonal Subspace Error Component:* Denote by  $\widetilde{\mathbf{W}}_2(k) := \mathbf{W}_2(k) - \mathbf{J}$ . The recursion is

$$\begin{aligned} (\mathbf{I} - \mathbf{J})\boldsymbol{\epsilon}_2(k+1) &= \frac{k}{k+1}\widetilde{\mathbf{W}}_2(k)(\mathbf{I} - \mathbf{J})\boldsymbol{\epsilon}_2(k) \\ &+ \frac{1}{k+1}\boldsymbol{\Gamma}_2(k) (b_0^2 \text{Vec}(\mathbf{S}_v) + \text{Vec}(\mathbf{S}_\eta)) \\ &+ (\mathbf{I} - \mathbf{J}) \left( -\frac{b_0^2 a (2k+a)}{(k+1)(a+k)^2} \text{Vec}(\mathbf{S}_v) \right. \\ &\left. + \frac{k}{k+1} \alpha_k^2 (\mathbf{L} \otimes \mathbf{L}) \boldsymbol{\sigma}_2(k) \right). \end{aligned}$$

Taking the norm:  $\|(\mathbf{I} - \mathbf{J})\boldsymbol{\epsilon}_2(k+1)\| \leq \frac{k}{k+1} \|\widetilde{\mathbf{W}}_2(k)\| \|(\mathbf{I} - \mathbf{J})\boldsymbol{\epsilon}_2(k)\| + \frac{c'_p}{k^2}$ , and the proof proceeds by again applying Lemma 6.

*Proof of (34):* We now show that the variance of agent  $i$ 's detection statistic  $\sigma_i^2(k) = \text{Var}(x_i(k))$  decays to zero as  $1/k$  and then we upper bound  $\limsup_{k \rightarrow \infty} k \sigma_i^2(k)$ . We work directly with the solution form of  $\mathbf{x}(k)$  in (26). To calculate  $x_i(k)$ , we multiply  $\mathbf{x}(k)$  by  $\mathbf{e}_i^\top$  on the left to obtain

$$x_i(k) = \frac{1}{k} \sum_{j=1}^k \mathbf{e}_i^\top \boldsymbol{\Phi}(k, j) \boldsymbol{\eta}(j) + \frac{1}{k} \sum_{j=1}^{k-1} (j \alpha_j) \mathbf{e}_i^\top \boldsymbol{\Phi}(k, j+1) \mathbf{v}(j). \quad (57)$$

We view the first sum above as the “signal” part and the second sum as the “noise” part. Indeed, the “signal” part involves the innovations (log-likelihood ratios), while the “noise” part involves only communication noise. We proceed as follows. In Step 1, we calculate the variance of  $x_i(k)$  from (57). In Step 2, we upper bound the “noise” part of the variance. In Step 3, we upper bound the “signal” part of the variance. Finally, in Step 4, we combine the two and finalize the bound.

*Step 1. Calculation of the Variance of  $x_i(k)$ :* Due to the mutual independence of: 1) the innovations  $\boldsymbol{\eta}(j)$  and  $\boldsymbol{\eta}(s)$  for  $j \neq s$ ; 2) the communication noises  $\mathbf{v}(j)$  and  $\mathbf{v}(s)$  for  $j \neq s$ ; and 3)  $\boldsymbol{\eta}(j)$  and  $\mathbf{v}(s)$  for all  $s, j$ , the variance of  $x_i(k)$  is the sum of the variances of the individual summands in (57):

$$\begin{aligned} \sigma_i^2(k) &= \frac{1}{k^2} \sum_{j=1}^k \text{Var}(\mathbf{e}_i^\top \boldsymbol{\Phi}(k, j) \boldsymbol{\eta}(j)) \\ &+ \frac{1}{k^2} \sum_{j=1}^{k-1} (\alpha_j j)^2 \text{Var}(\mathbf{e}_i^\top \boldsymbol{\Phi}(k, j+1) \mathbf{v}(j)). \quad (58) \end{aligned}$$

The first sum in (58) corresponds to the “signal” part, and the second corresponds to the “noise” part.

*Step 2. Variance of the “Noise” Part:* We have from (58)

$$\begin{aligned} &\frac{1}{k^2} \sum_{j=1}^{k-1} (\alpha_j j)^2 \text{Var}(\mathbf{e}_i^\top \boldsymbol{\Phi}(k, j+1) \mathbf{v}(j)) \\ &= \frac{1}{k^2} \sum_{j=1}^{k-1} (\alpha_j j)^2 \mathbf{e}_i^\top \boldsymbol{\Phi}(k, j+1) \mathbf{S}_v \boldsymbol{\Phi}(k, j+1)^\top \mathbf{e}_i, \quad (59) \end{aligned}$$

where we recall  $\mathbf{S}_v := \text{Cov}(\mathbf{v}(k)) \succ \mathbf{0}$ . The term

$$\begin{aligned} &\mathbf{e}_i^\top \boldsymbol{\Phi}(k, j+1) \mathbf{S}_v \boldsymbol{\Phi}(k, j+1)^\top \mathbf{e}_i \\ &\leq \|\mathbf{e}_i\| \|\boldsymbol{\Phi}(k, j+1)\| \|\mathbf{S}_v\| \|\boldsymbol{\Phi}(k, j+1)\| \|\mathbf{e}_i\| = \|\mathbf{S}_v\|, \end{aligned}$$

by the submultiplicative property of norms and because  $\|\boldsymbol{\Phi}(k, j+1)\| = 1$  (doubly stochastic matrix), and  $\|\mathbf{e}_i\| = 1$ . Thus, the “noise” part is bounded as

$$\begin{aligned} &\frac{1}{k^2} \sum_{j=1}^{k-1} (\alpha_j j)^2 \text{Var}(\mathbf{e}_i^\top \boldsymbol{\Phi}(k, j+1) \mathbf{v}(j)) \leq \|\mathbf{S}_v\| \frac{1}{k^2} \sum_{j=1}^{k-1} (j \alpha_j)^2 \\ &= \|\mathbf{S}_v\| \frac{1}{k^2} \sum_{j=1}^{k-1} \left( j \frac{b_0}{a+j} \right)^2 \leq \|\mathbf{S}_v\| \frac{1}{k^2} (k-1) b_0^2 \leq \frac{b_0^2 \|\mathbf{S}_v\|}{k} \quad (60) \end{aligned}$$

which completes the upper bound for the “noise” part.

*Step 3. Variance of the “Signal” Part:* We now upper bound the “signal” part of the variance:

$$\frac{1}{k^2} \sum_{j=1}^k \text{Var}(\mathbf{e}_i^\top \boldsymbol{\Phi}(k, j) \boldsymbol{\eta}(j)) = \frac{1}{k^2} \sum_{j=1}^k \mathbf{e}_i^\top \boldsymbol{\Phi}(k, j) \mathbf{S}_\eta \boldsymbol{\Phi}(k, j)^\top \mathbf{e}_i \quad (61)$$

where we recall  $\mathbf{S}_\eta = \text{Cov}(\boldsymbol{\eta}(k))$ . If all matrices  $\mathbf{W}(j)$  were equal to  $\mathbf{J}$  (fully connected network), and consequently,  $\boldsymbol{\Phi}(k, j) = \mathbf{J}$ , then (61) would equal (recall  $\mathbf{S}_\eta = \text{Diag}(\mathbf{S}_\zeta^{-1} \mathbf{m}) \mathbf{S}_\zeta \text{Diag}(\mathbf{S}_\zeta^{-1} \mathbf{m})$ )

$$\begin{aligned} &\frac{1}{k^2 N^2} k \mathbf{1}^\top \text{Diag}(\mathbf{S}_\zeta^{-1} \mathbf{m}) \mathbf{S}_\zeta \text{Diag}(\mathbf{S}_\zeta^{-1} \mathbf{m}) \mathbf{1} \\ &= \frac{1}{k N^2} \mathbf{m}^\top \mathbf{S}_\eta^{-1} \mathbf{m} = \frac{1}{k} \frac{\gamma_s}{N^2}. \quad (62) \end{aligned}$$

This is the variance of the agent  $i$ 's detection statistic if the network was fully connected at all times, i.e.,  $\mathbf{W}(k) \equiv \mathbf{J}$ , for all  $k$ . To further assess the “perturbation” of the non-ideal variance (61) w.r.t. the ideal variance (62), we split the matrix  $\boldsymbol{\Phi}(k, j) = \mathbf{J} + (\boldsymbol{\Phi}(k, j) - \mathbf{J}) = \mathbf{J} + \tilde{\boldsymbol{\Phi}}(k, j)$ . Recall  $\widetilde{\mathbf{W}}(k) = \mathbf{W}(k) - \mathbf{J}$ . It is easy to show, by (51), that:  $\tilde{\boldsymbol{\Phi}}(k, j) = \widetilde{\mathbf{W}}(k-1) \widetilde{\mathbf{W}}(k-2) \dots \widetilde{\mathbf{W}}(j)$ . Using the splitting and (62), the “signal” part variance (61) becomes

$$\begin{aligned} &\frac{1}{k^2} \sum_{j=1}^k \text{Var}(\mathbf{e}_i^\top \boldsymbol{\Phi}(k, j) \boldsymbol{\eta}(j)) \\ &= \frac{1}{N^2 k} \gamma_s + \frac{1}{k^2} \sum_{j=1}^k \mathbf{e}_i^\top \mathbf{J} \mathbf{S}_\eta \tilde{\boldsymbol{\Phi}}(k, j) \mathbf{e}_i \\ &+ \frac{1}{k^2} \sum_{j=1}^k \left( \mathbf{e}_i^\top \mathbf{J} \mathbf{S}_\eta \tilde{\boldsymbol{\Phi}}(k, j)^\top \mathbf{e}_i \right)^\top \\ &+ \frac{1}{k^2} \sum_{j=1}^k \mathbf{e}_i^\top \tilde{\boldsymbol{\Phi}}(k, j) \mathbf{S}_\eta \tilde{\boldsymbol{\Phi}}(k, j)^\top \mathbf{e}_i. \quad (63) \end{aligned}$$

We bound the “signal” part variance above by the submultiplicative property of norms and using  $\|\mathbf{e}_i\| = \|\mathbf{J}\| = 1$  and  $\|\tilde{\boldsymbol{\Phi}}(k, j)\| \leq \|\widetilde{\mathbf{W}}(k-1)\| \dots \|\widetilde{\mathbf{W}}(j)\| \leq 1$ , by (50):

$$\frac{1}{k^2} \sum_{j=1}^k \text{Var}(\mathbf{e}_i^\top \boldsymbol{\Phi}(k, j) \boldsymbol{\eta}(j)) \leq \frac{1}{k} \frac{\gamma_s}{N^2} + \frac{1}{k} \frac{3 \|\mathbf{S}_\eta\|}{k} \sum_{j=1}^k \|\tilde{\boldsymbol{\Phi}}(k, j)\|. \quad (64)$$

Note that the second summand in (64) is the additional ‘‘perturbation’’ when compared with the ideal fully connected network case.

*Step 4. Combining Bounds on the ‘‘Signal’’ Part and the ‘‘Noise’’ Part:* By (64) and (60), we get

$$\sigma_i^2(k) \leq \frac{b_0^2 \|\mathbf{S}_v\|}{k} + \frac{1}{k} \frac{\gamma_s}{N^2} + \frac{1}{k} 3 \|\mathbf{S}_\eta\| \left(1 - \frac{b_0 \lambda_2(\mathbf{L})}{a+k}\right)^{-1} \times \left\{ \frac{1}{k} \sum_{j=1}^k \|\tilde{\Phi}(k+1, j)\| \right\}. \quad (65)$$

We can see that  $\sigma_i^2(k) = O(1/k)$ , because the perturbation term:  $\mathcal{P}(k) = \frac{1}{k} \sum_{j=1}^k \|\tilde{\Phi}(k+1, j)\| \leq 1$ . To get a tighter bound, however, we now find the exact limit  $\lim_{k \rightarrow \infty} \mathcal{P}(k) = \frac{1}{1+b_0 \lambda_2(\mathbf{L})} = \frac{1}{1+b}$ . To do this, we rewrite  $\mathcal{P}(k)$  in a recursive form:

$$\mathcal{P}(k+1) = \left(1 - \frac{b}{a+k+1}\right) \left(\frac{k}{k+1} \mathcal{P}(k) + \frac{1}{k+1}\right),$$

$$\mathcal{P}(1) = 1 - \frac{b}{a+1} > 0. \quad (66)$$

By subtracting the assumed limit  $1/(1+b)$  from both sides of (66), algebraic manipulations show

$$\left(\mathcal{P}(k+1) - \frac{1}{1+b}\right) = \frac{k}{k+1} \left(1 - \frac{b}{a+k+1}\right) \times \left(\mathcal{P}(k) - \frac{1}{1+b}\right) + \frac{ab-b^2}{(a+k+1)(k+1)(b+1)}.$$

After taking the modulus on both sides and using the submultiplicative property of modulus, we get

$$\left|\mathcal{P}(k+1) - \frac{1}{1+b}\right| \leq \frac{k}{k+1} \left(1 - \frac{b}{a+k+1}\right) \times \left|\mathcal{P}(k) - \frac{1}{1+b}\right| + \frac{|ab-b^2|}{k^2(b+1)}$$

which parallels (54), for which we already showed that  $\|(\mathbf{I} - \mathbf{J})\epsilon(k)\| \rightarrow 0$ . Thus,  $|\mathcal{P}(k+1) - \frac{1}{1+b}| \rightarrow 0$ . We now revisit (65) to show (34):

$$\limsup_{k \rightarrow \infty} k \sigma_i^2(k) \leq \frac{\gamma_s}{N^2} + 3 \|\mathbf{S}_\eta\| \frac{1}{1+b_0 \lambda_2(\mathbf{L})} + b_0^2 \|\mathbf{S}_v\|$$

$$= \frac{\gamma_s}{N^2} \left\{ 1 + \frac{3N}{1+b_0 \lambda_2(\mathbf{L})} \frac{\|\mathbf{S}_\eta\|}{\frac{\gamma_s}{N}} + b_0^2 N \frac{\|\mathbf{S}_v\|}{\frac{\gamma_s}{N}} \right\}.$$

### B. Proof of Theorem 5

We now prove the upper bound on the exponential decay rate of the error probability for identical agents, when the weight  $\alpha_k$  is replaced by  $\beta_k = \frac{b_0}{a+k^\tau}$ ,  $\tau \geq 0$ . To find the upper bound on the rate, due to the Gaussianity of  $x_i(k)$  under either hypothesis, we only need to find an upper bound on the  $\limsup_{k \rightarrow \infty} \mathbb{E}[x_i(k) | H_1] = \limsup_{k \rightarrow \infty} \mu_i(k)$  (Step 1), and a lower bound on the variance  $\liminf_{k \rightarrow \infty} k \sigma_i^2(k)$  (Step 2). For Step 2, we have to distinguish three cases:  $\tau < 1$ ,  $\tau > 1$ , and  $\tau = 1$ . Before proceeding with Steps 1 and 2, we note that, to avoid cumbersome notation, we still use the same letter

$\mathbf{W}(k) = \mathbf{I} - \beta_k \mathbf{L}$  for the averaging matrix, although the weight changed; accordingly, we also use the symbols  $\tilde{\mathbf{W}}(k)$ ,  $\tilde{\Phi}(k, j)$ , and  $\tilde{\Phi}(k, j)$  unchanged. We recall that, for identical agents, the expressions for the mean and variance of the innovation vector  $\eta(k)$  simplify to  $\mathbf{m}_\eta = (1/2)\gamma_{s,i} \mathbf{1}$  and  $\mathbf{S}_\eta = \gamma_{s,i} \mathbf{I}$ .

*Step 1. The Mean:* Taking expectation in the solution equation (26) for the decision vector  $\mathbf{x}(k)$ :  $\mu(k) = \frac{1}{k} \sum_{j=1}^k \tilde{\Phi}(k, j) \frac{\gamma_{s,i}}{2} \mathbf{1} = \frac{\gamma_{s,i}}{2} \mathbf{1}$ ,  $\forall k$ . Hence, the mean  $\mu_i(k)$  remains time invariant under identical agents and equal to  $\frac{\gamma_{s,i}}{2}$ , irrespective of the choice of weights; clearly,

$$\lim_{k \rightarrow \infty} \mu_i(k) = \frac{\gamma_{s,i}}{2}. \quad (67)$$

*Step 2. The Variance:* We first carry out the analysis for arbitrary  $\tau \geq 0$ ; from a certain point on, we will need to distinguish between three cases:  $\tau < 1$ ,  $\tau > 1$ , and  $\tau = 1$ . We start with the exact expression for  $k \sigma_i^2(k)$ . We get the latter by combining (59), (63), multiplying these by  $k$ , and noting that the second summand in (63) is zero, because  $\mathbf{S}_\eta = \gamma_{s,i} \mathbf{I}$ , and so  $\mathbf{J} \mathbf{S}_\eta \tilde{\Phi}(k, j) = \gamma_{s,i} \mathbf{J} \tilde{\Phi}(k, j) = \mathbf{0}$ , by property (51). (Note that the property holds even though we use  $\beta_k$  (arbitrary  $\tau$ ) and not  $\alpha_k$  ( $\tau = 1$ ).) We have:

$$k \sigma_i^2(k) = \frac{\gamma_{s,i}}{N} + \frac{\gamma_{s,i}}{k} \sum_{j=1}^k \mathbf{e}_i^\top \tilde{\Phi}(k, j) \tilde{\Phi}(k, j)^\top \mathbf{e}_i$$

$$+ \frac{1}{k} \sum_{j=1}^{k-1} (\beta_j)^2 \mathbf{e}_i^\top \tilde{\Phi}(k, j+1) \mathbf{S}_v \tilde{\Phi}(k, j+1)^\top \mathbf{e}_i.$$

We next bound  $k \sigma_i^2(k)$  from above, using the following simple relations:  $\mathbf{e}_i^\top \tilde{\Phi}(k, j) \tilde{\Phi}(k, j)^\top \mathbf{e}_i \geq \lambda_2(\tilde{\Phi}(k, j) \tilde{\Phi}(k, j)^\top) \|(\mathbf{I} - \mathbf{J})\mathbf{e}_i\|^2 = \lambda_2(\tilde{\Phi}(k, j) \tilde{\Phi}(k, j)^\top) (1 - 1/N)$ ,  $\|\tilde{\Phi}(k, j)^\top \mathbf{e}_i\|^2 \geq \frac{1}{N^2}$ , where the latter inequality holds true because  $\tilde{\Phi}(k, j)^\top \mathbf{e}_i$  is a vector with nonnegative entries that sum to one, and so at least one entry is greater than or equal to  $1/N$ . The upper bound on  $k \sigma_i^2(k)$  is as follows:

$$k \sigma_i^2(k) \geq \frac{1}{N} \gamma_{s,i} + \frac{\gamma_{s,i}}{k} (1 - 1/N) \sum_{j=1}^k \lambda_2 \left( \tilde{\Phi}(k, j) \tilde{\Phi}(k, j)^\top \right)$$

$$+ \frac{1}{k} \sum_{j=1}^{k-1} (j \beta_j)^2 \lambda_1(\mathbf{S}_v) \|\tilde{\Phi}(k, j+1)^\top \mathbf{e}_i\|^2$$

$$\geq \frac{1}{N} \gamma_{s,i} + \frac{\gamma_{s,i}}{k} (1 - 1/N) \sum_{j=1}^k \left( \prod_{s=j}^{k-1} (1 - \lambda_N(\mathbf{L}) \beta_s)^2 \right)$$

$$+ \frac{1}{k} \frac{1}{N^2} \sum_{j=1}^{k-1} (j \beta_j)^2 \lambda_1(\mathbf{S}_v)$$

$$= \frac{\gamma_{s,i}}{N} + (1 - 1/N) \gamma_{s,i} (1 - \beta_k \lambda_N(\mathbf{L}))^{-2} \mathcal{P}_\beta(k)$$

$$+ \frac{\lambda_1(\mathbf{S}_v)}{N^2} \chi_\beta(k) \quad (68)$$

where we introduced  $\mathcal{P}_\beta(k) := \frac{1}{k} \sum_{j=1}^k \left( \prod_{s=j}^{k-1} (1 - \lambda_N(\mathbf{L}) \beta_s)^2 \right)$ ,  $\chi_\beta(k) := \frac{1}{k} \sum_{j=1}^{k-1} (j \beta_j)^2$ .

*Distinguishing Cases*  $\tau < 1$ ,  $\tau > 1$ , and  $\tau = 1$ : We now distinguish the three cases to prove

$$\lim_{k \rightarrow \infty} \mathcal{P}_\beta(k) \begin{cases} = 0, & \text{if } \tau < 1 \\ = 1, & \text{if } \tau > 1 \\ \geq \frac{1}{1+2b_0\lambda_N(\mathbf{L})}, & \text{if } \tau = 1 \end{cases}$$

$$\lim_{k \rightarrow \infty} \chi_\beta(k) \begin{cases} = +\infty, & \text{if } \tau < 1 \\ = 0, & \text{if } \tau > 1 \\ = b_0^2, & \text{if } \tau = 1. \end{cases} \quad (69)$$

The proofs for  $\chi_\beta(k)$  are trivial; we thus proceed with the proofs for  $\mathcal{P}_\beta(k)$ .

*Case*  $\tau < 1$ : We note that  $\mathcal{P}_\beta(k)$  updates according to the following recursion:

$$\mathcal{P}_\beta(k+1) = \left(1 - \frac{b'}{a + (k+1)^\tau}\right)^2 \left(\frac{k}{k+1}\mathcal{P}_\beta(k) + \frac{1}{k+1}\right)$$

$$\mathcal{P}_\beta(1) = \left(1 - \frac{b'}{a+1}\right)^2 > 0 \quad (70)$$

where  $b' := b_0\lambda_N(\mathbf{L})$ . By (70), for sufficiently large  $k_0$ , and for all  $k \geq k_0$ , we have that

$$\mathcal{P}_\beta(k+1) \leq \left(1 - \frac{b_\beta}{(k+1)^\tau}\right) \mathcal{P}_\beta(k) + \frac{1}{k+1},$$

for appropriately chosen  $b_\beta > 0$ . Now, applying [32, Lemma 4], we get that  $\lim_{k \rightarrow \infty} \mathcal{P}_\beta(k) = 0$ .

*Case*  $\tau > 1$ : By (70), and because  $\mathcal{P}_\beta(k) \leq 1, \forall k$ ,  $\mathcal{P}_\beta(k+1)$  can be bounded from below as

$$\begin{aligned} \mathcal{P}_\beta(k+1) &\geq \frac{k}{k+1} \left(1 - \frac{2b'}{a + (k+1)^\tau}\right) \mathcal{P}_\beta(k) \\ &\quad + \left(1 - \frac{2b'}{a + (k+1)^\tau}\right) \frac{1}{k+1} \\ &= \frac{k}{k+1} \mathcal{P}_\beta(k) - \frac{2b'k}{(k+1)(a + (k+1)^\tau)} \mathcal{P}_\beta(k) \\ &\quad + \left(1 - \frac{2b'}{a + (k+1)^\tau}\right) \frac{1}{k+1} \\ &\geq \frac{k}{k+1} \mathcal{P}_\beta(k) - \frac{2b''}{k^\tau} + \frac{1}{k+1}, \end{aligned} \quad (71)$$

for appropriately chosen  $b'' > 0$ , and for all  $k \geq k_1$ , where  $k_1$  is sufficiently large. Now, consider

$$\mathcal{U}(k+1) = \frac{k}{k+1} \mathcal{U}(k) + \frac{1}{k+1} - \frac{2b''}{k^\tau},$$

$$k = k_1, k_1+1, \dots, \mathcal{U}(k_1) = \mathcal{P}_\beta(k_1). \quad (72)$$

Clearly,  $\mathcal{P}_\beta(k) \geq \mathcal{U}(k)$ , for all  $k \geq k_1$ . Subtracting 1 from both sides in (72) and by Lemma 6,  $\mathcal{U}(k) \rightarrow 1$ , and so  $\liminf_{k \rightarrow \infty} \mathcal{P}_\beta(k) \geq 1$ ; also,  $\mathcal{P}_\beta(k) \leq 1$ , for all  $k$ , and, thus (69) for  $\tau > 1$ .

*Case*  $\tau = 1$ : Consider (71); as  $\tau = 1$ , we have

$$\mathcal{P}_\beta(k+1) \geq \frac{k}{k+1} \left(1 - \frac{2b'}{a + k+1}\right) \mathcal{P}_\beta(k)$$

$$+ \frac{1}{k+1} - \frac{2b'}{(k+1)^2}, \quad k = 1, 2, \dots$$

Now, define the recursion

$$\mathcal{V}(k+1) = \frac{k}{k+1} \left(1 - \frac{2b'}{a + k+1}\right) \mathcal{V}(k)$$

$$+ \frac{1}{k+1} - \frac{2b'}{(k+1)^2}, \quad k = 1, 2, \dots$$

Similarly to the proof of (31) in Corollary 4, it can be shown that  $\mathcal{V}(k) \rightarrow \frac{1}{1+2b'}$ . Noting that  $\mathcal{P}_\beta(k) \geq \mathcal{V}(k)$ ,  $k = 1, 2, \dots$  yields (69) for  $\tau = 1$ . Having proved (78), taking the  $\liminf$  in (68), and combining the obtained limit with (67), completes the proof of Theorem 5.

## REFERENCES

- [1] T. Derham, S. Doughty, K. Woodbridge, and C. Baker, "Design and evaluation of a low-cost multistatic netted radar system," *IET Radar, Sonar Navig.*, vol. 1, no. 5, pp. 362–368, Oct. 2007.
- [2] F. S. Cattivelli and A. H. Sayed, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 1917–1932, May 2011.
- [3] D. Bajovic, D. Jakovetic, J. M. F. Moura, J. Xavier, and B. Sinopoli, "Large deviations performance of consensus+innovations distributed detection with non-Gaussian observations," Nov. 2011 [Online]. Available: <http://arxiv.org/abs/1111.4555>
- [4] J. N. Tsitsiklis, "Decentralized detection," *Adv. Statist. Signal Process.*, vol. 2, pp. 297–344, 1993.
- [5] 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, Jan. 1997.
- [6] R. Viswanathan and P. R. Varshney, "Decentralized detection with multiple sensors: Part I—Fundamentals," *Proc. IEEE*, vol. 85, pp. 54–63, Jan. 1997.
- [7] J. F. Chamberland and V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 407–416, Feb. 2003.
- [8] S. A. Aldosari and J. M. F. Moura, "Detection in sensor networks: The saddlepoint approximation," *IEEE Trans. Signal Process.*, vol. 55, no. 1, pp. 327–340, Jan. 2007.
- [9] V. Saligrama, M. Alanyali, and O. Savas, "Distributed detection in sensor networks with packet losses and finite capacity links," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4118–4132, Nov. 2006.
- [10] S. Kar and J. M. F. Moura, "Consensus based detection in sensor networks: Topology design under practical constraints," presented at the IEEE Workshop Inf. Theory Sensor Netw., Santa Fe, NM, Jun. 2007.
- [11] S. Kar, S. A. Aldosari, and J. M. F. Moura, "Topology for distributed inference on graphs," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2609–2613, Jun. 2008.
- [12] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 355–369, Jan. 2009.
- [13] D. Jakovetic, J. Xavier, and J. M. F. Moura, "Weight optimization for consensus algorithms with correlated switching topology," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3788–3801, July 2010.
- [14] R. Olfati-Saber, "Distributed Kalman filtering for sensor networks," in *Proc. 46th IEEE Conf. Decision Control (CDC)*, New Orleans, LA, Dec. 2007, pp. 5492–5498.
- [15] R. Rahman, M. Alanyali, and V. Saligrama, "Distributed tracking in multihop sensor networks with communication delays," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4656–4668, Sep. 2007.
- [16] X. Zhao and A. H. Sayed, "Diffusion adaptation over networks under imperfect information exchange and non-stationary data," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3460–3475, 2012.
- [17] F. S. Cattivelli and A. H. Sayed, "Distributed detection over adaptive networks based on diffusion estimation schemes," in *Proc. IEEE 10th IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Perugia, Italy, Jun. 2009, pp. 61–65.
- [18] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS-based detection over adaptive networks," in *Proc. 39th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Oct. 2009, pp. 171–175.
- [19] A. Khalili, M. A. Tinati, A. Rastegarnia, and J. A. Chambers, "Steady-state analysis of diffusion LMS adaptive networks with noisy links," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 974–979, 2012.

- [20] S. Stankovic, N. Ilic, M. S. Stankovic, and K. H. Johansson, "Distributed change detection based on a consensus algorithm," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5686–5697, Dec. 2011.
- [21] P. Braca, S. Marano, V. Matta, and P. Willet, "Asymptotic optimality of running consensus in testing binary hypothesis," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 814–825, Feb. 2010.
- [22] D. Bajovic, D. Jakovetic, J. Xavier, B. Sinopoli, and J. M. F. Moura, "Asymptotic performance of distributed detection over random networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 3008–3011.
- [23] D. Bajovic, D. Jakovetic, J. Xavier, B. Sinopoli, and J. M. F. Moura, "Distributed detection via Gaussian running consensus: Large deviations asymptotic analysis," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4381–4936, Sep. 2011.
- [24] D. Bajovic, D. Jakovetic, J. Xavier, B. Sinopoli, and J. M. F. Moura, "Distributed detection over time varying networks: Large deviations analysis," in *Proc. 44th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, Sep. 2010, pp. 302–309.
- [25] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3575–3605, 2012.
- [26] S. Kar, R. Tandon, H. V. Poor, and S. Cui, "Distributed detection in noisy sensor networks," presented at the IEEE Int. Symp. Inf. Theory (ISIT), Saint Petersburg, Russia, Aug. 2011.
- [27] H. L. V. Trees, *Detection, Estimation, and Modulation Theory—Part I—Detection, Estimation, and Linear Modulation Theory*. New York: Wiley, 1968.
- [28] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Boston, MA: Jones and Barlett, 1993.
- [29] G. H. Hardy, Riesz, and Marcel, *The General Theory of Dirichlet's Series*. Cambridge, U.K.: Cambridge Univ. Press, 1915, Cambridge Tracts in Mathematics 18.
- [30] S. Cui, J.-J. Xiao, A. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Energy-efficient joint estimation in sensor networks: Analog vs. digital," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Philadelphia, PA, Mar. 2008, pp. iv/745–iv/748.
- [31] B. T. Polyak, *Introduction to Optimization*. New York: Optimization Software Inc., Publications Div., 1987.
- [32] S. Kar and J. M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and trade-offs," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 4, pp. 674–690, Aug. 2011.



**José M. F. Moura** (S'71–M'75–SM'90–F'94) received the Engenheiro Electrotécnico degree from the Instituto Superior Técnico (IST), Lisbon, Portugal, and the M.Sc., E.E., and D.Sc. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge.

He was formerly on the faculty at IST, has held visiting faculty appointments at the Massachusetts Institute of Technology, and was a visiting research scholar at the University of Southern California.

He is currently a University Professor with the Departments of Electrical and Computer Engineering and, by courtesy, BioMedical Engineering, at Carnegie Mellon University (CMU), Pittsburgh, PA. He is a founding Co-Director of the Center for Sensed Critical Infrastructures Research (CenSCIR) and manages a large education and research program between CMU and Portugal ([www.icti.cmu.edu](http://www.icti.cmu.edu)). His research interests include statistical and algebraic signal processing, image, bioimaging, and video processing, and digital communications. He has published over 450 technical journal and conference papers, filed 11 patents, eight of which are issued by the U.S. Patent Office, is the co-editor of two books, and has given numerous invited seminars at international conferences, U.S. and European universities, and industrial and government laboratories.

Dr. Moura is a member of the IEEE Board of Directors (2012–2013) and the IEEE Division IX Director and currently serves on the IEEE Publications Services and Products Board. He was the President (2008–2009) of the IEEE Signal Processing Society (SPS), Editor-in-Chief for the IEEE TRANSACTIONS IN SIGNAL PROCESSING, interim Editor-in-Chief for the IEEE SIGNAL PROCESSING LETTERS, and was on the Editorial Board of several journals, including the IEEE PROCEEDINGS, the *IEEE Signal Processing Magazine*, and the *ACM Transactions on Sensor Networks*. He has been on the steering and technical committees of several conferences. He is a Fellow of the American Association for the Advancement of Science (AAAS) and a corresponding member of the Academy of Sciences of Portugal (Section of Sciences). He was awarded the 2003 IEEE Signal Processing Society Meritorious Service Award and in 2000 the IEEE Millennium Medal. In 2007, he received the CMU's Philip L. Dowd Fellowship Award for Contributions to Engineering Education. In 2010, he was elected University Professor at CMU and received the IEEE Signal Processing Technical Achievement Award. He is affiliated with several IEEE societies, Sigma Xi, AAAS, AMS, IMS, and SIAM.



**Dušan Jakovetić** (S'10) received the Dipl.Ing. diploma from the School of Electrical Engineering, University of Belgrade, in 2007.

Currently, he is working towards the Ph.D. degree in electrical and computer engineering within the joint program of Carnegie Mellon University, Pittsburgh, PA, and Instituto de Sistemas e Robótica (ISR), Instituto Superior Técnico (IST), Lisbon, Portugal. His research interests include consensus, distributed inference, and distributed optimization.



**João Xavier** (S'97–M'03) received the Ph.D. degree in electrical and computer engineering from the Instituto Superior Técnico (IST), Lisbon, Portugal, in 2002.

Currently, he is an Assistant Professor in the Department of Electrical and Computer Engineering, IST. He is also a Researcher at the Institute of Systems and Robotics (ISR), Lisbon, Portugal. His current research interests are in the area of optimization and statistical inference for distributed systems.