

## LARGE DEVIATIONS ANALYSIS OF ADAPTIVE DISTRIBUTED DETECTION

Paolo Braca<sup>(1)</sup> Stefano Marano<sup>(2)</sup> Vincenzo Matta<sup>(2)</sup> Ali H. Sayed<sup>(3)\*</sup><sup>(1)</sup> Research Department, NATO-STO Centre for Maritime Research and Experimentation, La Spezia, Italy<sup>(2)</sup> Department of Information & Electrical Engineering and Applied Mathematics, University of Salerno, Fisciano (SA), Italy<sup>(3)</sup> Electrical Engineering Department, University of California, Los Angeles, CA, USA

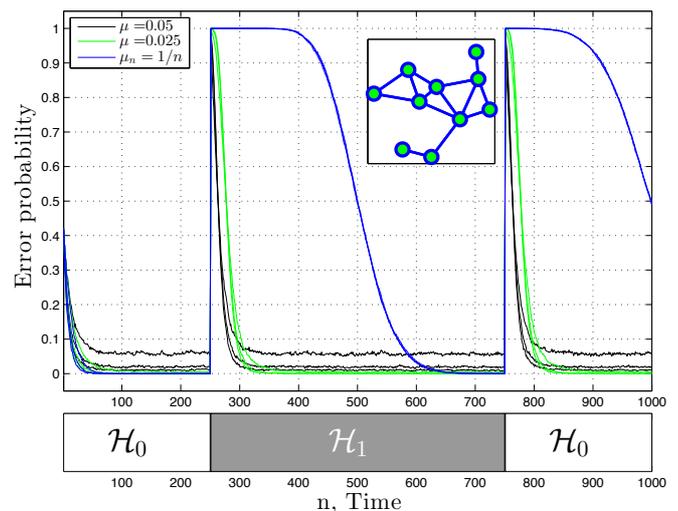
## ABSTRACT

In distributed inference, local cooperation among network nodes can be exploited to enhance the performance of each individual agent, but a challenging requirement for networks operating in dynamic real-world environments is that of *adaptation*. The interplay between these two fundamental aspects of cooperation and adaptation has been investigated in recent years in the context of estimation problems. Less explored in the literature is the case of *detection*, which is our focus. Capitalizing on the powerful tool of large deviations analysis, we show how to design and characterize the performance of diffusion strategies that reconcile both needs of adaptation and detection in decentralized systems.

## 1. MOTIVATION

Enabling continuous learning and adaptation over distributed networks engaged in inferential tasks is critical for the successful operation of decentralized solutions under dynamically changing conditions. In this work, we show how this objective can be achieved by means of diffusion strategies and, more importantly, we pursue a large deviations analysis to establish that the proposed structure leads to an exponential decay in the probability of error. We characterize this rate of decay by deriving a closed form expression for the steady-state error exponent.

We thus consider a collection of  $S$  agents that are assumed to monitor a certain phenomenon of interest. As time elapses, the nodes collect an increasing amount of data, whose statistical properties depend upon an *unknown* state of nature, formally represented by the hypotheses  $\mathcal{H}_0$  and  $\mathcal{H}_1$ . At each time instant, each sensor must make a decision, based on its local observations and on the information exchanged with neighboring nodes. We emphasize the need for *adaptation* since in the assumed model, the true hypothesis may drift over time, and the network must be able to react to this situation. This scenario is illustrated in Fig. 1, where we show the simulated time-evolution of the error probabilities achieved by certain agents of the network (three nodes out of ten), for different distributed strategies. We show in the lower part of the figure that the true (unknown) hypothesis changes at certain (unknown) epochs following the pattern  $\mathcal{H}_0 \rightarrow \mathcal{H}_1 \rightarrow \mathcal{H}_0$ . The error probabilities corresponding to the different strategies



**Fig. 1.** Adaptive diffusion algorithms at work. The network shown in the inset plot is engaged in a detection problem. *Top panel:* time-evolution of the error probability at three nodes with  $i$ ) the diffusion strategy with different step-sizes  $\mu = 0.025, 0.05$ , and  $ii$ ) the running consensus strategy (decaying step-size  $\mu_n = 1/n$ ). *Bottom panel:* Actual pattern of the true hypothesis.

show different degrees of adaptation, quantifiable in terms of the delay needed for identifying a change in the state of nature. The simulation shows the learning curves for two adaptive diffusion implementations [1–3] with *constant* step-sizes and one adaptive running consensus implementation with a decaying [4–9] step-size sequence of the form  $\mu_n = 1/n$ , where  $n$  is the time index. It is seen that this latter case is not apt at tracking drifts in the underlying state of nature. This inability to track drifts degrades further as time progresses since the step-size sequence  $\mu_n = 1/n$  decays to zero as  $n \rightarrow \infty$ . For this reason, in this work, we shall set the step-sizes to constant values to enable continuous adaptation and learning by the distributed network of detectors. In order to evaluate how well these adaptive networks perform, we need to be able to assess the goodness of the inference performance (reliability of the decisions), so as to exploit the trade-off between adaptation and learning capabilities. This is the main theme of the paper.

## 2. SUMMARY OF MAIN RESULTS

Recent literature on the subject focuses on the detection performance of decentralized consensus algorithms with decay-

\*The work of A. H. Sayed was supported in part by NSF grant CCF-1011918.

ing step-size [4–9]. For the case of constant step-size, while several results have been obtained for the mean-square-error (MSE) *estimation* performance of adaptive networks [1, 10], less is known about the performance of distributed detection networks (see, e.g., [11]). This is mainly due to the fact that results on the asymptotic distribution of the error quantities under constant step-size adaptation over networks are largely unavailable in the literature. While [12] argues that the error in single-agent LMS adaptation converges in distribution, the resulting distribution is not characterized. Only recently these questions have been studied in [13, 14] in the context of distributed estimation. Nevertheless, these results on the asymptotic distribution of the errors are still insufficient to characterize the rate of decay of the probability of error over distributed networks. To do so, it is necessary to pursue a large deviations analysis in the constant step-size regime. Motivated by these remarks, we therefore provide a thorough statistical characterization of the diffusion network in a manner that enables detector design and analysis.

The main result established in this paper can be summarized by the following formulas:

$$\alpha_{k,\mu} \doteq e^{-(1/\mu)S\mathcal{E}_0}, \quad \beta_{k,\mu} \doteq e^{-(1/\mu)S\mathcal{E}_1}, \quad (1)$$

where  $\alpha_{k,\mu}$  and  $\beta_{k,\mu}$  represent the Type-I (false-alarm) and Type-II (miss-detection) error probabilities at the  $k$ -th sensor, corresponding to the *steady-state* (i.e.,  $n \rightarrow \infty$ ) output of the distributed diffusion strategy with *constant* step-size  $\mu$ . The notation  $\doteq$  in (1) means equality to the first-order in the exponent as  $\mu$  goes to zero. Moreover, the factors  $\mathcal{E}_0$  and  $\mathcal{E}_1$  are functions of the moment generating function of the single-sensor data, and of the decision regions; they are *independent* of the step-size, the number of sensors  $S$ , and the network connectivity.

Result (1) has several insightful ramifications. At a fundamental level, Eq. (1) reveals that the error probabilities are driven to zero exponentially fast as functions of  $1/\mu$ , and that the error exponents governing this decay increase linearly in the number of sensors. It is instructive to compare this detection scaling law to the already known results for adaptive distributed estimation over diffusion networks, which state that the MSE attained by sensor  $k$  obeys [1, 10]:  $\text{MSE}_k \propto \mu/S$ . Thus, the scaling laws governing errors of detection and estimation over diffusion networks behave very differently, the former being exponential with decay proportional to  $1/\mu$ , while the latter is linear with decay proportional to  $\mu$ . This reveals an interesting analogy with other more traditional inferential schemes, such as, for example, (a) the centralized, non-adaptive inferential system with  $N$  independent and identically distributed (i.i.d.) data points [15, 16], where error probabilities vanish exponentially fast as functions of  $N$ , and the MSE decays as  $1/N$ ; and (b) the case of multiterminal inference with bit-rate  $R$  [17], where error probabilities vanish exponentially fast as functions of  $R$ , and the MSE decays as  $1/R$ . The step-size  $\mu$  emerges thus as the basic parameter quantifying the cost of the information used by the network for inference purposes, much as the number of data  $N$  or the bit-rate  $R$  in the considered examples.

On a more practical level, Eq. (1) characterizes the asymptotic performance of each agent in the network, revealing in

particular that: *i*) the inferential diffusion strategy equalizes the detection performance of the agents, in the sense that the error exponents do not depend on the particular sensor  $k$ ; *ii*) cooperation offers *exponential gains* in terms of detection performance, as the error exponents increase linearly with  $S$ ; *iii*) the diffusion strategy performs asymptotically as well as the centralized stochastic-gradient solution, because this latter is equivalent to a *fully* connected case [1, 18]. We now move on to describe the adaptive distributed solution and to establish its main property (1). Throughout the paper, we use boldface letters to denote random variables, and normal font letters for their realizations. Capital letters refer to matrices, small letters to both vectors and scalars. Exceptions to these rules will be obvious from the context.

### 3. PROBLEM SETUP

The scalar observation collected by the  $k$ -th sensor at time  $n$  is denoted by  $\mathbf{x}_k(n)$ ,  $k = 1, \dots, S$ , and it arises from a distribution with mean  $\mathbb{E}\mathbf{x}$  and variance  $\sigma_x^2$ . The data across the sensors are assumed to be spatially and temporally i.i.d., conditioned on the hypothesis. It is well-known that for the i.i.d. model, an optimal centralized (and non-adaptive) detection statistic is the sum of the log-likelihoods. When the latter are not available, detection statistics obtained as a sum of some suitably chosen functions of the observations are often employed as alternatives [19, 20]. Accordingly, we assume that  $\mathbf{x}_k(n)$  represents the local statistic that is available to sensor  $k$  at time  $n$ . Given the objective of mimicking weighted averages, we therefore resort to the class of diffusion strategies for adaptation over networks [1–3, 11]. Among the various forms, we consider the Adapt-then-Combine (ATC) form, due to some inherent advantages in terms of MSE performance [1, 3]. In the ATC diffusion implementation, each node  $k$  updates its state from  $\mathbf{y}_k(n-1)$  to  $\mathbf{y}_k(n)$  through local cooperation with its neighbors as follows:

$$\mathbf{r}_k(n) = \mathbf{y}_k(n-1) + \mu[\mathbf{x}_k(n) - \mathbf{y}_k(n-1)], \quad (2)$$

$$\mathbf{y}_k(n) = \sum_{\ell=1}^S a_{k,\ell} \mathbf{r}_\ell(n). \quad (3)$$

In this construction, node  $k$  first uses its local statistic,  $\mathbf{x}_k(n)$ , to update its state from  $\mathbf{y}_k(n-1)$  to an intermediate value  $\mathbf{r}_k(n)$ . All other nodes in the network are performing similar updates simultaneously using their local statistics. Subsequently, node  $k$  aggregates the intermediate states of its neighbors using nonnegative convex combination weights  $\{a_{k,\ell}\}$  that add up to one. Again, all other nodes in the network perform a similar calculation. The above adaptation and aggregation steps can be combined into a single equation as follows:

$$\mathbf{y}_k(n) = \sum_{\ell=1}^S a_{k,\ell} (\mathbf{y}_\ell(n-1) + \mu[\mathbf{x}_\ell(n) - \mathbf{y}_\ell(n-1)]), \quad (4)$$

where  $0 < \mu \ll 1$  is a small step-size parameter. If we collect the combination coefficients into a matrix  $A = [a_{k,\ell}]$ , then  $A$  is right-stochastic satisfying  $A\mathbf{1} = \mathbf{1}$ , with  $\mathbf{1}$  being a column-vector with all entries equal to one. Since we

are interested in reaching a *balanced* fusion of the observations, we further assume that  $A$  is *doubly-stochastic* with second largest eigenvalue magnitude strictly less than one, which yields  $A^n \xrightarrow{n \rightarrow \infty} \mathbb{1}\mathbb{1}^T/S$  [21, 22].

At time  $n$ , the  $k$ -th sensor runs a decision algorithm based upon  $\mathbf{y}_k(n)$ , whose performance is measured according to the Type-I and Type-II error probabilities:

$$\alpha_k(n) \stackrel{\text{def}}{=} \mathbb{P}_0[\mathbf{y}_k(n) \in \Gamma_1], \quad \beta_k(n) \stackrel{\text{def}}{=} \mathbb{P}_1[\mathbf{y}_k(n) \in \Gamma_0], \quad (5)$$

where  $\mathbb{P}_h[\cdot]$ , with  $h = 0, 1$ , is the probability operator under  $\mathcal{H}_h$ , and the decision regions in favor of  $\mathcal{H}_0$  and  $\mathcal{H}_1$  are denoted by  $\Gamma_0$  and  $\Gamma_1$ , respectively.

Computation of the exact distributions of  $\mathbf{y}_k(n)$  is generally intractable, implying that the structure of the test is unknown. However, we are able to address this problem in the slow adaptation regime for sufficiently small step-sizes. In this case, we are able to show that (a)  $\mathbf{y}_k(n)$  has a *limiting distribution* as  $n$  goes to infinity (Theorem 1); (b) the distribution of  $\mathbf{y}_k(n)$  approaches a Gaussian, i.e., it is *asymptotically normal* (Theorem 2); (c) the *large deviations* of the steady-state output  $\mathbf{y}_k(n)$  can be characterized (Theorem 3); and (d) these findings are key for designing the distributed inference system and characterizing its performance (Theorem 4). The proofs of the theorems are omitted for space constraints, and can be found in [24].

#### 4. STEADY-STATE DISTRIBUTION

Let  $\mathbf{x}_n$  and  $\mathbf{y}_n$  denote the  $S \times 1$  vectors that collect, respectively, the local statistics  $\{\mathbf{x}_k(n)\}$  and the state variables  $\{\mathbf{y}_k(n)\}$  from across the network at time  $n$ . It is then straightforward to verify from the diffusion strategy (4) that the state of the  $k$ -th sensor can be written as:

$$\mathbf{y}_k(n) = \text{transient} + \sum_{i=1}^n \mathbf{z}_k(i), \quad (6)$$

where

$$\text{transient} = (1 - \mu)^n \sum_{\ell=1}^S b_{k,\ell}(n) \mathbf{y}_\ell(0), \quad (7)$$

$$\mathbf{z}_k(i) = \mu(1 - \mu)^{i-1} \sum_{\ell=1}^S b_{k,\ell}(i) \mathbf{x}_\ell(n - i + 1), \quad (8)$$

and the scalars  $b_{k,\ell}(n)$  are the entries of  $B(n) \stackrel{\text{def}}{=} A^n$ . The rightmost term in (6) involves a sum of independent, but *not* identically distributed, random variables. It is however straightforward to verify that

$$\mathbb{E} \left( \sum_{i=1}^n \mathbf{z}_k(i) \right) \xrightarrow{n \rightarrow \infty} \mathbb{E} \mathbf{x}, \quad \text{VAR} \left( \sum_{i=1}^n \mathbf{z}_k(i) \right) \leq \frac{\sigma_x^2 S \mu}{2 - \mu} < \infty. \quad (9)$$

In view of the Infinite Convolution Theorem [23, p. 266], these two conditions are sufficient to establish that the second term on the RHS of (6), i.e., the sum of the random variables  $\mathbf{z}_k(i)$ , converges in distribution to a random variable  $\mathbf{y}_{k,\mu}^*$  as

$n \rightarrow \infty$ , and that the first two moments of the limiting distribution are equal to  $\mathbb{E} \mathbf{x}$  and  $\sum_{i=1}^{\infty} \text{VAR}(\mathbf{z}_k(i))$ . Note that we are making explicit the dependence of  $\mathbf{y}_{k,\mu}^*$  on the step-size  $\mu$  for later use. Since the first term on the RHS of (6) vanishes with  $n$ , by application of Slutsky's Theorem [15] we have in fact shown that, at the steady-state, the diffusion output  $\mathbf{y}_k(n)$  is distributed as  $\mathbf{y}_{k,\mu}^*$ . The above findings are summarized in the following result (where the symbol  $\rightsquigarrow$  means convergence in distribution).

**THEOREM 1:** (*Steady-state distribution of  $\mathbf{y}_k(n)$* ). *The state variable  $\mathbf{y}_k(n)$  that is generated by the diffusion strategy (4) is asymptotically stable in distribution, namely,*

$$\mathbf{y}_k(n) \xrightarrow{n \rightarrow \infty} \mathbf{y}_{k,\mu}^*. \quad (10)$$

□

So far we have only proved that a steady-state distribution for  $\mathbf{y}_k(n)$  exists. While the exact form of the steady-state distribution is generally impossible to evaluate, it is nevertheless possible to approximate it well for small values of the step-size parameter.

#### 5. THE SMALL- $\mu$ REGIME

We start by stating the following result (where  $\mathcal{N}(a, b)$  denotes a Gaussian distribution with mean  $a$  and variance  $b$ ).

**THEOREM 2:** (*Asymptotic normality of  $\mathbf{y}_{k,\mu}^*$  as  $\mu \rightarrow 0$* ). *Under the assumption  $\mathbb{E}|\mathbf{x}_k(n)|^3 < \infty$ , we have, for all  $k = 1, \dots, S$ :*

$$\frac{\mathbf{y}_{k,\mu}^* - \mathbb{E} \mathbf{x}}{\sqrt{\mu}} \xrightarrow{\mu \rightarrow 0} \mathcal{N} \left( 0, \frac{\sigma_x^2}{2S} \right). \quad (11)$$

□

Theorem 2 provides an approximation of the diffusion output distribution for small step-sizes. At first glance, this may seem enough to approximate the detection error probabilities. A closer inspection reveals that this is not the case. From Theorem 2 we learn that, as  $\mu \rightarrow 0$ , the diffusion output shrinks down to its limiting expectation  $\mathbb{E} \mathbf{x}$  and that the *small* (of order  $\sqrt{\mu}$ ) deviations around this value have a Gaussian shape. Consider instead the evaluation of terms like

$$\mathbb{P}[|\mathbf{y}_{k,\mu}^* - \mathbb{E} \mathbf{x}| > \delta] \xrightarrow{\mu \rightarrow 0} 0, \quad \delta > 0. \quad (12)$$

While the above convergence to zero can be inferred from (11), it is well known that (11) is not sufficient in general to characterize the rate at which the probability vanishes. Assessing the rate of convergence is critical for the accurate design and characterization of reliable inference systems [25, 26].

In order to fill this gap, the study of the *large deviations* of  $\mathbf{y}_{k,\mu}^*$  is needed. We will be showing in the sequel that the process  $\mathbf{y}_{k,\mu}^*$  obeys a Large Deviation Principle (LDP), namely, that the following limit exists [25, 26]:

$$\lim_{\mu \rightarrow 0} \mu \ln \mathbb{P}[\mathbf{y}_{k,\mu}^* \in \Gamma] = - \inf_{\gamma \in \Gamma} I(\gamma), \quad (13)$$

for some  $I(\gamma)$  that is called *rate function*. We determine the expression for  $I(\gamma)$  in Theorem 3. The above equation can be equivalently rewritten as

$$\mathbb{P}[\mathbf{y}_{k,\mu}^* \in \Gamma] \doteq e^{-(1/\mu) \inf_{\gamma \in \Gamma} I(\gamma)}, \quad (14)$$

which shows how the LDP generally implies an exponential scaling law for probabilities, with an exponent governed by the rate function. One basic ingredient of Theorem 3 is the Gärtner-Ellis Theorem [25, 26], which is stated next in a form that uses directly the set of assumptions relevant for our purposes.

**GÄRTNER-ELLIS THEOREM [25].** *Let  $z_\mu$  be a family of random variables with Logarithmic Moment Generating Function (LMGF)  $\phi_\mu(t) = \ln \mathbb{E} \exp\{tz_\mu\}$ . If*

$$\phi(t) \stackrel{\text{def}}{=} \lim_{\mu \rightarrow 0} \mu \phi_\mu(t/\mu) \quad (15)$$

*exists, with  $\phi(t) < \infty$  for all  $t \in \mathbb{R}$ , and  $\phi(t)$  is differentiable in  $\mathbb{R}$ , then  $z_\mu$  satisfies the LDP property (13) with rate function given by the Fenchel-Legendre transform of  $\phi(t)$ , namely,*

$$\Phi(\gamma) \stackrel{\text{def}}{=} \sup_{t \in \mathbb{R}} [\gamma t - \phi(t)]. \quad (16)$$

□

We now show how the result allows deriving the asymptotic performance of the diffusion output in our inferential network. Let us introduce the LMGF of the data  $x_k(n)$ , and that of the steady-state variable  $y_{k,\mu}^*$ , respectively:

$$\psi(t) \stackrel{\text{def}}{=} \ln \mathbb{E} \exp\{tx_k(n)\}, \quad \phi_{k,\mu}(t) \stackrel{\text{def}}{=} \ln \mathbb{E} \exp\{ty_{k,\mu}^*\}. \quad (17)$$

Our main theorem about the large deviations of the diffusion output is the following.

**THEOREM 3: (Large deviations of  $y_{k,\mu}^*$  as  $\mu \rightarrow 0$ ).** *Assume that  $\psi(t) < +\infty \forall t \in \mathbb{R}$ , and define  $\omega(t) \stackrel{\text{def}}{=} \int_0^t \frac{\psi(\tau)}{\tau} d\tau$ . Then, for all  $k = 1, \dots, S$ :*

$$\phi(t) \stackrel{\text{def}}{=} \lim_{\mu \rightarrow 0} \mu \phi_{k,\mu}(t/\mu) = S \omega(t/S). \quad (18)$$

*Also, the steady-state variable  $y_{k,\mu}^*$  obeys the LDP with rate function  $I(\gamma) = S \Omega(\gamma)$ .* □

## 6. ADAPTIVE DISTRIBUTED DETECTION

In view of Theorem 1, we consider the behavior of (5) as  $n \rightarrow \infty$ , and write the steady-state detection performance as follows (assuming that  $y_{k,\mu}^*$  lies on the boundaries of the decision regions with probability zero [16, 27]):

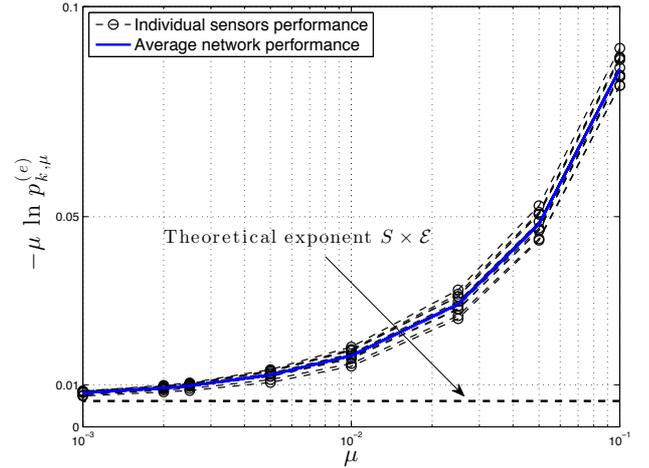
$$\alpha_{k,\mu} \stackrel{\text{def}}{=} \mathbb{P}_0[y_{k,\mu}^* \in \Gamma_1], \quad \beta_{k,\mu} \stackrel{\text{def}}{=} \mathbb{P}_1[y_{k,\mu}^* \in \Gamma_0], \quad (19)$$

where the dependence upon  $\mu$  has been made explicit. The large-deviations result offered by Theorem 3 can now be tailored to our detection setup as follows.

**THEOREM 4: (Detection error exponents).** *For  $h \in \{0, 1\}$ , let  $\Gamma_h$  be the decision regions – independent of  $\mu$  – and assume that  $\psi_h(t) < \infty$ , for all  $t \in \mathbb{R}$ . Also, let  $\omega_h(t) \stackrel{\text{def}}{=} \int_0^t \frac{\psi_h(\tau)}{\tau} d\tau$ . Then, for all  $k = 1, \dots, S$ , Eq. (1) holds with:*

$$\mathcal{E}_0 = \inf_{\gamma \in \Gamma_1} \Omega_0(\gamma), \quad \mathcal{E}_1 = \inf_{\gamma \in \Gamma_0} \Omega_1(\gamma). \quad (20)$$

□



**Fig. 2.** Shift-in-mean detection problem with Laplace noise. Here  $\sigma = 1$ ,  $\theta = 0.05$ , and the number of Monte Carlo runs is  $10^5$ .

Let us now provide some tangible examples showing applications of Theorem 4. We assume that  $x_k(n)$  is the local log-likelihood computed by the  $k$ -th sensor, and that a threshold detector is employed with  $\Gamma_0 = \{\gamma : \gamma \leq \eta\}$ . We start with a classical Gaussian shift-in-mean problem where the data arise from  $\mathcal{N}(0, \sigma^2)$  under  $\mathcal{H}_0$ , and from  $\mathcal{N}(\theta, \sigma^2)$  under  $\mathcal{H}_1$ , with  $\theta > 0$ . The sensors compute the corresponding log-likelihoods, which yields  $x_k(n) \sim \mathcal{N}(-D, 2D)$  under  $\mathcal{H}_0$ , and  $x_k(n) \sim \mathcal{N}(D, 2D)$  under  $\mathcal{H}_1$ , where  $D = \frac{\theta^2}{2\sigma^2}$  is the Kullback-Leibler divergence between the two hypotheses [16]. For this example, the rate functions can be found in closed form by applying the recipe in Theorem 4, yielding:

$$\Omega_0(\gamma) = \frac{(\gamma + D)^2}{2D}, \quad \Omega_1(\gamma) = \frac{(\gamma - D)^2}{2D}. \quad (21)$$

Setting the threshold  $\eta$  in the interval  $(-D, D)$  – a different choice will otherwise nullify one of the exponents – the minimization in (20) is easily performed using convexity properties of the rate functions (21). The final result is:

$$\alpha_{k,\mu} \stackrel{\text{def}}{=} e^{-(1/\mu) S \frac{(\eta+D)^2}{2D}}, \quad \beta_{k,\mu} \stackrel{\text{def}}{=} e^{-(1/\mu) S \frac{(\eta-D)^2}{2D}}. \quad (22)$$

In order to test the theoretical findings we now present some evidences arising from Monte Carlo simulations. The network employed is made of ten sensors, arranged so as to form the topology in the inset plot of Fig. 1, with combination weights  $a_{k,\ell}$  following the Laplacian rule [21]. Going beyond the Gaussian example, we consider a shift-in-mean problem with double-exponential (Laplace) noise. The probability density functions of the data under  $\mathcal{H}_0$  and  $\mathcal{H}_1$  are:

$$f_0(\xi) = \frac{1}{2\sigma} e^{-\frac{|\xi|}{\sigma}}, \quad f_1(\xi) = f_0(\xi - \theta). \quad (23)$$

Each sensor computes the local log-likelihood  $x_k(n)$ , and compares it to a threshold  $\eta = 0$ . By symmetry arguments,  $\alpha_{k,\mu} = \beta_{k,\mu} \stackrel{\text{def}}{=} p_{k,\mu}^{(e)}$ , and  $\mathcal{E}_0 = \mathcal{E}_1 \stackrel{\text{def}}{=} \mathcal{E}$ . In Fig. 2 we show the empirical error exponents  $-\mu \ln p_{k,\mu}^{(e)}$ , for  $k = 1, 2, \dots, S$ , along with the average network performance. It is seen that, in perfect agreement with Theorem 4, all sensors tend to reach the theoretical exponent  $S \times \mathcal{E}$  as the step-size  $\mu$  decreases.

## 7. REFERENCES

- [1] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, "Diffusion strategies for adaptation and learning over networks," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 155–171, May 2013.
- [2] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3122–3136, Jul. 2008.
- [3] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1035–1048, Mar. 2010.
- [4] P. Braca, S. Marano, and V. Matta, "Enforcing consensus while monitoring the environment in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3375–3380, Jul. 2008.
- [5] P. Braca, S. Marano, V. Matta, and P. Willett, "Asymptotic optimality of running consensus in testing statistical hypotheses," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 814–825, Feb. 2010.
- [6] —, "Consensus-based Page's test in sensor networks," *Signal Processing*, vol. 91, no. 4, pp. 919–930, Apr. 2011.
- [7] 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. 4381–4396, Sep. 2011.
- [8] 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," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5987–6002, Nov. 2012.
- [9] D. Jakovetic, J. M. F. Moura, and J. Xavier, "Distributed detection over noisy networks: Large deviations analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4306–4320, Aug. 2012.
- [10] X. Zhao and A. H. Sayed, "Performance limits for distributed estimation over LMS adaptive networks," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5107–5124, Oct. 2012.
- [11] 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.
- [12] R. R. Bitmead, "Convergence in distribution of LMS-type adaptive parameter estimates," *IEEE Trans. Autom. Control*, vol. 28, no. 1, pp. 54–60, Jan. 1983.
- [13] X. Zhao and A. H. Sayed, "Probability distribution of steady-state errors and adaptation over networks," in *Proc. IEEE International Workshop on Statistical Signal Processing (SSP)*, Nice, France, Jun. 2011, pp. 253–256.
- [14] J. Chen and A. H. Sayed, "On the probability distribution of distributed optimization strategies," in *Proc. IEEE GlobalSIP*, Austin, TX, USA, Dec. 2013.
- [15] H. Shao, *Mathematical Statistics*, 2nd ed., Springer, 2003.
- [16] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*. Springer, 2005.
- [17] H. Viswanathan and T. Berger, "The quadratic CEO problem," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1549–1559, Sep. 1997.
- [18] X. Zhao and A. H. Sayed, "Attaining optimal batch performance via distributed processing over networks," in *Proc. IEEE ICASSP*, Vancouver, Canada, May 2013, pp. 1–5.
- [19] S. A. Kassam, *Signal Detection in Non-Gaussian Noise*. Springer-Verlag, 1987.
- [20] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Springer-Verlag, 1988.
- [21] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems and Control Letters*, vol. 53, pp. 65–78, 2004.
- [22] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1985.
- [23] W. Feller, *An Introduction to Probability and Its Applications*, vol. 2, Wiley, NY, 1971.
- [24] P. Braca, S. Marano, V. Matta, and A. H. Sayed, "Asymptotic performance of adaptive distributed detection over networks," *submitted for publication*, January 2014. Available as arXiv:1401.5742v2 [cs.IT].
- [25] F. den Hollander, *Large Deviations*. American Mathematical Society, 2008.
- [26] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer, 1998.
- [27] P. Billingsley, *Convergence of Probability Measures*, 2nd ed., Wiley, 1999.