Large Deviations Analysis of Consensus+Innovations Detection in Random Networks

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

***Abstract*—We study the large deviations performance of consensus+innovations distributed detection over random networks, where each sensor, at each time k, weight averages its decision variable with its neighbors decision variables (consensus), and accounts for its new observation (innovation). Sensor observations are independent identically distributed (i.i.d.) both in time and space, but have generic (non Gaussian) distributions. The underlying network is random, described by a sequence of i.i.d. stochastic, symmetric weight matrices $W(k)$; we measure the corresponding speed of consensus by $|\log r|$, where r is the second largest eigenvalue of the second moment of $W(k)$. We show that distributed detection exhibits a phase transition behavior with respect to $|\log r|$: when $|\log r|$ is above a threshold, distributed detection is equivalent to the optimal centralized detector, i.e., has the error exponent equal to the Chernoff information. We explicitly quantify the optimality threshold for $|\log r|$ as a function of the log-moment generating function $\Lambda_0(\cdot)$ of a sensor's log-likelihood ratio. When below the threshold, we analytically find the achievable error exponent as a function of r and $\Lambda_0(\cdot)$. Finally, we illustrate by an example the dependence of the optimality threshold on the type of the sensor observations distribution.**

I. Introduction

We study the large deviations performance (error exponent) of consensus+innovations distributed detection over random networks, when sensors have generic, non-Gaussian observations. With consensus+innovations distributed detection, each sensor i, at each time k, updates its decision variable two-fold: 1) by weight averaging its decision variable with the neighbors decision variables (consensus); and 2) by accounting for its new observation (innovation.) The network is random; its connectivity is described by a sequence of independent identically distributed (i.i.d.) stochastic, symmetric matrices $\{W(k)\}$. We measure the speed of consensus by $r := \lambda_2\left(\mathbb{E}\left[W^2(k)\right]\right)$, where λ_2 denotes the second largest eigenvalue.[1] The sensors' observations are i.i.d., both in time and space.

We show that the consensus+innovations distributed detector exhibits a phase transition behavior with respect to r: when $|\log r|$ is above a threshold, the distributed detector achieves the optimal (centralized) error exponent. We explicitly quantify the optimality threshold as a function of the log-moment generating function (LMGF) Λ_0 of a sensor's log-likelihood ratio. We also quantify what is the fraction of the optimal error exponent the distributed detector can achieve, when below the threshold.

We show that the error exponent and the optimality threshold of the consensus+innovations detector depends on the probability distribution of the sensors' observations. This contrasts with the centralized (fusion–based) detector[2] where the error exponent is equal for two different types of distributions of the sensors' observations, provided that the two corresponding per sensor Chernoff informations are equal.

Our analysis assumes that the LMGF $\Lambda_0(\lambda)$ is finite on $\lambda \in \mathbb{R}$, which is satisfied in a wide range of problems, including the binary simple hypothesis problem, when the sensor noise satisfies a mild technical condition (detailed in Section II); other examples are discussed in the journal version of this paper [1].

Brief review of the literature. Distributed detection has been extensively studied in the context of parallel

The work of D. Bajovic, D. Jakovetic, J. Xavier, and B. Sinopoli is partially supported by: the Carnegie Mellon—Portugal Program under a grant from the Fundação para a Ciência e Tecnologia (FCT) from Portugal; and by grant CMU-PT/SIA/0026/2009. This work is also supported by ISR/IST plurianual funding (POSC program, FEDER). The work of J. M. F. Moura is partially supported by NSF under grants CCF-1011903 and CCF-1018509, and by AFOSR grant FA95501010291. D. Bajović and D. Jakovetić hold fellowships from FCT.

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

J. M. F. Moura is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA moura@ece.cmu.edu

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

B. Sinopoli is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA brunos@ece.cmu.edu

[1] The quantity r is always in $[0, 1]$; the smallest r, the faster consensus.

[2] In the centralized detection, all sensors, at all times, transmit their observations to a fusion node.

fusion architectures, e.g., [2], [3], [4], [5], [6], consensus-based detection, e.g., [7], [8], and, more recently, consensus+innovations type distributed detection, e.g., [9], [10], [11], [12], [13], [14].

Different variants of consensus+innovations distributed detection algorithms have been proposed; we consider here running consensus, the variant in [12]. Reference [12] also considers the asymptotic optimality of the running consensus distributed detector, but in a very different framework than the framework of this paper and the framework in [15]; see [15] for a comment on these differences.

We contrast this paper with our prior work in [15] and [9]. Here the network is random, while in [9] it is deterministic, time varying, where the union of networks over a finite time window is connected. With respect to [15], we consider here generic distributions (with finite LMGF on $\mathbb{R}$,) while in [15] the sensor observations are Gaussian. Further, reference [15] allows for spatially correlated observations, while here the observations are spatially independent.

The remainder of the paper is organized as follows. Section II gives preliminaries on the centralized detector. Section III presents the consensus+innovations distributed detector. Section IV presents our main results on the asymptotic performance of the distributed detector. Section V illustrates our results with Gaussian and Laplace distributions. Finally, Section VI concludes the paper.

Throughout, we denote by: Z_{ij} the entry in the i-th row and j-th column of a matrix Z; z_i the i-th entry of a vector z; $|a|$ the modulus of a scalar a; 1 the $N \times 1$ vector with unit entries; J the $N \times N$ ideal consensus matrix $J := \frac{1}{N}11^\top$; $\|\cdot\| = \|\cdot\|_2$ the Euclidean (respectively, spectral) norm of its vector (respectively, matrix) argument; $\lambda_i(\cdot)$ the i-th largest eigenvalue; $\mathbb{E}[\cdot]$ and $\mathbb{P}(\cdot)$ the expected value and probability, respectively.

II. Centralized detection

This section reviews the centralized log-likelihood ratio detector and its error exponent; the section also introduces the LMGF of a sensors log-likelihood ratio and reviews its relevant properties.

A. Log-likelihood ratio test

Consider the centralized (fusion-based) binary detection, where N sensors, at each time t, send their observations $Y_i(t)$ to a fusion node. Nature can be in one of two states: H_1 – event occurring and H_0 – event not occurring. The sensors' observations are independent and identically distributed (i.i.d.) both in time and in space, with distribution μ_l under hypothesis H_l, $l = 0, 1$, i.e., for $i = 1, \ldots, N$ and $t = 1, 2, \ldots$:

$$Y_i(t) \sim \begin{cases} \mu_1, & H_1 \\ \mu_0, & H_0. \end{cases} \tag{1}$$

Here μ_1 and μ_0 are mutually absolutely continuous, distinguishable measures. The prior probability of hypothesis H_1 is $\pi_1 = \mathbb{P}(H_1) \in (0, 1)$, and $\pi_0 = \mathbb{P}(H_0) = 1 - \pi_1$.

The log-likelihood ratio of the sensor i at time t, denoted by $L_i(t)$, is given by

$$L_i(t) = \log \frac{d\mu_1}{d\mu_0} (Y_i(t)),$$

where $\frac{d\mu_1}{d\mu_0}(\cdot)$ is the Radon-Nikodym derivative of μ_1 with respect to μ_0. The log-likelihood ratio for the vector $Y(t) := (Y_1(t), \ldots, Y_N(t))$ of all sensors' observations is given by:

$$\sum_{i=1}^{N} L_i(t). \tag{2}$$

Thus, the centralized log-likelihood ratio test for the observation interval of size k and based on all sensors' observations takes the form:

$$D(k) := \sum_{t=1}^{k} \sum_{i=1}^{N} L_i(t) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_k, \tag{3}$$

where γ_k is a chosen threshold.

B. Log-moment generating function (LMGF)

The error exponent for the optimal centralized detector can be expressed in terms of the LMGF of a sensor's log-likelihood ratio, e.g., [16]. We now introduce the LMGF and its relevant properties. Denote by Λ_0 the LMGF for the log-likelihood ratio under hypothesis H_0:

$$\Lambda_0(\lambda) = \log \mathbb{E}\left[e^{\lambda L_1(1)} | H_0\right]. \tag{4}$$

Also, define

$$\Lambda_1(\lambda) = \log \mathbb{E}\left[e^{-\lambda L_1(1)} | H_1\right].$$

It can be shown [16] that Λ_0 is convex and $\Lambda_1(\lambda) = \Lambda_0(1 - \lambda)$, for $\lambda \in \mathbb{R}$.

Throughout the paper, we assume that $\Lambda_0(\lambda) < +\infty$, $\forall \lambda \in \mathbb{R}$. The latter condition holds, if, e.g., $Y_i(t) = m + n_i(t)$, under H_1, and $n_i(t)$, under H_0, where $m \in \mathbb{R}$ is a constant signal, and $n_i(t)$ is a zero-mean additive noise with density $f(\cdot)$, supported on $\mathbb{R}$, that satisfies a mild technical condition. We give a complete, formal account for the condition on $f(\cdot)$ in the companion journal paper [1]. Examples of $f(\cdot)$ that yield finite Λ_0 include the following. Let $f(y) = c\, e^{-g(y)}$, where $c > 0$ is a constant. Then, Λ_0 is finite on $\mathbb{R}$ if $g(y)$ is a polynomial in y of arbitrary finite degree; or $g(y) = y^\theta$, $\theta \in (0, 1)$ or $g(y) = c \log y$, $c \in [2, +\infty)$. The last case covers power laws with decay coefficient greater or equal two.

C. Error exponent

Denote by $P_{\mathrm{e}}(k)$ the (Bayes) error probability of the optimal centralized detector and the observation interval of size k. When k grows unbounded, the probability of error with the optimal centralized detector decays exponentially fast to zero. The rate of the decay (the error exponent) is given by the Chernoff lemma [17] and equals the Chernoff information between the two joint distributions of all N sensors' observations under H_1 and H_0.

It can be shown that, under spatially and temporally i.i.d. sensors' observations (conditioned on either hypothesis,) the error exponent for the optimal centralized detector is given by:

$$\lim_{k\to\infty} -\frac{1}{k}\log P_{\mathrm{e}}(k) = N C_{\mathrm{ind}}, \tag{5}$$

where $C_{\mathrm{ind}} := \max_{\lambda\in[0,1]}\{-\Lambda_0(\lambda)\}$ is the per sensor Chernoff information.

III. Distributed detection: Consensus+innovations

We now consider distributed detection when sensors cooperate through a randomly varying network. Specifically, we consider the running consensus distributed detection, proposed in [12]. At each time k, each sensor i improves its decision variable, call it $x_i(k)$, in two ways: 1) by incorporating its new observation at time k; and 2) by exchanging the decision variable (with incorporated new observation) locally with its neighbors and computing the weighted average of its own and the neighbors' variables.

More precisely, the update of $x_i(k)$ is as follows:

$$x_i(k) = \sum_{j\in O_i(k)} W_{ij}(k)\left(\frac{k-1}{k}x_j(k-1)+\frac{1}{k}L_j(k)\right)$$

$$k = 1,2,...,\;\; x_i(0)=0. \tag{6}$$

Here $O_i(k)$ is the (random) neighborhood of sensor i at time k (including i), and $W_{ij}(k)$ are the (random) averaging weights. The local sensor i's decision test at time k is given by:

$$x_i(k) \underset{H_0}{\overset{H_1}{\gtrless}} 0, \tag{7}$$

i.e., H_1 (respectively, H_0) is decided when $x_i(k)\geq 0$ (respectively, $x_i(k)<0$).

Let $x(k) = (x_1(k), x_2(k), ..., x_N(k))^\top$ and $L(k) = (L_1(k), ..., L_N(k))^\top$. Also, collect the averaging weights $W_{ij}(k)$ in $N\times N$ matrix $W(k)$, where, clearly, $W_{ij}(k) = 0$ if the sensors i and j do not communicate at time step k. The algorithm (6) in matrix form becomes:

$$\begin{aligned} x(k) &= W(k)\left(\frac{k-1}{k}x(k-1)+\frac{1}{k}L(k)\right) \\ k &= 1,2,...,\;\; x(0)=0. \end{aligned} \tag{8}$$

We allow the averaging matrices $\{W(k)\}_{k=1}^{\infty}$ to be an i.i.d. sequence, each $W(k)$ to be symmetric and stochastic (row-sums are equal to 1 and the entries are nonnegative,) with probability one and $W(t)$ and $Y(k)$ to be mutually independent over all times k and t.

IV. Main result

In this section, we analyze the performance of the consensus+innovations distributed detector in terms of the detection error exponent, when the size k of the observation interval tends to $+\infty$. Denote by $P_{\mathrm{e},i}(k)$ the error probability at sensor i, with algorithm (6). We have the following result on the error exponent, proof of which is left for the companion journal paper [1].

Theorem 1 Consider the distributed detector in (6). Suppose that the sensors' observations are spatially and temporally i.i.d., conditioned on either hypothesis and that the LMGF Λ_0 is finite on $\mathbb{R}$. Then, at each sensor i, the error exponent is bounded from below as follows:

$$\liminf_{k\to\infty} -\frac{1}{k}\log P_{\mathrm{e},i}(k) \geq \left\{\begin{array}{ll} N C_{\mathrm{ind}}, & \text{if} \;\; |\log r| \geq \mathrm{thr}\,(\Lambda_0, N) \\ -\max\{B_0, B_1\}, & \text{otherwise} \end{array}\right.,$$

where

$$\begin{aligned} &\mathrm{thr}\,(\Lambda_0, N) = \max\{\Lambda_0(N\lambda^\star) - N\Lambda_0(\lambda^\star), \\ &\Lambda_0(1-N(1-\lambda^\star)) - N\Lambda_0(\lambda^\star)\}; \end{aligned} \tag{9}$$

$\lambda^\star$ is the minimizer of Λ_0 over $\mathbb{R}$; $\lambda_0^\star = 1-\lambda_1^\star = \lambda^\star$; $\lambda_l^{\mathrm{SW}}\geq 0$ is the zero of the function[3]

$$\Delta_l(\lambda) := \Lambda_l(N\lambda) - |\log r| - N\Lambda_l(\lambda),\;\; l=0,1;$$

and

$$B_l = \left\{\begin{array}{ll} \Lambda_l(N\lambda_l^{\mathrm{SW}}) - |\log r|, & \text{if} \;\; \frac{\lambda_l^\star}{N} < \lambda_l^{\mathrm{SW}} \leq \lambda_l^\star \\ \Lambda_l(\lambda_l^\star) - |\log r|, & \text{if} \;\; \lambda_l^{\mathrm{SW}} \leq \frac{\lambda_l^\star}{N} \end{array}\right..$$

Moreover, if $|\log r| \geq \mathrm{thr}\,(\Lambda_0, N)$, the distributed detector (6) is asymptotically optimal at each sensor i.

Theorem 1 says that when the speed of consensus $|\log r|$ is above a threshold, the distributed detector in (6) is asymptotically equivalent to the optimal centralized detector; when below the threshold, Theorem 1 says what distributed detector (at least) can achieve.

[3] It can shown that, if $|\log r| < \mathrm{thr}\,(\Lambda_0, N)$, there exists a unique zero of the function $\Delta_l(\lambda)$ on $\mathbb{R}$.

Theorem 1 establishes that to achieve a desired level of detection performance there is a minimum level of connectivity, say $|\log r^\star|$, above which the distributed detection performance cannot improve. Theorem 1 is valuable in the practical design of a sensor network, as it says how much connectivity (resources) one needs to achieve asymptotically optimal detection.

Equation (9) says that the sensor observations distribution (through the LMGF) plays a role in determining the distributed detector performance. We illustrate and explain by an example the effect of the distribution on the distributed detector performance in the next Section.

V. An example: Gaussian versus Laplace distribution

This section illustrates Theorem 1 with the Gaussian and Laplace distributions.

Gaussian distribution. Consider detection of a signal in additive Gaussian noise; $Y_i(t)$ has the following density:

$$f_{\mathrm{G}}(y) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_{\mathrm{G}}} e^{-\frac{(y-m_{\mathrm{G}})^2}{2\sigma_{\mathrm{G}}^2}}, & H_1 \\ \frac{1}{\sqrt{2\pi}\sigma_{\mathrm{G}}} e^{-\frac{y^2}{2\sigma_{\mathrm{G}}^2}}, & H_0. \end{cases} \tag{10}$$

It can be shown that, for this case, the LMGF equals $\Lambda_{0,\mathrm{G}}(\lambda) = -\frac{\lambda(1-\lambda)}{2}\frac{m_{\mathrm{G}}^2}{\sigma_{\mathrm{G}}^2}$, and attains the minimum at $\lambda^\star = 0.5$. The per sensor Chernoff information equals:

$$C_{\mathrm{ind,G}} = \frac{m_{\mathrm{G}}^2}{8\sigma_{\mathrm{G}}^2}.$$

Now, applying Theorem 1, it is easy to get the sufficient condition for optimality of the detector (6) (at each sensor):

$$|\log r| \geq N(N-1)C_{\mathrm{ind,G}}. \tag{11}$$

Laplace distribution. Consider now the case when the sensors' observations have Laplace distribution; the density of $Y_i(t)$ is:

$$f_{\mathrm{L}}(y) = \begin{cases} \frac{1}{2b_{\mathrm{L}}} e^{-\frac{|y-m_{\mathrm{L}}|}{b_{\mathrm{L}}}}, & H_1 \\ \frac{1}{2b_{\mathrm{L}}} e^{-\frac{|y|}{b_{\mathrm{L}}}}, & H_0. \end{cases} \tag{12}$$

The LMGF in this case equals: $\Lambda_{0,\mathrm{L}}(\lambda) = \log\left(\frac{1-\lambda}{1-2\lambda}e^{-\lambda\frac{m_{\mathrm{L}}}{b_{\mathrm{L}}}} - \frac{\lambda}{1-2\lambda}e^{-(1-\lambda)\frac{m_{\mathrm{L}}}{b_{\mathrm{L}}}}\right)$, and it attains its minimum at $\lambda^\star = 0.5$. The per sensor Chernoff information is

$$C_{\mathrm{ind,L}} = \frac{m_{\mathrm{L}}}{2b_{\mathrm{L}}} - \log\left(1 + \frac{m_{\mathrm{L}}}{2b_{\mathrm{L}}}\right).$$

Applying again Theorem 1, the optimality condition for detector (6) becomes:

$$\begin{aligned} |\log r| \quad &\geq \quad \log\left(\frac{2-N}{2-2N}e^{-\frac{N}{2}\frac{m_{\mathrm{L}}}{b_{\mathrm{L}}}}\right. \\ &- \quad \left.\frac{N}{2-2N}e^{-(1-\frac{N}{2})\frac{m_{\mathrm{L}}}{b_{\mathrm{L}}}}\right) \\ &- \quad N\log\left(1+\frac{m_{\mathrm{L}}}{2b_{\mathrm{L}}}\right) + N\frac{m_{\mathrm{L}}}{2b_{\mathrm{L}}}. \end{aligned}$$

We now compare through a numerical example the Gaussian and the Laplace distribution under equal underlying networks (equal r) and equal per sensor Chernoff informations $C_{\mathrm{ind,L}} = C_{\mathrm{ind,G}} = C_{\mathrm{ind}}$. The latter condition ensures that the two corresponding centralized detectors have equal error exponents ($= NC_{\mathrm{ind}}$). We consider a network with $N = 50$ sensors, $C_{\mathrm{ind}} = C_{\mathrm{ind,L}} = C_{\mathrm{ind,G}} = 0.0014$, $b_{\mathrm{L}} = 0.0373$, $m_{\mathrm{L}} = 0.004$, and $m_{\mathrm{G}}^2/\sigma_{\mathrm{G}}^2 = 0.011 = 8C_{\mathrm{ind}}$. We calculate the optimality thresholds for r in (9); they equal $r_{\mathrm{G}} = 0.9667$ ($|\log r_{\mathrm{G}}| = 3.4009$), for the Gaussian case, and $r_{\mathrm{L}} = 0.8613$ ($|\log r_{\mathrm{L}}| = 1.9752$), for the Laplace case. We can see that the optimality thresholds for the Gaussian and Laplace cases are different. Also, the Laplace distribution requires less connectivity (requires smaller $|\log r|$) to achieve asymptotic optimality than the Gaussian distribution[4]. Further, for the range $|\log r| \in [\,|\log r_{\mathrm{L}}|, |\log r_{\mathrm{G}}|\,)$, the distributed detector with Laplace sensors is asymptotically optimal, while the distributed detector with same network infrastructure (equal r), equal per sensor Chernoff information, but Gaussian sensors may not be optimal.

VI. Conclusion

We analyzed the large deviations performance (error exponent) of consensus+innovations distributed detection over random networks. The sensors' observations have generic (non-Gaussian) distribution, i.i.d. both over time and space, with finite LMGF Λ_0 of a sensor's log-likelihood ratio. We showed that the distributed detector exhibits phase transition behavior with respect to the speed of consensus, measured by $|\log r|$, where $r = \lambda_2\left(\mathbb{E}\left[W^2(k)\right]\right)$. When $|\log r|$ is above the threshold, the distributed detector has the same error exponent as the optimal centralized detector. We determined the optimality threshold as a function of Λ_0. When $|\log r|$ is below the threshold, we quantified the achievable performance of the distributed detector. We demonstrated that the optimality threshold depends on the sensor observations' distribution. We illustrated this dependence by comparing Gaussian and Laplace distributions.

[4] We showed in [15] that, for the Gaussian case (10), the threshold in (11) is exact for a certain type of $W(k)$, for the so-termed switching fusion type. That is, $|\log r|$ that is ϵ less than $N(N-1)C_{\mathrm{ind}} = 3.4009$ yields the error exponent strictly less than NC_{ind}. On the other hand, the Laplace distribution needs at most $|\log r_{\mathrm{L}}| = 1.9752$ connectivity, strictly less than for the Gaussian case.

REFERENCES

[1] D. Bajović, D. Jakovetić, J. M. F. Moura, J. Xavier, and B. Sinopoli, "Large deviation analysis of consensus+innovations detection with non-Gaussian observations," October 2011, to be submitted for publication.

[2] S. A. Aldosari and J. M. F. Moura, "Detection in sensor networks: the saddlepoint approximation," *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 327–340, January 2007.

[3] R. Viswanatan and P. R. Varshney, "Decentralized detection with multiple sensors: Part I–fundamentals," *Proc. IEEE*, vol. 85, pp. 54–63, January 1997.

[4] R. S. Blum, S. A. Kassam, and H. V. Poor, "Decentralized detection with multiple sensors: Part II–advanced topics," *Proc. IEEE*, vol. 85, pp. 64–79, January 1997.

[5] J. F. Chamberland and V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE Journal on selected areas in communications*, vol. 22, no. 6, pp. 1007–1015, August 2004.

[6] D. Bajovic, B. Sinopoli, and J. Xavier, "Sensor selection for event detection in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4938–4953, Oct. 2011.

[7] S. Kar, S. A. Aldosari, and J. M. F. Moura, "Topology for distributed inference on graphs," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2609–2613, June 2008.

[8] M. Alanyali, V. .Saligrama, O. Savas, and S. Aeron, "Distributed detection in sensor networks with packet losses and finite capacity links," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4118–4132, July 2004.

[9] D. Bajovic, D. Jakovetic, J. Xavier, B. Sinopoli, and J. M. F. Moura, "Distributed detection over time varying networks: large deviations analysis," in *48th Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2010.

[10] D. Jakovetic, J. M. F. Moura, and J. Xavier, "Distributed detection over noisy networks: large deviations analysis," August 2011, submitted for publication. [Online]. Available: arXiv:1108.1410v1 [cs.IT]

[11] F. S. Cattivelli and A. H. Sayed, "Distributed detection over adaptive networks based on diffusion estimation schemes," in *Proc. IEEE SPAWC '09, 10th IEEE International Workshop on Signal Processing Advances in Wireless Communications*, Perugia, Italy, June 2009, pp. 61–65.

[12] P. Braca, S. Marano, V. Matta, and P. Willet, "Asymptotic optimality of running consensus in testing binary hypothesis," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 814–825, February 2010.

[13] F. Cattivelli and A. Sayed, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 1917–1932, May 2011.

[14] S. Kar, R. Tandon, H. V. Poor, and S. Cui, "Distributed detection in noisy sensor networks," in *Proc. ISIT 2011, Internetional Symposium on Information Theory*, Saint Petersburgh, Russia, August 2011.

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

[16] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*. Boston, MA: Jones and Barlett, 1993.

[17] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: John Wiley and Sons, 1991.