

ASYMPTOTIC PERFORMANCE OF DISTRIBUTED DETECTION OVER RANDOM NETWORKS

Dragana Bajović^{*†}, Dušan Jakovetić^{*†}, João Xavier^{*}

Bruno Sinopoli[†], José M. F. Moura[†]

^{*}Instituto Superior Técnico
Instituto de Sistemas e Robótica
1049-001 Lisboa, Portugal

[†]Carnegie Mellon University
Dept. of Electrical and Computer Engineering
Pittsburgh, PA 15213, USA

ABSTRACT

We show that distributed detection over random networks, or using a random protocol, e.g., of the gossip type, is asymptotically optimal, if the rate of information flow across the random network is large enough. Asymptotic optimality is in the sense of Chernoff information; in other words, we determine when the exponential rate of decay of the error probability for distributed detection is the best possible and equal to the rate of decay of the best centralized detector. The rate of information flow is defined by $|\log r|$, where r is the second largest eigenvalue of the second moment of the random, consensus weight matrix. We quantify interesting tradeoffs in distributed detection, between the rate of information flow and the achievable detection performance.

Index Terms—Chernoff information, distributed detection, random network, running consensus, information flow.

1. INTRODUCTION

We show that distributed detection over random networks is asymptotically equivalent to optimal centralized detection, if the rate of information flow across the random network is large enough. That is, if the rate of information flow is above a threshold, then the exponential rate of decay of the error probability of distributed detection equals the Chernoff information—the best possible rate of the optimal centralized detector. When the random network has slower information flow (asymptotic optimality cannot be achieved,) we find what fraction of the best possible rate of decay of the error probability distributed detection can achieve; hence, we *quantify the tradeoff between network connectivity and achievable detection performance*.

Specifically, we consider the problem where sensors cooperate over a network and sense the environment to decide

^{*}Partially supported by: the Carnegie Mellon|Portugal Program under a grant from the Fundação para a Ciência e Tecnologia (FCT) from Portugal; by FCT grants SIPM PTDC/EEA-ACR/73749/2006; and by ISR/IST plurianual funding (POSC program, FEDER). Dragana Bajović and Dušan Jakovetić hold fellowships from the Carnegie Mellon|Portugal Program.

[†]Work of José M. F. Moura is partially supported by NSF under grants CCF-1011903 and CCF-1018509, and by AFOSR grant FA95501010291.

between two hypothesis. The network is random, varying over time k , or uses a random communication protocol, like gossip. The network connectivity is described by $\{W(k)\}_{k=1}^{\infty}$, the sequence of independent identically distributed (i.i.d.) consensus weight matrices. The sensors' observations are Gaussian, correlated in space, and uncorrelated in time. We consider a stochastic approximation type distributed detection algorithm. At each time k , each sensor: 1) communicates with its single-hop neighbors to compute the weighted average of its own and the neighbors' decision variables; and 2) accounts for its new observation acquired at time k . The network's rate of information flow (i.e., the speed of averaging,) is then measured by $|\log r|$, where r is the second largest eigenvalue of the expected value of $W(k)^2$. We show that, if $|\log r|$ exceeds a threshold, then distributed detection is asymptotically optimal. If $|\log r|$ is below the threshold, we find the achievable rate of decay of the error probability as a function of r .

Several recent references consider different variants of stochastic approximation type distributed algorithms, e.g., references [1, 2] in the context of distributed detection, and reference [3] in the context of distributed estimation. We consider here running consensus, the variant in [2]. Recent work [4, 5] also examines exponential decay rate of error probability (error exponent) for distributed detection. Differently than stochastic approximation type detection, [4] considers tree networks with a fusion node, and [5] considers networks with a fusion node and feedback.

We now contrast this paper to reference [2], which considers a different problem. Reference [2] studies performance of the running consensus detection when the means of the sensors' observations under the two hypothesis become closer, at $1/\sqrt{k}$ rate, and there is an asymptotic, non zero, probability of a miss, and an asymptotic, non zero, probability of false alarm. Within this framework, the running consensus achieves the efficacy [6] of the optimal centralized detector, under a mild assumption on the underlying network being connected on average. In contrast, we assume that the means of the distributions stay fixed with k ; the Bayes error probability exponentially decays to zero, and we examine its rate

of decay. To achieve the optimal decay rate, distributed detection needs an assumption *stronger* than connectedness on average, namely, that the random network or random communication protocol lead to stronger connectivity (or averaging speed as measured by $|\log r|$) than connectedness on average.

In prior work [7], we considered running consensus detection when the underlying network is *deterministically* time varying; we showed that asymptotic optimality holds if the graph that collects the union of links that are online at least once over a finite time window is connected. In contrast, we consider here the case when the underlying network or the communication protocol are *random*, and we establish a sufficient condition for optimality in terms of averaging speed (measured by $|\log r|$.)

Paper organization. Section 2 describes the sensors observations' model and the optimal centralized detector, as if there was a fusion node that collects and processes all sensors' observations. Section 3 describes the random network model and the distributed detection algorithm. Section 4 presents our main results on the asymptotic performance of distributed detection. Finally, Section 5 concludes the paper.

2. CENTRALIZED DETECTION

Subsection 2.1 describes the sensors observations model; Subsection 2.2 describes the (asymptotically) optimal centralized detector; and Subsection 2.3 studies the asymptotic performance of this detector.

2.1. Sensor observations model

We assume that N sensors are deployed to sense the environment and to decide between the two hypotheses, H_1 and H_0 . Each sensor i measures a scalar quantity $y_i(k)$ at each time step k ; for simplicity, we assume that all sensors measure synchronously at $k = 1, 2, \dots$. Collect all sensors observations at time k , $y_i(k)$, $i = 1, \dots, N$, into the $N \times 1$ vector $y(k) = (y_1(k), \dots, y_N(k))^T$, that we assume:

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

The quantity m_l is the constant signal; the quantity $\zeta(k)$ is a zero-mean, Gaussian, spatially correlated noise, i.i.d. across time, with distribution $\zeta(k) \sim \mathcal{N}(0, S)$, where S is a positive definite covariance matrix. The noise can be spatially correlated, e.g., due to dense deployment of the sensor network. Denote also by f_l the probability density function of $y(k)$ under H_l , $l = 0, 1$.

2.2. Centralized detector: Log-likelihood ratio test

The optimal centralized detector forms the log-likelihood (LLR) decision test:

$$\frac{1}{k} \sum_{j=1}^k L(j) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_k, \quad (2)$$

where $L(k) = \log(f_1(y(k))/f_0(y(k)))$ is the LLR and γ_k is

the test threshold. Under the Gaussian observations in Subsection 2.1, $L(k)$ is the linear sufficient statistic: $L(k) = (m_1 - m_0)^T S^{-1} (y(k) - \frac{m_1 + m_0}{2})$. Define $v \in \mathbb{R}^N$:

$$v := S^{-1}(m_1 - m_0). \quad (3)$$

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

$$L(k) = \sum_{i=1}^N v_i \left(y_i(k) - \frac{[m_1]_i + [m_0]_i}{2} \right) = \sum_{i=1}^N \eta_i(k) \quad (4)$$

The LLR at time k is separable, i.e., it is the sum of the $\eta_i(k)$ that depend affinely on the individual observations $y_i(k)$. We will exploit this fact in Section 3 to derive the distributed, running consensus, detection algorithm.

2.3. Asymptotic performance of centralized detection

We are interested in the Bayes probability of error $P_{\text{cen}}^e(k)$ after k samples are processed by the centralized detector (2). Specifically, we are interested in the exponential rate of decay of $P_{\text{cen}}^e(k)$ as $k \rightarrow \infty$, i.e., $\lim_{k \rightarrow \infty} -\frac{1}{k} \log P_{\text{cen}}^e(k)$. The Chernoff lemma (see, e.g., [8]) states that the best achievable rate, among all possible decision tests, is attained by the LLR test with $\gamma_k = 0$, for all k , and this rate equals the Chernoff information (see, e.g., [8]), which is, for the Gaussian case studied here, given by:

$$\mathbf{C} = \mathbf{C}_{\text{tot}} = \frac{1}{8} (m_1 - m_0)^T S^{-1} (m_1 - m_0). \quad (5)$$

In eqn. (5), the subscript tot designates the total Chernoff information of the network, i.e., the Chernoff information for the observations collected from all sensors. Specifically, if the sensor observations are uncorrelated (the noise covariance matrix $S = \text{Diag}(\sigma_1^2, \dots, \sigma_N^2)$), then $\mathbf{C}_{\text{tot}} = \sum_{i=1}^N \frac{[m_1 - m_0]_i^2}{8\sigma_i^2} = \sum_{i=1}^N \mathbf{C}_i$, where \mathbf{C}_i is the Chernoff information of the individual sensor i . That is, \mathbf{C}_i equals the best achievable rate of the Bayes error probability, as if sensor i worked as an independent detector and did not cooperate with other sensors. In this paper, we find the exponential decay rate for the error probability for the running consensus distributed detector. Moreover, we find conditions on the network connectivity when this rate equals \mathbf{C}_{tot} . When the exponential decay rate of the error probability for distributed detector equals \mathbf{C}_{tot} at each sensor, we say that distributed detection is *asymptotically optimal*.

3. DISTRIBUTED DETECTION

We consider distributed detection, under the same assumptions on the sensor observations as in Subsection 2.1; now, the fusion node is no longer available, and the sensors cooperate through a randomly varying network. A similar analysis will hold if we consider a random communication protocol like gossip. We study running consensus distributed detection, proposed in [2], extending it to spatially correlated observations. Recall the definition of the vector v in eqn. (3)

and the scalar $\eta_i(k)$ in eqn. (4). The update of $x_i(k)$, the local decision variable of sensor i at time k , is then as follows:

$$x_i(k+1) = \frac{k}{k+1} W_{ii}(k) x_i(k) + \frac{k}{k+1} \sum_{j \in \Omega_i(k)} W_{ij}(k) x_j(k) + \frac{1}{k+1} \eta_i(k+1), \quad (6)$$

for $k = 1, 2, \dots$, and $x_i(1) = \eta_i(1)$. Here $\Omega_i(k)$ is the (random) neighborhood of sensor i at time k , and $W_{ij}(k)$ are the (random) averaging weights. Hence, at each k , sensor i averages its current decision variable with the neighbors' decision variables, and adds $\eta_i(k+1)$ to account for new observation $y_i(k+1)$. The local sensor i 's decision test at time k is as follows: if $x_i(k) > 0$ (respectively, $x_i(k) \leq 0$), then H_1 (respectively, H_0) is accepted. Let $x(k) = (x_1(k), x_2(k), \dots, x_N(k))^T$ and $\eta(k) = (\eta_1(k), \dots, \eta_N(k))^T$. Also, collect the averaging weights $W_{ij}(k)$ in the $N \times N$ matrix $W(k)$, where $W_{ij}(k) = 0$ if i and j do not communicate at time step k . The algorithm in matrix form becomes:

$$x(k+1) = \frac{k}{k+1} W(k) x(k) + \frac{1}{k+1} \eta(k+1), \quad (7)$$

for $k = 1, 2, \dots$, and $x(1) = \eta(1)$. We allow the averaging matrices $W(k)$ (and hence, the network connectivity) to be random. We next summarize the assumptions on $W(k)$. First, let $J := (1/N)(11^T)$ and $\widetilde{W}(k) := W(k) - J$. Here, 1 is the $N \times 1$ vector with unit entries.

Assumption 1 1) The sequence $\{W(k)\}_{k=1}^\infty$ is i.i.d.; 2) $W(k)$ are symmetric and stochastic (row-sums are equal to 1 and the entries are nonnegative), with prob. one; and 3) $W(l)$ and $y(s)$ are independent, $\forall l, \forall s$.

In section 4, we examine what (additional) conditions the matrices $W(k)$ have to satisfy, to achieve asymptotic optimality of the distributed detection algorithm. Recall the definition of the $N \times 1$ vector v in (3). The sequence of $N \times 1$ random vectors $\{\eta(k)\}$, conditioned on H_l , is i.i.d. The vector $\eta(k)$ (under hypothesis H_l , $l = 0, 1$) is Gaussian with mean $m_\eta^{(l)}$ and covariance S^η :

$$m_\eta^{(l)} = (-1)^{(l+1)} \text{Diag}(v) \frac{1}{2} (m_1 - m_0) \quad (8)$$

$$S^\eta = \text{Diag}(v) S \text{Diag}(v). \quad (9)$$

Here, $\text{Diag}(v)$ is a diagonal matrix with the diagonal v .

4. DISTRIBUTED DETECTION: CONDITIONS FOR ASYMPTOTIC OPTIMALITY

This section provides necessary and sufficient conditions for asymptotic optimality for distributed detection. When distributed detection is not guaranteed to be optimal, this section

finds a lower bound on the exponential decay rate of the error probability, in terms of network connectivity and observations' distributions. We start by finding a sufficient condition for optimality and evaluating the lower bound on the decay rate of the error probability.

4.1. Sufficient condition

Recall that $r := \lambda_2(\mathbb{E}[W(k)^2]) = \|\mathbb{E}[W(k)^2] - J\|$, where $\lambda_2(\cdot)$ and $\|\cdot\|$ are, respectively, the second largest eigenvalue and the spectral norm. It is well known that r measures the speed of averaging across network, like with standard consensus and gossip, e.g., [9]. (The smaller r is, the faster averaging is.) Theorem 2 shows that distributed detection is asymptotically optimal if r is small enough. The Theorem finds a lower bound on the rate of decay of the error probability, even when the sufficient condition for asymptotic optimality does not hold. Denote by $m_L^{(l)}$ and σ_L^2 , respectively, the mean and the variance of the LLR in eqn. (4) under H_l , $l = 0, 1$; it can be shown that $m_L^{(l)} = \frac{(-1)^{l+1}}{2} (m_1 - m_0)^T S^{-1} (m_1 - m_0)$ and $\sigma_L^2 = (m_1 - m_0)^T S^{-1} (m_1 - m_0)$. Denote by $\overline{m}^{(l)} = \max_{i=1, \dots, N} |[m_\eta^{(l)}]_i| = \overline{m}$, i.e., the maximum-in-modulus of the entries of $m_\eta^{(l)}$. The proof of the Theorem 2 is in [10].

Theorem 2 Let Assumption 1 hold and consider the distributed detection algorithm in eqn. (6). Then, for the exponential decay rate of the error probability, we have:

$$\liminf_{k \rightarrow \infty} \frac{1}{k} \log P_{i, \text{dis}}^e(k) \geq \begin{cases} \mathbf{C}_{\text{tot}} & \text{if } |\log r| \geq \frac{1}{8} N^2 (1 + (1 - \frac{1}{N})K) \|S^\eta\| \\ \mathbf{C}_{\text{dis}} & \text{otherwise, where} \end{cases} \quad (10)$$

$$\mathbf{C}_{\text{dis}} = - \left(\frac{1}{2N^2} \sigma_L^2 \overline{\mu}^2 + \frac{1}{N} m_L^{(0)} \overline{\mu} \right), \quad (11)$$

$$\overline{\mu} = \begin{cases} \frac{1}{4} \frac{K}{K+1} + \frac{1}{4} \frac{\sqrt{K^2 + \frac{32|\log r|}{\|S^\eta\|} (1+K)}}{K+1}, \\ \text{if } \frac{1}{8} \|S^\eta\| < |\log r| < \frac{1}{8} N^2 (1 + (1 - \frac{1}{N})K) \|S^\eta\|; \\ \frac{1}{4} \sqrt{K^2 + \frac{32|\log r|}{\|S^\eta\|}} - \frac{1}{4} K, \\ \text{if } |\log r| \leq \frac{1}{8} \|S^\eta\|. \end{cases}$$

Here $K = (8\overline{m}) / \|S^\eta\|$. Moreover, each sensor i is asymptotically optimal, i.e., $\lim_{k \rightarrow \infty} -\frac{1}{k} \log P_{i, \text{dis}}^e(k) = \mathbf{C}_{\text{tot}}$, $\forall i$, provided that:

$$|\log r| \geq \frac{1}{8} N^2 \left(1 + (1 - \frac{1}{N})K \right) \|S^\eta\|. \quad (12)$$

Theorem 2 says that the network connectivity should be large enough (i.e., r -small enough,) in order to achieve asymptotic optimality. Specifically, there exists a threshold on $|\log r|$, above which asymptotic optimality is achieved. Consider further (12) and, for simplicity, assume that the sensor observations are spatially uncorrelated and that the individual Chernoff information at each sensor is equal, i.e., $\mathbf{C}_i = \mathbf{C}_j$, $i \neq j$.

Then, it can be shown that the sufficient condition for asymptotic optimality becomes: $\frac{|\log r|}{3N-1} \geq C_{\text{tot}} = NC_i$. The quantity $\frac{|\log r|}{3N-1}$ depends solely on $W(k)$, i.e., the network connectivity, and may be interpreted as the “information capacity” of the network. If the information capacity is greater than the total Chernoff information C_{tot} offered by the sensors’ observations, then asymptotic optimality is achieved. Further, $r = 0$ corresponds to the fully connected network ($W(k) \equiv J$), and, hence, to the optimal centralized detector. In this case, the capacity is $+\infty$, and asymptotic optimality is achieved for any value of C_{tot} . Figure 1 plots the lower bound in eqn. (10) for a network with $N = 20$ sensors, spatially uncorrelated observations, C_i equal for all sensors, and $C_{\text{tot}} = NC_i = 0.0047$, illustrating what fraction of C_{tot} is achievable for given network connectivity (for given r .)

4.2. Necessary condition

We have shown that a sufficient condition for asymptotic optimality is that the network connectivity (measured by $|\log r|$) exceeds a certain threshold that depends on the Chernoff information. We now show that, for asymptotic optimality, it is also necessary that certain measure of the network connectivity exceeds a Chernoff-information dependent threshold. Denote by $E_i(k)$ the event that, at time k , the sensor i is connected to at least one of the remaining sensors in the network; that is, $E_i(k) := \{\max_{j=1, \dots, N, j \neq i} W_{ij}(k) > 0\}$. Further, denote by $P_i(k) = P_i = \mathbb{P}(E_i(k))$. We state the Theorem (Theorem 3) for spatially uncorrelated sensors’ observations, the proof of which is in [10].

Theorem 3 (Asymptotic optimality: Necessary condition)

Consider the distributed detection algorithm in eqn. (6), let Assumption 1 hold, and let the sensor observations be spatially uncorrelated, i.e., $S = \text{Diag}(\sigma_1^2, \dots, \sigma_N^2)$. Then, a necessary condition for the asymptotic optimality of distributed detection at sensor i is: $|\log(1 - P_i)| > C_{\text{tot}} - C_i$.

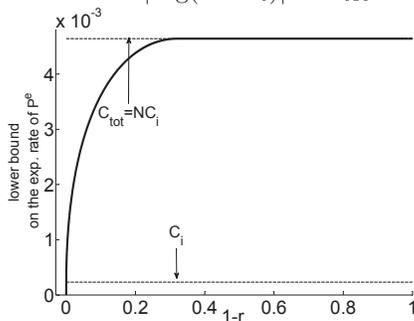


Fig. 1. Lower bound on the exponential decay rate of the error probability (eqn. (10)) for a network with $N = 20$ sensors, spatially uncorrelated observations, and $C_{\text{tot}} = NC_i = 0.0047$. The rate is optimal (C_{tot}) for $(1 - r) \geq 0.29$.

5. CONCLUSION

We studied distributed detection over random networks when at each time step k each sensor: 1) averages its decision vari-

able with the neighbors’ decision variables; and 2) accounts on-the-fly for its new observation. We found necessary and sufficient conditions for asymptotic optimality of distributed detection, in the sense of the exponential decay rate of the error probability, in terms of the network connectivity and the Chernoff information of the sensors’ observations. When the sufficient condition for optimality is not met, we find a lower bound on this rate, as a function of the network connectivity and the Chernoff information.

6. REFERENCES

- [1] F. Cattivelli and A. H. Sayed, “Diffusion LMS-based detection over adaptive networks,” in *Asilomar Conf. on Sig., Sys. and Comp.*, Pacific Grove, Oct. 2009.
- [2] P. Braca, S. Marano, V. Matta, and P. Willet, “Asymptotic optimality of running consensus in testing binary hypotheses,” *IEEE Trans. Sig. Process.*, vol. 58, no. 2, pp. 814–825, Feb. 2010.
- [3] S. Kar, J. M. F. Moura, and K. Ramanan, “Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication,” Submitted, available at arxiv.org/abs/0809.0009, Aug. 2008.
- [4] W. P. Tay and J. N. Tsitsiklis, *Error Exponents for Decentralized Detection in Tree Networks*, Networked Sensing Information and Control, V. Saligrama (Ed.), Springer-Verlag, 2008.
- [5] O. P. Kreidl, J. N. Tsitsiklis, and S. I. Zoumpoulis, “Decentralized detection in sensor network architectures with feedback,” in *48th Allerton Conf. on Comm., Control and Computing*, Monticello, IL, Sep. 2010, pp. 1605–1609.
- [6] S. A. Kassam, *Signal Detection in Non-Gaussian Noise*, Springer-Verlag, New York, 1987.
- [7] 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 Conf. on Comm., Control, and Computing*, Monticello, IL, Sep. 2010.
- [8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, New York, 1991.
- [9] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, “Randomized gossip algorithms,” *IEEE Trans. Info. Theory*, vol. 52, no. 6, pp. 2508–2530, June 2006.
- [10] D. Bajovic, D. Jakovetic, J. Xavier, B. Sinopoli, and J. M. F. Moura, “Distributed detection over random networks: Large deviations performance analysis,” Submitted to the *IEEE Trans. Sig. Process.*, available at: <http://arxiv.org/abs/1012.4668>, Dec. 2010.