Running Consensus for Decentralized Detection

Paolo Braca

Coordinatore: Prof. Maurizio Longo
Tutor: Prof. Stefano Marano

Anno Accademico 2009-2010, VIII Ciclo – Nuova Serie
Facoltà di Ingegneria  
Dipartimento di Ingegneria dell’Informazione e Ingegneria Elettrica

Dottorato di Ricerca in Ingegneria dell’Informazione  
VIII Ciclo – Nuova Serie

TESI DI DOTTORATO

Running Consensus for Decentralized Detection

CANDIDATO: PAOLO BRACA

COORDINATORE: PROF. MAURIZIO LONGO

TUTOR: PROF. STEFANO MARANO

Anno Accademico 2009 – 2010
# Contents

1 Introduction 5

2 Mathematical Properties of Running Consensus 9
   2.1 Basic Procedure 9
   2.2 The Ideal Centralized System 10
   2.3 Some Bounds 13
   2.4 Binary Hypothesis Test by Running Consensus 15
      2.4.1 Fixed Sample Size Test 16
      2.4.2 Sequential Test 17

3 Sequential Estimation 21
   3.1 Parameter Estimation 21
   3.2 Performance Metrics 23
      3.2.1 Consensus in the Network 23
      3.2.2 Comparison with the Ideal Centralized System 25
   3.3 Asymptotic Optimality 25
   3.4 Examples: Pairwise Averaging 28

4 Sequential Detection 33
   4.1 Centralized Hypothesis Testing 33
   4.2 Hypothesis Testing in Sensor Networks with Running Consensus 34
   4.3 Asymptotic Performances 37
      4.3.1 Fixed Sample Size Test 37
      4.3.2 Locally Optimum Sequential Detection 40
   4.4 Examples and Numerical Experiments 45
4.4.1 Neyman-Pearson FSS Tests . . . . . . . . . 47
4.4.2 Sequential Tests . . . . . . . . . . . . . . . . 48
  Gaussian example . . . . . . . . . . . . . . . . . 48
  A Non-Gaussian Example . . . . . . . . . . . . . 53
4.5 Some Practical Issues . . . . . . . . . . . . . . . . 55

5 Quickest Change Detection 61
  5.1 Problem Formalization: Centralized Page’s Test . . 61
  5.2 Decentralized Page’s Test by Running Consensus . .66
  5.3 Bank of Parallel Page’s Detectors . . . . . . . . . . .72
  5.4 Relative Efficiencies . . . . . . . . . . . . . . . . . . 75
  5.5 Examples and Numerical Experiments . . . . . . . .77

6 Conclusions 85

A 89
  A.1 Proof of Proposition 2 . . . . . . . . . . . . . . . . . 89
  A.2 Proof of Theorem 1 . . . . . . . . . . . . . . . . . . 93

B 97
  B.1 Proof of Theorem 2 . . . . . . . . . . . . . . . . . . 97
  B.2 Proof of Theorem 3 . . . . . . . . . . . . . . . . . . 98
Preface

This thesis represents a culmination of work and learning that has taken place over a period of almost three years (2007 - 2010) at the University of Salerno, and at the University of Connecticut (2009). It is mostly an unified mathematical dissertation of the running consensus procedures \([1, 6]\), which has been applied to the problem fields of sequential estimation (see \([1, 2]\) and Chapter 3), sequential/non-sequential detection (see \([3, 4]\) and Chapter 4), and change detection (see \([5, 6]\) and Chapter 5).

In the recent years, the detection using the paradigm of the running consensus has been recognized as one of the three possible classes of distributed detection (e.g. see \([7–10]\)) in which the phases of sensing and communication need not be mutually exclusive, i.e., sensing and communication occur simultaneously.

Considering that the running consensus paradigm is just an intuitive inference procedure, i.e. suboptimal w.r.t. an ideal centralized system scheme which is optimal, the most important result is that it asymptotically reaches the performance of this ideal scheme. There are two asymptotic frameworks. In the first one the running consensus is locally efficient \([11, 12]\) as the centralized procedure, see \([4]\) and Chapter 4. The limit is in the number of observations (which is proportional to the time duration of the algorithm), and while this number diverges the two statistical hypotheses are closer and closer. In the second framework, that of large deviations \([13]\), while the procedure time duration diverges the two statistical hypotheses are fixed and it is studied the convergence rate of the error probability. Interestingly, this rate can be equal or below that of the ideal system depending on the connectivity of the network \([7, 10]\).
Recently, the running consensus has been also extended and generalized in order to cover more general problems, related for instance to the noisy networks [14].

Beyond the detection problems some connections with the running consensus have been highlighted w.r.t the so-called “consensus+innovations” distributed inference procedures, e.g. see [15]. One of the last consensus+innovations distributed algorithms in developed in [16], and faces the multitarget tracking problem, where the number of targets is unknown and time-varying, generalizing the work in [17], where just a single target has been considered. In [16] the multisensor Probability Hypothesis Density (PHD) filter is approximately computed in a distributed fashion using the consensus paradigm. It is important to note that the “optimality” in a consensus tracking problem cannot be reached as in [4,7,10] just waiting a sufficiently large time because the state of the nature (targets’ positions, velocities, etc.) evolves in time instead to be fixed. In this case only increasing the number of sensors allows to obtain an asymptotic optimality property [18,19].

Paolo Braca

May, 2013,
La Spezia, Italy.
Acknowledgments

It is anything but easy to thank all the people who contributed to my personal and professional development during my Ph.D. studies. More than anyone else, however, my Tutor Stefano Marano and Vincenzo Matta have been the ones who made me think research was what I really wanted to do. I learned a lot, not only on a professional plane, and I owe to them most of what I know about doing research. A special thanks goes to Peter Willett, he allowed me to be exposed to a stimulating research environment during my stay at the University of Connecticut.

I should have never started my Ph.D. studies without the support of my Department (DIIIE), and in particular its director Maurizio Longo.

Besides the above, my colleagues Marco Guerriero, Gianluca Gennarelli, Domenico Attianese, Paolo Addesso have also been good friends and wonderful people to work with.

I am grateful to friends inside and outside the professional environment, people I simply shared good moments and laughter with: Antonio De Costanza, Igor Negri, Giuseppe Zambrano, Antonio Fortunato, Domenico De Pascale, Christian Berger, Vishal Cholapadi Ravindra, Fabio Postiglione, Rocco Restaino, Roberto Conte, Fabio Mazzarella, Luigi Bruno, Pierpaolo D’Agostino, Pasquale Pistillo, Michele Grimaldi, CoRiTeL boys, Ozgur Erdinc, Ramona Georgescu, Bill Donat, Naomi Thonakkaraparayil, Sumit Narayan, Xin Tian, Javier and Carol Areta.

More than anyone else, however, my family has been fundamental. Even when we had to face problems way more important than anything I could write in this thesis, I never missed your un-
conditional support to my professional development. Eventually, I hope to have a chance to return what you gave me.

Last but not least, a very special thanks to Loredana, without her love, encouragement and support, I would not have finished this Thesis.

Thank you.

Paolo

March, 2010,
Salerno, Italy.
Chapter 1

Introduction

Recent advances in hardware technology have led to the emergence of small, low-power, and possibly mobile devices with limited on-board processing and wireless communication capabilities. Typically, these devices, called sensors, consist of a radio frequency circuit, a low-power digital signal processor, a sensing unit, and a battery. Due to their low cost and low complexity design requirement, individual sensors can only perform simple local computation and communicate over a short range at low data rates. But when deployed in a large number across a spatial domain, these primitive sensors can form an intelligent network to monitor the physical environment with high performance.

Sensor networks are suited for situation awareness applications such as environmental monitoring (air, water, and soil), biomedical engineering, home applications, radar, smart factory instrumentation, military surveillance, precision agriculture, space exploration, and intelligent transportation.

On the other side, the design of sensor network poses new challenges and requires multifaceted, interdisciplinary, and cross-layer approaches. In particular it becomes of paramount importance to develop energy and bandwidth efficient signal processing algorithms that can be implemented in a fully distributed manner. Distributed signal processing in a wireless sensor network differs from the traditional signal processing framework in several impor-
1. Introduction

Sensor measurements are collected in a distributed fashion across a network, and this needs an appropriate data sharing among the sensors to save the energy and bandwidth, that are critical resources for this kind of system. Indeed, in wireless sensor networks applications, the typical goal is to study inference problems with bandwidth and energy constraints. The sensor measurements are appropriately transformed and transmitted to a fusion center, that can be static or mobile, see sensor network with mobile access (SENMA) concept. The research has been focused on finding good quantization at sensors and suitable fusion strategies. A very small list of papers would include Achievable performance, in terms of estimation error as a function of the data rate is of particular interest, and in the Gaussian case this has become known as the CEO problem for which the answers are both simple and elegant. The distributed target tracking problem with power-limited sensors is studied in , while the static target estimation problem with data association and bandwidth constraint is studied in .

In a wireless sensor network, sensors may enter or leave the network dynamically, resulting in unpredictable changes in network size and topology. Sensors may disappear permanently either due to damage to the nodes or drained batteries. Temporarily due to topology, traffic, and channel conditions. This makes it necessary for distributed signal processing algorithms to be robust to the changes in network topology. This motivates us to consider fully decentralized sensor network without fusion center, where the remote units sense the environment and collect data, but, due to the lack of the fusion center, they are also programmed to run a consensus protocol aimed at corroborating the local measurements with observations made at the neighboring nodes. The process of data exchange updates the locally computed statistics and (asymptotically) leads to the agreement about a common value, shared by all the nodes, that represents the final statistic.

The “consensus” problem is studied from the ’70s in the social sciences, in which the behavior of some people that try to reach consensus is mathematically modeled. Then this math-
ematical framework was transferred in the engineer field to design
the “behavior” (cooperation and flocking) of artificial agents, for
example sensors or mobile vehicles. A consensus algorithm is, in
general, the set of cooperation rules in a network of agents.

The consensus scheme for the computation of statistic functions has also attracted much interest in the last years, especially
for its properties of robustness and scalability. We mention as
useful entry points to the topical literature [17, 37–49].

In this thesis we propose an innovative modification of classical
consensus protocols, which we call running consensus, see [1–5].
While in the classical scheme, first the environmental sensing stage
is performed, and only successively the acquired data are ex-
changed among nodes to reach the consensus, in the running con-
sensus paradigm, this limitation is relaxed and the data acquisition
proceeds simultaneously with data exchange. In this way, there
is no need to fix in advance when the environmental monitoring
stage is to be terminated, which may represent a drawback in ap-
lication where the system operates in dangerous environments.
Indeed, with running consensus, even though many nodes of the
system are abruptly destroyed or impaired, the information they
collected up to that time still contributes to the final statistic being
shared across the system on-the-fly.

Both in the classical scheme and its modification, one would
like that the final statistic achieved asymptotically be the same
statistic that an ideal centralized entity would compute, provided
that all the measurements collected by the entire sensor network
are available to it. As we will see, with the consensus protocols,
the state of each node (namely, the locally computed statistic)
approaches asymptotically the state of the ideal centralized sys-
tem, as the number of consensus steps (data exchanges) diverges.
However, the exponentially fast convergence that the classical con-
sensus algorithms usually exhibits is lost and the convergence, ex-
hibited by the running consensus, is much slower due to the pres-
ence of the new measurements made at each step, that must be
incorporated on-the-fly into the statistics.

The above consideration raises the following question. Suppose
that the ideal centralized statistic is asymptotically optimal, in the sense of achieving unbeatable performance as the number of overall measurements grows. Is such asymptotic optimality retained by a sensor network implementing the running consensus algorithm? The main contribution of this thesis is to answer (affirmatively) the above question. In more generality, we show under what conditions the statistic computed by running consensus attains the same asymptotic performance as an ideal centralized system.

The remainder of this work is so organized. In chapter 2 the mathematical properties of the running consensus are developed, with the related physical insights deferred to chapters. In chapter 3 the sequential estimation problem, implemented via the running consensus is considered [1,2]. Analytical performances are characterized in terms of bounds, and it is shown that each sensor reaches asymptotically the consensus and the ideal estimation performance. In chapter 4 the detection problem is considered using the running consensus [3,4]. Here the sensors are cooperating by the running consensus scheme to discriminate among two simple hypotheses. First, we consider the Neyman-Pearson setup in which the number of measurements used for the decision is fixed, and for this reason this is also called the FSS (fixed sample size) setup. Then, we focus on the the sequential case in which there is a virtually unbounded number of data available, and an appropriate number of these is used to make the final decision according to prescribed error probabilities. In this case, the asymptotic scenario is essentially that where the average number of samples needed for a decision tends to infinity. In chapter 5 the quickest change detection problem is addressed [5]. The problem is to detect an event (that is modeled by a change in the data distribution) as soon as possible, see also [50]. Approximate performance evaluation is considered and the running consensus performance is close to the ideal system. The comparison with a bank of parallel Page’s test is also provided.
Chapter 2

Mathematical Properties of Running Consensus

In this chapter the mathematical properties of the running consensus are developed and the main properties are formally proved. In Section 2.1 we introduce the running consensus procedure. In Section 2.2 we define the performance benchmark for the running consensus: an ideal centralized system that observes all the data shared in the network. In Section 2.3 upper and lower bounds on the performance metrics are developed, then the asymptotic optimality for hypothesis testing problems is proved in Section 2.4.

2.1 Basic Procedure

Let us consider a slotted system where in the n\textsuperscript{th} time slot each sensor of the network collects a new measurement, and these measurements are modeled as iid (independent, identically distributed) random variables. The collected random values are stored for successive processing; the value stored by a node is hereafter also referred to as the state of that sensor. We denote by \( \mathbf{x}_n \) the column vector whose entries are the measurements collected by the sensors at time \( n \). Its size is \( M \), the number of nodes in the network.

Let \( \mathbf{s}_n \) be the \( M \)-vector whose entries represent the state of the
nodes at the $n^{th}$ time slot. We assume that this (random) state vector is updated as follows:

$$
\begin{pmatrix}
  s_{n,1} \\
  s_{n,2} \\
  \vdots \\
  s_{n,M}
\end{pmatrix} = W_n \alpha_n \begin{pmatrix}
  s_{n-1,1} \\
  s_{n-1,2} \\
  \vdots \\
  s_{n-1,M}
\end{pmatrix} + W_n \beta_n \begin{pmatrix}
  t(x_{n,1}) \\
  t(x_{n,2}) \\
  \vdots \\
  t(x_{n,M})
\end{pmatrix}, \quad (2.1)
$$

the above can be cast in a compact vector form as

$$s_n = W_n (\alpha_n s_{n-1} + \beta_n t(x_n)), \quad (2.2)$$

note that $t(x)$ is a real function and $\alpha_n, \beta_n$ are deterministic time-varying weights. Specifically, it is assumed that (i) data are acquired by sensors at time instants $n = 1, 2, \ldots$, and (ii) for each $n$, a randomly chosen subset of nodes is allowed to exchange data, according to the so-called gossip algorithm [42]. The matrices $W_n, n = 1, 2, \ldots$, are assumed iid and doubly stochastic, so that the connection protocol among nodes (formalized by the product $W_n s_{n-1}$ in the above equation) amounts to a weighted average of the nodes’ states [42]. The $W_n$’s are also assumed statistically independent of the sensors’ observations.

The basic assumption made throughout this work is that the statistical average $E[W_n W_n^T]$ (which is obviously doubly stochastic as well) has unitary eigenvalue with algebraic multiplicity 1, see also [42]. Due to the identical distribution with respect to the time slot $n$, to simplify the notation, in the following we write $E[W W^T]$ for $E[W_n W_n^T]$.

### 2.2 The Ideal Centralized System

An ideal system to which all the collected observations would be made available is able to compute the following quantity:

$$s_n^{(c)} = \chi_n \sum_{i=1}^{n} 1^T t(x_i), \quad (2.3)$$
2.2. The Ideal Centralized System

where the suffix \( c \) emphasizes that this is the *ideal centralized system* state and \( \mathbf{1} \) is the vector of all ones. The choice of \( \chi_n \) and \( t(x) \) depends on the application. Now if we are interested to estimate \( \mu = E[x_{i,j}] \) then, choosing \( \chi_n = 1/(nM) \) and \( t(x) = x \), the ideal centralized estimator is the simple arithmetic mean:

\[
\hat{\mu}_n \overset{\text{def}}{=} s_n^{(c)}, \quad \chi_n = \frac{1}{nM}, \quad t(x) = x; \quad (2.4)
\]

If we are interested to compute the random walk \( \sum_{i=1}^n \sum_{j=1}^M t(x_{i,j}) \) (which will play when we introduce the sequential tests), then we choose \( \chi_n = 1 \).

Now our interest is to analyze the differences between the ideal centralized system state and the running consensus state. Let us define \( \Phi_{n,i} \) and \( \tilde{\Phi}_{n,i} \) as the product of \( W_n \) and \( \tilde{W}_n \overset{\text{def}}{=} W_n - \mathbf{1}\mathbf{1}^T/M \), respectively, i.e.:

\[
\Phi_{n,i} \overset{\text{def}}{=} \begin{cases} W_n & \text{if } i = n, \\ W_n W_{n-1} \ldots W_i & \text{if } i < n; \end{cases} \quad (2.5)
\]

\[
\tilde{\Phi}_{n,i} \overset{\text{def}}{=} \begin{cases} \tilde{W}_n & \text{if } i = n, \\ \tilde{W}_n \tilde{W}_{n-1} \ldots \tilde{W}_i & \text{if } i < n. \end{cases} \quad (2.6)
\]

The properties of \( \Phi_{n,i} \) and \( \tilde{\Phi}_{n,i} \) are investigated in [42], here we stress that \( \Phi_{n,i} \) is doubly stochastic, being the product of doubly stochastic matrices, this implies from the above definitions

\[
\tilde{\Phi}_{n,i} = \Phi_{n,i} - \frac{\mathbf{1}\mathbf{1}^T}{nM}. \quad (2.7)
\]

To quantify the difference between the sensor state \( s_{n,j} \) and the centralized state, we introduce the *running consensus error* [24].

**Definition 1.** The running consensus error is defined as

\[
e_n \overset{\text{def}}{=} \beta_n \sum_{i=1}^n \tilde{\Phi}_{n,i} t(x_i). \quad (2.8)
\]
2. Mathematical Properties of Running Consensus

The following proposition remarks the relationship between the running consensus state (2.1) and the ideal centralized system state (2.3).

**Proposition 1.**

1. *s* \(_n\) is a linear combination of *t*(x) and \(\Phi_{n,i}\):
   
   for \(\alpha_n = (n - 1)/n\) and \(\beta_n = 1/n\) we have
   
   \[ s_n = \frac{1}{n} \sum_{i=1}^{n} \Phi_{n,i} t(x_i); \quad (2.9) \]
   
   for \(\alpha_n = 1\) and \(\beta_n = M\) we have
   
   \[ s_n = M \sum_{i=1}^{n} \Phi_{n,i} t(x_i); \quad (2.10) \]

2. the state vector \(s_n\) is the sum of the ideal centralized system state \(s_n^{(c)}\) and the error term \(e_n\)
   
   \[ s_n = s_n^{(c)} \mathbf{1} + e_n; \quad (2.11) \]
   
   with
   
   \[ \chi_n = \begin{cases} \frac{1}{nM} & \alpha_n = \frac{n-1}{n}, \beta_n = \frac{1}{n}; \\
   \quad 1 & \alpha_n = 1, \beta_n = M \end{cases} \quad (2.12) \]

3. The expected value of \(s_n\) is:

   \[ E[s_n] = \begin{cases} 1 \mu & \alpha_n = \frac{n-1}{n}, \beta_n = \frac{1}{n}; \\
   1 n M \mu & \alpha_n = 1, \beta_n = M \end{cases} \quad (2.13) \]

   and consequently the running consensus error is a zero-mean random variable:

   \[ E[e_n] = 0. \quad (2.14) \]
Proof. Equations (2.9)-(2.10) follow immediately from the (2.1) with the particular choose of \( \alpha_n \) and \( \beta_n \). Suppose \( \alpha_n = (n - 1)/n \) and \( \beta_n = 1/n \) (the proof for \( \alpha_n = 1 \) and \( \beta_n = M \) is similar), from the equation (2.9) we have \( s_n = \frac{11T}{nM} \sum_{i=1}^{n} t(x_i) + \frac{1}{n} \sum_{i=1}^{n} (\Phi_{n,i} - \frac{11T}{M}) t(x_i) \) that gives us equation (2.11) by the error definition (2.8) and by (2.7). Finally we have \( E[s_n] = E[ \frac{1}{n} \sum_{i=1}^{n} \Phi_{n,i} x_i ] = \frac{1}{n} \sum_{i=1}^{n} E[\Phi_{n,i}] 1 \mu \). Since \( \Phi_{n,i} \) is doubly stochastic, the same property holds for its statistical expectation, thus implying that \( E[\Phi_{n,i}] 1 = 1 \), yielding \( E[s_n] = E[x_n] = 1 \mu \). ▽

2.3 Some Bounds

In this section the basic properties of the error \( e_{n,j} \), defined in (2.8), are investigated. This will be key to prove many results in the optimality of running consensus. Let us define:

- \( \sigma^2 \) \( \text{def} = \text{VAR}[t(x)] \);
- \( \xi^3 \) \( \text{def} = E[\|t(x) - \mu 1\|^3] \), where \( \|z\| \) is the Euclidean norm of the vector \( z \);
- \( \lambda_U \) the second largest eigenvalue of \( E[W^T W] \);
- \( \lambda_L \) the minimum eigenvalue of \( E[W^T W] \);

where \( \text{VAR}[t(x)] \) is the variance of \( t(x) \). The fundamental hypothesis \( \lambda_U < 1 \) means that the graph of the network is connected [42], i.e. the data sharing involves each sensor.

**Proposition 2.** Suppose \( \alpha_n = 1, \beta_n = M \) and \( \lambda_U < 1 \) then for all \( n = 1, 2, \ldots \) and \( j = 1, \ldots, M \) we have

\[
E[e_{n,j}^2] \leq C_1(M, \lambda_U) \sigma^2, \tag{2.15}
\]

\[
E[|e_{n,j}|^3] \leq C_1(M, \lambda_U) \xi^3 + C_2(M, \lambda_U) \sigma^3, \tag{2.16}
\]
where

\[ C_1(M, \lambda_U) = M^3 \frac{\lambda_U}{1 - \lambda_U}; \tag{2.17} \]
\[ C_2(M, \lambda_U) = M^{\frac{2}{3}} \left( \frac{\sqrt{\lambda_U}}{1 - \sqrt{\lambda_U}} + \frac{1}{1 - \lambda_U} \right). \tag{2.18} \]

\[ \bullet \]

**Proof.** See appendix A.1

Suppose, without loss of generality, that \( \mu = 0 \). Let us define the covariance matrix of the state vector

\[ C_n \overset{\text{def}}{=} E [s_n s_n^T], \tag{2.19} \]

denoting by \((C_n)_{ij}\) its entries. Now we define

\[ \rho_{c,n,ij} \overset{\text{def}}{=} \frac{(C_n)_{ij}}{\sqrt{(C_n)_{ii}(C_n)_{jj}}}, \quad \text{and} \quad \rho_{e,n,ij} \overset{\text{def}}{=} \frac{\sqrt{(C_n)_{ii}(C_n)_{jj}}}{(C_n)_{ii} + (C_n)_{jj}}. \]

The first quantity is the standard statistical correlation coefficient, and the second is the ratio between the geometric and the arithmetic mean of two diagonal entries of the matrix \( C_n \).

**Definition 2.** The consensus coefficient between nodes \( i \) and \( j \) is defined as:

\[ \rho_{n,ij} \overset{\text{def}}{=} \rho_{c,n,ij} \rho_{e,n,ij} = \frac{2 (C_n)_{ij}}{(C_n)_{ii} + (C_n)_{jj}}. \tag{2.20} \]

**Definition 3.** The performance coefficient of sensor \( i \) is defined as:

\[ \gamma_{n,i} \overset{\text{def}}{=} \frac{(C_n)_{ii}}{\sigma_n^2}. \tag{2.21} \]

where \( \sigma_n^2 \overset{\text{def}}{=} \sigma^2/(n \, M) \).

The rationale of the above definition is explained in Chapter 3.
2.4 Binary Hypothesis Test by Running Consensus

**Theorem 1.** If \( \lambda_U < 1 \) then for all \( i = 1, \ldots, M \) and \( j \neq i \) the following bounds hold

\[
\begin{align*}
(M - 1) \psi_n^L &\leq \gamma_{n,i} - 1 \leq (M - 1) \psi_n^U, \\
\frac{M \psi_n^L}{1 + (M - 1) \psi_n^U} &\leq 1 - \rho_{n,ij} \leq \frac{M \psi_n^L}{1 + (M - 1) \psi_n^U},
\end{align*}
\]  

(2.22)

where we define

\[
\psi_n^U \overset{\text{def}}{=} \frac{\lambda_U - \lambda_n^0}{n - 1 - \lambda_U} \quad \text{and} \quad \psi_n^L \overset{\text{def}}{=} \frac{\lambda_L - \lambda_n^0}{n - 1 - \lambda_L}.
\]

\[\blacksquare\]

**Proof.** See appendix A.1 \(\nabla\)

\[\text{Proof.}\]

2.4 Binary Hypothesis Test by Running Consensus

Consider the problem to discriminate between the hypotheses

\[
\begin{align*}
\mathcal{H}_0 &: \ x_{i,j} \sim f_{\theta_0}(x), \\
\mathcal{H}_1 &: \ x_{i,j} \sim f_{\theta}(x),
\end{align*}
\]  

(2.24)

where \( i = 1, 2, \ldots, n, \ j = 1, 2, \ldots, M \), the symbol \( \sim \) stands for “is distributed according to”, and \( f_{\theta_0}(x), f_{\theta}(x) \) are the marginal probability density functions of the data, parametrized by \( \theta \). Note that the moments of \( t(x) \) will depend on \( \theta \), then

\[
\begin{align*}
\mu(\theta) &= E_{\theta}[t(x)]; \\
\sigma^2(\theta) &= \text{VAR}_{\theta}[t(x)]; \\
\xi^3(\theta) &= E_{\theta} \left[ ||t(x) - \mu(\theta) 1||^3 \right];
\end{align*}
\]  

(2.25)

(2.26)

(2.27)

where \( E_{\theta}[\cdot] \) stands for expectation under \( f_{\theta}(x) \), including as particular case \( \theta = \theta_0 \). We define the efficacy \( d = \sqrt{M \mu'(\theta_0) / \sigma(\theta_0)} \), the role of the efficacy will be explained in Chapter 3. The statistic \( T_n \) of the observations

\[ [x_1, x_2, \ldots, x_n], \]
2. Mathematical Properties of Running Consensus

can be \( T_n = s_n^c \) and \( T_n = s_{n,j} \), with \( \alpha_n = 1, \beta_n = M \), and \( \chi_n = 1 \).

We are interested to study the asymptotic behaviour \( (\theta \to \theta_0) \) of:

- the fixed sample size test or Neyman-Pearson paradigm [12], in which the number of measurements is fixed;
- the sequential test [51], in which the number of measurements is a random variable.

2.4.1 Fixed Sample Size Test

For the fixed sample size case we use the model (2.24), with \( \theta_n = \theta_0 + \gamma \frac{\sqrt{n}}{\sqrt{M}}, \ \gamma > 0, \) (2.28)

and the performances are studied in the limit of \( n \to \infty \). For a given \( n \) the detector is identified by the pair \( (T_n, \delta_n) \), in the sense that \( \mathcal{H}_1 \) is declared whenever \( T_n \geq \delta_n \) and \( \mathcal{H}_0 \) is declared otherwise. The detection and false alarm probabilities are accordingly defined as \( p_{dn} = \Pr[T_n \geq \delta_n | \mathcal{H}_1], \ p_{fn} = \Pr[T_n \geq \delta_n | \mathcal{H}_0] \).

**Theorem 2.** Suppose that:

- \( \lambda_U < 1; \)
- \( \mu(\theta) \) is differentiable with \( \mu'(\theta_0) \neq 0, \sigma(\theta) \) is a continuous function;
- the centralized test statistic \( s_n^c \) fulfills

\[
\begin{align*}
\frac{s_n^c - n M \mu(\theta_0)}{\sqrt{n M \sigma(\theta_0)}} & \xrightarrow{f_{\theta_0}} N(0,1), \\
\frac{s_n^c - n M \mu(\theta_n)}{\sqrt{n M \sigma(\theta_n)}} & \xrightarrow{f_{\theta_n}} N(0,1),
\end{align*}
\]

(2.29)

where \( f_{\theta_0} \rightarrow (\text{resp. } f_{\theta_n} \rightarrow) \) denotes convergence in distribution under \( f_{\theta_0}(x) \) (resp. under \( f_{\theta_n}(x) \)) as \( n \) diverges.
Then, the decentralized detection statistic $s_{n,j}$, for any sensor $j = 1, 2, \ldots, M$, is asymptotically equivalent to $s_n^{(c)}$ in the sense that, for any fixed asymptotic probability of false alarm $p_f = \lim_{n \to \infty} p_{fn}$, they achieve the same asymptotic probability of detection

$$p_d = \lim_{n \to \infty} p_{dn} = Q(Q^{-1}(p_f) - \gamma d) \quad (2.30)$$

\[\blacksquare\]

**Proof.** The proof is in Appendix B.1 \[\nabla\]

### 2.4.2 Sequential Test

For the sequential case we use the model (2.24), with \[51\]

$$\theta_r = \theta_0 + \frac{1}{\sqrt{r}} \quad (2.31)$$

where $r$ is a positive number, and the asymptotic performances will be studied in the limit $r \to \infty$. Consider the sequential test

$$\begin{cases} T_n - n M \eta_r \geq b_r, & \text{declare } \mathcal{H}_1, \\ T_n - n M \eta_r \leq a_r, & \text{declare } \mathcal{H}_0, \\ T_n - n M \eta_r \in (a_r, b_r), & \text{take another sample} \end{cases} \quad (2.32)$$

with stopping time

$$N_r = \inf \{ n : T_n - n M \eta_r \notin (a_r, b_r) \}. \quad (2.33)$$

The false alarm and detection probabilities are

$$p_{dr} = \Pr[\text{declare } \mathcal{H}_1 | \mathcal{H}_1], \quad (2.34)$$

$$p_{fr} = \Pr[\text{declare } \mathcal{H}_1 | \mathcal{H}_0], \quad (2.35)$$

and their asymptotic values are

$$p_d = \lim_{r \to \infty} p_{dr}, \quad (2.36)$$

$$p_f = \lim_{r \to \infty} p_{fr}. \quad (2.37)$$
In (2.32) we define
\[
\eta_r \overset{\text{def}}{=} \frac{\mu(\theta_r) + \mu(\theta_0)}{2},
\]
(2.38)
\[
a_r \overset{\text{def}}{=} \sqrt{r M \sigma(\theta_0)} \log \frac{1 - p_d}{1 - p_f},
\]
(2.39)
\[
b_r \overset{\text{def}}{=} \sqrt{r M \sigma(\theta_0)} \log \frac{p_d}{p_f}.
\]
(2.40)

**Theorem 3.** Suppose that:
- \(\lambda_U < 1\);
- \(\mu(\theta)\) is differentiable, \(\mu'(\theta_0) \neq 0\) and \(\sigma(\theta)\) is continuous;
- the centralized test statistic \(s^{(c)}_{\bar{n}}\) fulfills
\[
\frac{s^{(c)}_{[rt]} - [rt] M \eta_r}{\sqrt{r M \sigma(\theta_0)}} \overset{f_{\theta_0}}{\longrightarrow} \mathcal{W}_{-d/2}(t),
\]
and \(\frac{s^{(c)}_{[rt]} - [rt] M \eta_r}{\sqrt{r M \sigma(\theta_0)}} \overset{f_{\theta_r}}{\longrightarrow} \mathcal{W}_{+d/2}(t),
\]
(2.41)
where \(f_{\theta_r} \rightarrow \) (resp. \(f_{\theta_0} \rightarrow\)) denotes weak convergence to a random process \([52]\), under \(f_{\theta_0}(x)\) (resp. \(f_{\theta_0}(x)\)) as \(r\) diverges, and \(\mathcal{W}_z(t)\) is a Wiener process with drift \(z\).
- for any \(\epsilon > 0\), there exist \(r_\epsilon\) and a function \(g_\epsilon(t) > 0\) such that, for all \(r \geq r_\epsilon\) and \(t \geq 1,
\[
P_{\theta_0} \left[ \frac{s^{(c)}_{[rt]} - [rt] M \mu(\theta_0)}{\sqrt{r M \sigma(\theta_0)}} > \epsilon t \right] \leq g_\epsilon(t),
\]
(2.42)
\[
P_{\theta_r} \left[ \frac{s^{(c)}_{[rt]} - [rt] M \mu(\theta_r)}{\sqrt{r M \sigma(\theta_r)}} \leq -\epsilon t \right] \leq g_\epsilon(t),
\]
with \(\int_1^\infty g_\epsilon(t) \, dt < \infty\), and by \(P_\theta\) we mean the probability corresponding to distribution \(f_\theta(x)\).
Then, the decentralized statistic $s_{n,j} \forall j = 1, 2, \ldots, M$, is asymptotically equivalent to $s_n^{(c)}$ in the sense that, for any fixed $p_d$ and $p_f$, they achieve the same asymptotic expected stopping time

$$
\lim_{r \to \infty} r^{-1} E_{\theta_0}[N_r] = 2 \frac{D_b(p_f, p_d)}{d^2}, \\
\lim_{r \to \infty} r^{-1} E_{\theta_r}[N_r] = 2 \frac{D_b(p_d, p_f)}{d^2},
$$

where $D_b(p, q)$ stands for the Kullback-Leibler divergence between the binary probability mass functions $[p, 1-p]$ and $[q, 1-q]$ \[53]\.

Proof. The proof is in Appendix B.2. ▼
2. Mathematical Properties of Running Consensus
Chapter 3

Sequential Estimation

In this chapter the sequential estimation problem is solved by the running consensus procedure and a new suitable definition of consensus is proposed related to the asymptotic optimality of our scheme, see [1, 2]. The chapter is so organized. The problem is precisely formalized in Section 3.1, the performance metrics are defined in Section 3.2 and the related analytical bounds are developed in Section 3.3. Examples of applications are provided in Section 3.4.

3.1 Parameter Estimation

Let \( n = 1, 2, \ldots \) and \( M \), be respectively the discrete time index and the number of sensors. For each time slot \( n \), the sensors of the network collect measurements that are iid realizations of a random variable with mean \( \mu \) and with variance \( \sigma^2 \). These observations are iid over time and across the sensors. The goal is to estimate \( \mu \). The arithmetic mean of the \( nM \) measurements is the statistic of the ideal centralized system, and it represents the benchmark for the running consensus estimator.\(^1\) The ideal centralized system, \(^1\)Note that the arithmetic mean is a consistent estimator of \( \mu \).
3. Sequential Estimation

defined in (2.3) with \( \chi_n = 1/n \) and \( t(x) = x \), can be recast as

\[
s_n^{(c)} = \frac{n-1}{n} s_{n-1}^{(c)} + \frac{1}{n} x_n. \tag{3.1}
\]

Consider the running consensus update rule with \( \alpha_n = (n-1)/n, \beta_n = 1/n \)

\[
s_n = \frac{n-1}{n} W_n s_{n-1} + \frac{x_n}{n}, \tag{3.2}
\]

in which there is a light difference with (2.1): the new measurements taken at the moment time \( n \) are not exchanged among the sensors. It is clear that all the asymptotic results and the physical insights are the same in both the cases. Note that the first term on the right-hand side (RHS) of (3.2) enforces consensus among the nodes in other words one would like that for \( n \) sufficiently large

\[
W_n s_{n-1} \approx 1 s_{n-1}^{(c)}, \tag{3.3}
\]

that means that the state of each node is close to each other and close to ideal centralized state. The second term on the RHS of (3.2) accounts for the new measurements. These two terms are properly weighted by \((n-1)/n\) and \(1/n\), respectively. The state \( s_{n,j} \), for the generic sensor \( j \), is the local estimator for the parameter \( \mu \) at time \( n \).

In Figure 3.1 the system evolution is depicted with \( \mu = 0 \). Note that for \( n \) sufficiently large the running consensus algorithm seems to force \( s_{n,j} \approx s_{n}^{(c)} \) as expressed in eq. (3.3). The mathematical explanation of this phenomenon is given in the next sections.

In the example it is supposed

\[
W_n = W_{i_1 j_1} W_{i_2 j_2} \ldots W_{i_v j_v}, \tag{3.4}
\]

in which the random matrices \( \{W_{ij}\} \) can be expressed as

\[
W_{ij} = I - \frac{(e_i - e_j)(e_i - e_j)^T}{2}, \tag{3.5}
\]

where \( e_i \) denotes a vector of zeros with only the \( i^{th} \) entry equal to 1 and, accordingly, the product \( W_{ij} s_{n-1} \) amounts to replace
the entries $i$ and $j$ of the vector with their arithmetic mean, which is just the pairwise averaging. From equation (3.4) in the $n$th slot the pairs of nodes performing the $v$ pairwise averages are $(i_1, j_1), \ldots, (i_v, j_v)$. The fact that $W_n$ is actually the product of several pairwise matrices (i.e., $v > 1$) is a minor aspect, while the important fact with eq. (3.2) is the simultaneous presence of the sensing stage and of the averaging step. The state of each sensor at time $n$ encompasses the current measurement made in that time slot. On the other hand, note that the $v$ pairwise averages performed within the $n$ time slot involves the state of the node at the previous time slot. In these regards, different formalizations are possible, leading to minor modifications of the formulas while leaving substantially unchanged the physical insights.

### 3.2 Performance Metrics

By observing Figure 3.1 it seems that when the time is large, the state of each sensor is approximately close to each other and close to the ideal centralized state. We can quantify analytically this effect by the performance figures $\gamma_{n,i}$ and $\rho_{n,ij}$ defined in Chapter 2. In the following two subsections these metrics are discussed.

#### 3.2.1 Consensus in the Network

Consider the consensus coefficient $\rho_{n,ij} = \rho_{n,ij}^c \rho_{n,ij}^e$, defined in (2.20). As well known, $\rho_{n,ij}^c$ quantifies the degree of statistical dependence between random variables (state of the nodes, in our case), and attains its maximum value of 1 only if the variables are linearly dependent almost everywhere (a.e.). On the other hand, $\rho_{n,ij}^e$ is an index of equality between two positive numbers (here variances), and takes the value 1 only when these numbers are equal. Consequently, the product $\rho_{n,ij}^c \rho_{n,ij}^e$ belongs to $(-1, 1)$ and the value 1 is attained only when the two random variables coincide a.e.: having the same mean and being linearly related ($\rho_{n,ij}^c = 1$) they can only differ for a scale factor, which must by unitary because the variables share the same variance ($\rho_{n,ij}^e = 1$).
Figure 3.1: State evolution of the sensors in a network made of $M = 3$ nodes, with $v = 10$ pairwise averages per single time slot. $s_n^{(c)}$ is the arithmetic mean of all processed measurements in the network, up to time $n$. The zoom emphasizes the behavior in the first five time slots.
3.3 Asymptotic Optimality

This legitimates the adoption of $\rho_{n,ij}$ as a quantitative measure of the consensus degree: when $\rho_{n,ij} = 1$ the state of the two nodes is a.e. identical. While a unitary consensus coefficient for all the pair of nodes would mean that all nodes share the same state (a.e.), it is also necessary to check that such state is that desired.

3.2.2 Comparison with the Ideal Centralized System

If all the observations collected by the entire network up to time slot $n$ were simultaneously available to a single device, this latter could compute the following arithmetic mean, which represents the state of an ideal centralized system: $s_n^{(c)} = \frac{1}{nM} \sum_{i=1}^{nM} x_i^n$, see Section 2.2 for the mathematical details. Computing $s_n^{(c)}$ turns out to be the main goal of many inference problems in WSNs where the optimal decision/estimation statistic is obtained by averaging the network observations, or functions thereof, see, e.g., [47]. Accordingly, in our fully decentralized architecture, the ideal goal would be that any node attained the same performances of the centralized scheme. To this aim, let us refer to the mean square error. As to the statistical mean, $E[s_n^{(c)}]$ is just $\mu$ (that we set to zero for simplicity), and being $E[s_n] = 1\mu$ we recognize that the state of all the nodes takes the same mean value of the optimal centralized scheme, at any time $n$. Consequently, all the nodes should hopefully share the same variance of the centralized scheme $\sigma_n^2 = \sigma^2/(nM)$. When $\gamma_{n,i} = (C_n)_i / \sigma_n \to 1$ the optimal centralized performance is reached for the node $i$.

3.3 Asymptotic Optimality

Here we introduce upper and lower bounds for the metrics $\rho_{n,ij}$ and $\gamma_{n,i}$, using the analogous of Theorem 1 when the update rule is (3.2).
Theorem 4. If $\lambda_U < 1$ then for all $i = 1, \ldots, M$ and $j \neq i$ the following bounds hold

$$(M - 1) \psi_n^L \leq \gamma_{n,i} - 1 \leq (M - 1) \psi_n^U,$$

$$(3.6)$$

$$\frac{M \psi_n^L}{1 + (M - 1) \psi_n^U} \leq 1 - \rho_{n,ij} \leq \frac{M \psi_n^U}{1 + (M - 1) \psi_n^L},$$

$$(3.7)$$

where

$$\psi_n^U = \frac{1}{n} \frac{1 - \lambda_U^n}{1 - \lambda_U}$$

and

$$\psi_n^L = \frac{1}{n} \frac{1 - \lambda_L^n}{1 - \lambda_L}.$$
3.3. Asymptotic Optimality

Using the update rule (2.1), with $\alpha_n = (n-1)/n$, $\beta_n = 1/n$, the approximate bounds are

$$B_{n}^{U,L} = \frac{M}{n} \frac{\lambda_{U,L}}{1 - \lambda_{U,L}},$$

(3.10)

with rate

$$r = M \frac{\lambda_{U}}{1 - \lambda_{U}}.$$  

(3.11)

- Useful insights about the differences with classical consensus algorithms can be gained by allowing multiple averaging steps per time slot, namely, by assuming that the weighting matrix $W_n$ in eq. (3.2) is actually the product of $v > 1$ iid doubly stochastic matrices, say $W_n = M_{n,1} \ldots M_{n,v}$. Now we can obtain bounds like those in the theorem, in terms of the smallest and the second largest eigenvalues of $E[MM^T]$, say $\xi_L \geq 0$ and $\xi_U < 1$: the final result amounts to substitute in the claims of the theorem $\lambda_{U,L}$ with $\xi_{U,L}$, respectively. In the classical consensus scenario, see e.g. [42], $v$ diverges so that $\xi_{U,L}$ become negligible with exponential rate.

- In the same spirit of the previous item, one may also refer to the case that each sensor just spends time $n$ to gather data and then exchange the locally averaged data using classic algorithms [42] which reach consensus in a time negligible compared to the data gathering time $n$. Of course, this scheme does not take into account possible sensors’ failures occurring before the data exchange process, which is the main motivation of our analysis.

- A distinct feature of the running consensus scheme is the speed of convergence of the performance figures. In fact, this is substantially different from the exponential law that governs the classic consensus algorithms [42]. Furthermore, in our setup, the specific network topology/connectivity (which rules the system eigenvalues) is less crucial with respect to the classical case. Indeed, in this latter, a network design
yielding a larger value of $\xi_U$ is expected to exponentially outperform a system with a smaller eigenvalue. Oppositely, in our case, the universal scaling law is $n^{-1}$, and only the value of the rate coefficient can be tuned by the eigenvalues of the system.

### 3.4 Examples: Pairwise Averaging

In this section we limit the analysis to the simple pairwise protocol described in section 3.1. Specifically, we make reference to the network topologies schematically depicted in Figure 3.2 and we assume that only the admissible pairs of nodes (those connected by straight lines) can be selected for the pairwise averaging. Any such pair is selected with one and the same probability, so that $W_n = W_{ij}$ and any realization of such random matrix (any choice of an admissible pair $(i, j)$) has the same chance of occurrence.

In this case, the eigenvalues appearing in Theorem 4 admit a simple interpretation. Indeed, from eq. (3.5) we immediately see that $W_{ij}$ is doubly stochastic, symmetric and idempotent. The last two properties imply that $E[WW^T] = E[W]$, with the consequence that the eigenvalues $\lambda_L$ and $\lambda_U$ can be equivalently referred to this latter matrix. In addition, it can be easily shown that our basic requirement, namely $\lambda_U < 1$, is fulfilled provided that the graph associated to $E[W]$ is strongly connected [54], and this is certainly true in the architectures of Figure 3.2.

In Figure 3.2 (a), $M = 15$ sensors are arranged to form a ring, and all the pairs of node are admissible. In Figure 3.2 (b) the same ring topology of Figure 3.2 (a) is considered, with the difference that each sensor can only communicate with four neighbors, two in one direction and two in the opposite one, and this result in a lower number of admissible pairs. Similarly, Figures. 3.2 (c)-(d) refer to a network made of $M = 30$ sensors with a topology typical of randomly deployed sensors, and in (c) more pairs of nodes are admissible than in (d).

The considered architectures determine $E[W]$ and, specifically,
Panel (a) represents a completely connected ring of $M = 15$ sensors. The same ring is considered in panel (b) where each node is connected only to four neighbors. In panels (c) and (d), which refer to $M = 30$, the node position is that typical of randomly deployed sensors, with the network in (c) having a larger number of admissible pairs than that depicted in (d).
Figure 3.3: Bounds for the normalized variance $\gamma_{n,i}$ and for the consensus coefficient $\rho_{n,ij}$, as provided by Theorem 1. The upper bounds are drawn as solid curves, while the lower bounds as dashed lines. The labeling (a)-(d) refers to the networks in Figure 5.3 (in (a) the two bounds coincide). Recall that (a) and (b) refer to $n = 15$, while (c) and (d) refer to $n = 30$, whence the differences of $\gamma_{n,i}$ at the initial point $n = 1$. 
its eigenvalues $\lambda_U$ and $\lambda_L$. For the scenario in Figure 3.2(a), we have $\lambda = \lambda_U = \lambda_L = (M-2)/(M-1) \approx 0.9286$ in our case. In fact, we can easily find $E[W] = \lambda I + \frac{1-\lambda}{M}11^T$, and the eigenvalues of such matrix are $\lambda, \lambda, \ldots, \lambda, 1$. The bounds in both the eqs. (2.22) and (2.23) coincide, implying that $\gamma_{n,i}$ and $\rho_{n,ij}$ can be computed exactly. These functions are drawn in Figure 3.3 as solid lines without markers in top and bottom panels, respectively. We see that $\gamma_{n,i}$ starts from $M = 15$ and decreases monotonically toward 1, while $\rho_{n,ij}$ grows monotonically from 0 to 1.

Consider now the network of Figure 3.2(b): here we have $\lambda_U \approx 0.9868$ and $\lambda_L \approx 0.8921$. The bounds on $\gamma_{n,i}$, see eq. (2.22), are shown in the top plot of Figure 3.3 while the bounds in eq. (2.23) for the consensus coefficient $\rho_{n,ij}$ are drawn in the bottom plot. The vertical axis of this latter is limited to the meaningful range $(-1, 1)$; in fact the lower bound for $\rho_{n,ij}$ may occasionally fall below $-1$, thus losing significance.

The eigenvalues ($\lambda_U, \lambda_L$) of the networks depicted in Figure 3.2(c) and (d) are (0.9964, 0.9426) and (0.9994, 0.9262), respectively. The bounds in eqs. (2.22) and (2.23) are also drawn in Figure 3.3. Note that the performance index $\gamma_{n,i}$ now starts from $M = 30$.

The asymptotic behavior of the network performances is better highlighted in Figure 3.4, where the panels (a)-(d) refer to the their analogue in Figure 3.2, and the bounds for $\gamma_{n,i} - 1$ and $1 - \rho_{n,ij}$ are drawn on the same plot. Note that, for large $n$, the bounds simplify as in eq. (3.8), giving the portion of the curves marked by dots. Clearly, in Figure 3.4(a) the upper and the lower bound coincide so that only two curves are drawn. In Figure 3.4 we also check the derived bounds by means of computer simulations based on a standard Monte Carlo counting procedure. For each time slot $n = 1, 2, \ldots$, the shown simulation points involve $10^3$ program runs for computing the entries $(C_n)_{ij}$ of the covariance matrix. Then, the estimated values of $\rho_{n,ij}$ and $\gamma_{n,i}$ are obtained as arithmetic averages of the pertinent entries; for instance, the values of $\gamma_{n,i}$ result from averaging the $n$ diagonal entries $(C_n)_{ii}$, see eq. (2.21).
3. Sequential Estimation

Figure 3.4: Bounds for the performance indices $\epsilon_n = \gamma_{n,i} - 1$ and $\epsilon_n = 1 - \rho_{n,ij}$ as function of $n$, for the same four scenarios (a)-(d) of Figure 5.3. For large $n$, the bounds of the two performance figures coincide, as prescribed by eq. (3.8). The upper bounds are drawn as solid curves, while the lower bounds as dashed lines. The points marked with “+” and “×” result from computer simulations for estimating respectively $\gamma_n = \sum_i \gamma_{n,i}/M$ and $\rho_n = \sum_{ij} \rho_{n,ij}/J$, where $J$ is the number of admissible pairs.
Chapter 4
Sequential Detection

The detection problem is considered using the running consensus \[3,4\]. Here the sensors are cooperating by the running consensus scheme to discriminate among two hypotheses. The problem formalization is presented in Section 4.1 then the decentralized statistic is developed in Section 4.2. The asymptotic optimality is proved in Section 4.3 where we study the performance for the Neyman-Pearson setup, in which the number of measurements used for the decision is fixed, and for the sequential setup in which there is a virtually unbounded number of data available, and an appropriate number of these is used to make the final decision according to prescribed error probabilities. Examples of application and computer experiments are given in Section 4.4 while Section 4.5 addressed some specific issues of practical relevance.

4.1 Centralized Hypothesis Testing

We assume that a network of wireless sensors monitors a phenomenon of physical interest, modeled as a binary state of the nature. Each sensor collects \( n \) samples; all data are iid. By denoting with \( x_{i,j} \) the \( i^{th} \) sample collected by the \( j^{th} \) sensor, the goal of the network is to discriminate between the two hypotheses

\[
\mathcal{H}_0 : x_{i,j} \sim f_{\theta_0}(x), \\
\mathcal{H}_1 : x_{i,j} \sim f_{\theta}(x),
\]

(4.1)
where \( i = 1, 2, \ldots, n, \ j = 1, 2, \ldots, M \) and \( f_{\theta_0}(x), f_{\theta}(x) \) are the marginal probability density functions of the data, parametrized by \( \theta \).

A centralized statistic \( T_n^{(c)} \) must be able to discriminate among \( \mathcal{H}_0 \) and \( \mathcal{H}_1 \). Two main different kinds of detection strategies are available: the Neyman-Pearson setup \([11, 55]\) and the sequential setup \([56, 57]\).

In the Neyman-Pearson setup the time index \( n \) is fixed and non-random, then the number of measurements is \( nM \). The centralized detector is identified by the pair \( (T_n^{(c)}, \delta_n) \), in the sense that \( \mathcal{H}_1 \) is declared whenever \( T_n^{(c)} \geq \delta_n \) and \( \mathcal{H}_0 \) is declared otherwise. Typically the choice of \( \delta_n \) takes account of the desired false alarm probability.

In the sequential paradigm the number of measurements cannot be fixed in advance, and the decision rule is defined as \([56, 57]\)

\[
\begin{cases}
    T_n^{(c)} \geq b, & \text{declare } \mathcal{H}_1, \\
    T_n^{(c)} \leq a, & \text{declare } \mathcal{H}_0, \\
    T_n^{(c)} \in (a,b), & \text{take another sample,}
\end{cases}
\]  

(4.2)

where \( a \) and \( b \) determine the error probabilities of the test \([56]\).

The time index, in which a decision is taken, is called stopping time. In Figure 4.1 an example of sequential test, defined in (4.2), is schematically depicted. Under \( \mathcal{H}_0 \) the detection statistic tends to decrease to eventually declare \( \mathcal{H}_0 \) when the lower threshold is crossed. Under \( \mathcal{H}_1 \) the detection statistic tends to increase up to eventually declare \( \mathcal{H}_1 \) if the upper threshold is crossed. The stopping time is the first instant index in which a threshold is crossed.

### 4.2 Hypothesis Testing in Sensor Networks with Running Consensus

The focus is on detectors that are asymptotically equivalent when the hypotheses come close to each other, as the number of sam-
Figure 4.1: The sequential test, defined in (4.2) is depicted here. Under $\mathcal{H}_0$ the detection statistic tends to decrease down to eventually declare $\mathcal{H}_0$ when the lower threshold is crossed. Under $\mathcal{H}_1$ the detection statistic tends to increase up to eventually declare $\mathcal{H}_1$ if the upper threshold is crossed. The stopping time is defined as the first instant index in which a threshold is crossed.
ples increases. We consider, as is standard in these contexts \[12\],
detection statistics of the form:

\[
T_n^{(c)} = \sum_{i=1}^{n} \sum_{j=1}^{M} t(x_{i,j});
\]

note that \(T_n^{(c)}\) is the ideal centralized state, defined in (2.3) choosing \(\chi_n = 1\). Since data are iid, the first two moments of the ideal centralized statistic are the sum of the moments of \(t(x_{i,s})\):

\[
\mathbb{E}_\theta[T_n^{(c)}] = n M \mu(\theta), \quad \text{and} \quad \text{VAR}_\theta[T_n^{(c)}] = n M \sigma^2(\theta),
\]
where \(\mathbb{E}_\theta[\cdot]\) stands for expectation under \(f_\theta(x)\), including as particular case \(\theta = \theta_0\). In order to compute statistics of the form (4.3), all data should be available at a common site. The focus is on a fully decentralized flat architecture, without a fusion center and we would like to obtain some surrogate form of \(T_n^{(c)}\) made available to any sensor of the network.

We assume that sensors measure and exchange data according to the running consensus procedure, as detailed in Chapter 2. Let us denote with \(T_{n,j}\) the state (detection statistic) computed at the \(n^{th}\) epoch by the \(j^{th}\) node. The updating rule for the states of the nodes according to the eq. (2.1) with \(\alpha_n = 1\) and \(\beta_n = M\)

\[
T_n = W_n \left[ T_{n-1} + M \, t(x_n) \right],
\]
where the definition of vectors \(T_n\) and \(t(x_n)\) is equivalent of that in Chapter 2.

For our purposes, it is convenient to make explicit the relationships between the statistic available at the \(j\) sensors and the centralized statistic \(T_n^{(c)}\): we can always write, \(\forall n\),

\[
T_n = T_n^{(c)} \, 1 + e_n,
\]
or, by zooming on a single sensor:

\[
T_{n,j} = T_n^{(c)} + e_{n,j},
\]
\(e_{n,j}\) being the running consensus error at instant \(n\) for the \(j^{th}\) sensor, see (2.8).
4.3 Asymptotic Performances

The main result we are going to present is that the decentralized statistic $T_{n,j}$, for any node $j$, is asymptotically equivalent to $T_{n}^{(c)}$, as claimed by the theorems in the next section. Such result follows from the basic properties of the error $e_{n,j}$, see Chapter 2. Roughly speaking, the consensus error becomes less relevant with respect to the centralized statistic $T_{n}^{(c)}$ when the time index $n$ grows. In fact $T_{n}^{(c)}$ has an expectation proportional to $n$ while the average square of the consensus error is bounded by a constant, independent from $n$, see equation (2.15). Now supposing that $n$ is sufficiently large the consensus error becomes negligible with respect to the centralized statistic.

4.3 Asymptotic Performances

A meaningful setup for the asymptotic design and characterization of the detector is obtained by studying test (4.1), as the parameter $\theta$ of the alternative hypothesis approaches $\theta_0$, and we investigate such issue by considering two different kinds of detection strategies: the Neyman-Pearson setup (FSS test) and the sequential paradigm (see e.g., [56]).

In the former case, we refer to the standard framework of Pitman ARE [58], and asymptotic optimality is proved by simply exploiting the boundedness of the second moment of $e_{n,j}$. Addressing the sequential case is considerably more involved. Therefore, we first introduce the proper asymptotic framework proposed by Lai [51], and then offer a rigorous proof of asymptotic optimality, as detailed in Appendices A.1 and B.2.

4.3.1 Fixed Sample Size Test

Let us consider the hypothesis test of a simple alternative $\theta = \theta_0$ against the one-sided alternative $\theta > \theta_0$, with $\theta$ being a real, unknown parameter. Formally, we use model (4.1), with [12]

$$\theta_n = \theta_0 + \frac{\gamma}{\sqrt{n}}, \quad \gamma > 0, \quad (4.7)$$
and the performance figures are studied in the limit of $n \to \infty$. For a given $n$ the ideal centralized detector is identified by the pair $(T_n^{(c)}, \delta_n)$, in the sense that $\mathcal{H}_1$ is declared whenever $T_n^{(c)} \geq \delta_n$ and $\mathcal{H}_0$ is declared otherwise. The detection and false alarm probabilities are accordingly defined as $p_{dn} = \Pr[T_n^{(c)} \geq \delta_n | \mathcal{H}_1]$, $p_{fn} = \Pr[T_n^{(c)} \geq \delta_n | \mathcal{H}_0]$, and the detector is said of (asymptotic) size $p_f$ if $\lim_{n \to \infty} p_{fn} = p_f$ [12].

\begin{equation}
\begin{align*}
\frac{T_n^{(c)} - n M \mu(\theta_0)}{\sqrt{n M \sigma(\theta_0)}} & \xrightarrow{f_{\theta_0}} N(0, 1), \\
\frac{T_n^{(c)} - n M \mu(\theta_n)}{\sqrt{n M \sigma(\theta_n)}} & \xrightarrow{f_{\theta_n}} N(0, 1),
\end{align*}
\end{equation}

where $\xrightarrow{f_{\theta_n}}$ (resp. $\xrightarrow{f_{\theta_0}}$) denotes convergence in distribution under $f_{\theta_n}(x)$ (resp. under $f_{\theta_0}(x)$) as $n$ diverges.

Convergence in distribution requires that the cumulative distribution function of a sequence of random variables converges toward the cumulative distribution function (the standard normal, in our case) of the limiting random variable. Now, asymptotic normality under $\theta_0$ is expected by simple application of the Central Limit Theorem (CLT), due to the additive nature of the centralized statistic $T_n^{(c)}$. Under the alternative, the convergence is less immediate since the normalization terms and the underlying distributions vary with $n$. However, exploiting extensions of CLT to triangular arrays or Le Cam’s contiguity theory, such convergence is usually met under very mild technical conditions, see e.g., [11, 52].

Under the assumption (4.8), it can be shown that, for any detector of size $p_f$, $p_d = \lim_{n \to \infty} p_{dn} = Q(Q^{-1}(p_f) - \gamma d)$, where $Q(\cdot)$ is the area under the right tail of a standard Gaussian function, $Q^{-1}(\cdot)$ is its inverse function, and $d = \sqrt{M} \mu'(\theta_0)/\sigma(\theta_0)$ is called efficacy [11, 12]. We note explicitly that the asymptotic formula

\begin{footnotesize}
\begin{itemize}
\item[\footnote{In the literature sometimes $d^2$ is referred to as the efficacy, see also footnote 3 in [12]. Note that, in general, the normalizing functions $n M \mu(\theta)$ and $\sqrt{n M \sigma(\theta)}$ yielding asymptotic normality need not to be the moments of the}]
\end{itemize}
\end{footnotesize}
depends on the specific relationship between $\theta_n$ and $\theta_0$ through the factor $\gamma$.

Comparison of two detectors in the asymptotic regime is usually accomplished by their Asymptotic Relative Efficiency (ARE). The ARE of detector 2 with respect to detector 1 is defined as the asymptotic ratio of the sample size of detector 1, say $n_1$, divided by that of detector 2, $n_2$, for the same asymptotic probabilities $p_f$ and $p_d$. We have $\text{ARE} = (n_1/n_2) = (d_2/d_1)^2$, where $d_1$ and $d_2$ are the efficacies of detector 1 and 2, respectively [58, 59]. As a consequence, we shall say that two detectors are asymptotically equivalent if they share the same efficacy $d$, and that a detection statistic is asymptotically optimal if it reaches the best attainable efficacy, which, under suitable regularity conditions, is $d_{\text{max}} = \sqrt{MI(\theta_0)}$, where $I(\theta_0)$ is the Fisher information at $\theta_0$ [11, 60].

Since now, we have essentially summarized some known results referred to the ideal centralized statistic $T_n(c)$. Now, we want to characterize the asymptotic detection performances of the decentralized statistic $T_{n,j}$ when the running consensus algorithm comes into scene. To this end, let us consider the detector $(T_{n,j}, \delta_n)$, that is, we design the running consensus test using the same threshold $\delta_n$ which is used for the ideal system.

**Theorem 2.** Under mild technical conditions, see Subsection 2.4.1, if the network graph is connected then the decentralized detection statistic $T_{n,j}$ is asymptotically equivalent to the centralized detection statistic $T_n(c)$.

Before ending this section, we would like to make a brief comment on the main hypotheses of the theorem. The technical regularity conditions, detailed in Subsection 2.4.1 are usually adopted in the context of asymptotic detection [12], and focused on the asymptotic normality of the detection statistic. The condition on the graph connectivity is a basic requirement [42] of having that the information can flow toward/from each sensor.
4. Sequential Detection

4.3.2 Locally Optimum Sequential Detection

First, some known results about sequential testing between continuous time processes are recalled. Let $W(t)$ be a standard Wiener process, where $t$ ranges over the reals, and suppose that one wants to test the hypothesis that the Wiener process has a negative drift $-d/2$ against that of a positive drift $+d/2$. In formulas:

\[ H_0: \ W(t) = W_{-d/2}(t), \]
\[ H_1: \ W(t) = W_{+d/2}(t), \]

Adopting a sequential approach, it is known that the optimum test, in the sense of minimizing the expected sample size for a prescribed pair of error probabilities $p_f$ and $p_d$, is

\[
\begin{align*}
&\begin{cases}
W(t) \geq \beta, & \text{declare } H_1, \\
W(t) \leq \alpha, & \text{declare } H_0, \\
W(t) \in (\alpha, \beta), & \text{take another sample}
\end{cases}
\end{align*}
\]

where the thresholds exactly take the form

\[
\alpha = \frac{1}{d} \log \frac{1-p_d}{1-p_f}, \quad \beta = \frac{1}{d} \log \frac{p_d}{p_f}, \quad (4.9)
\]

The above test implicitly defines a (continuous) stopping time, viz.:

\[ \tau = \inf \{ t : W(t) \notin (\alpha, \beta) \}. \]

and the average times for making a decision are

\[
E_{-d/2}[\tau] = 2 \frac{D_b(p_f, p_d)}{d^2}, \quad E_{d/2}[\tau] = 2 \frac{D_b(p_d, p_f)}{d^2}, \quad (4.10)
\]

where we used $E_{\pm d/2}$ to denote expectation under the two hypotheses, and where $D_b(p, q)$ stands for the Kullback-Leibler divergence between the binary probability mass functions $[p, 1-p]$ and $[q, 1-q]$.

Let us now come back to our discrete-time problem, involving the statistic $T_n^{(c)}$, see (4.3). In paralleling the reasoning used in the FSS case, we study the asymptotic properties of a sequential
4.3. Asymptotic Performances

decision rule as the hypotheses come close to each other. However, in sequential tests the number of samples is a random quantity, and a natural modification of (4.7), see [51], is

$$\theta_r = \theta_0 + \frac{1}{\sqrt{r}},$$

(4.11)

where $r$ is a positive number, and the asymptotic performances will be studied in the limit $r \to \infty$.

The sequential test is then implemented as follows. We fix a value of the parameter $r$ in (4.11), and the decision rule is designed for such $r$. To this aim, it is expedient to shift the detection statistic such that its expectations under the two hypotheses are opposite in sign. This is easily accomplished by defining

$$\frac{E_{\theta_r}[T_n^{(c)}] + E_{\theta_0}[T_n^{(c)}]}{2} = n M \frac{\mu(\theta_r) + \mu(\theta_0)}{2} \overset{\text{def}}{=} n M \eta_r,$$

so that a sequential decision rule can be formulated as

$$\begin{cases} T_n^{(c)} - n M \eta_r \geq b_r, & \text{declare } H_1, \\ T_n^{(c)} - n M \eta_r \leq a_r, & \text{declare } H_0, \\ T_n^{(c)} - n M \eta_r \in (a_r, b_r), & \text{take another sample.} \end{cases}$$

(4.12)

Such a test clearly defines implicitly a random variable $N_r$ representing the number of samples needed to terminate the testing procedure:

$$N_r = \inf \left\{ n : T_n^{(c)} - n M \eta_r \notin (a_r, b_r) \right\}.$$

(4.13)

The false alarm and detection probabilities are

$$p_{dr} = \Pr[\text{declare } H_1 | H_1],$$

(4.14)

$$p_{fr} = \Pr[\text{declare } H_0 | H_0].$$

(4.15)

A suitable mathematical tool for dealing with the asymptotic performance of sequential tests, in the limit of $r \to \infty$, is provided in [51], which essentially formulates a LOD (Locally Optimum Detection) theory in the sequential framework. The rationale behind
the development used in [51] is as follows. For a prescribed pair of error probabilities, as the two hypotheses come close to each other (i.e., as $r$ diverges), the (average) number of samples needed by the detection statistic for exceeding one of the thresholds is expected to increase. Therefore, the time evolution of the statistic can be regarded as a random walk which moves inside two barriers, with single steps that become smaller and smaller with respect to the distance between the barriers, as the hypotheses approach each other. Otherwise stated, as $r$ grows, the random walk approaches a continuous time process.

In [51] the above simple intuition is made precise by considering the following detection statistic

$$T^{(c)}_{[rt]} - [rt] M \eta_r = T^{(c)}_n - n M \eta_r,$$

(4.16)

where $[x]$ stands for the integer part of $x$, and $t$ defines a continuous time axis. This means that we are looking at a piecewise constant random process which changes values at the integer time instants $n/r$ (for integer $n$), where the elementary step $1/r$ goes to zero as $r$ diverges. As to the limiting distribution of the above statistic for $r \to \infty$, we can invoke the functional central limit theorem, namely, the convergence to Wiener processes. In paralleling eq. (4.8) suppose that the following convergences hold ($d$ is the efficacy):

$$\frac{T^{(c)}_{[rt]} - [rt] M \eta_r}{\sqrt{r M \sigma(\theta_0)}} \xrightarrow{f_{\theta_r}} \mathcal{W}_{-d/2}(t),$$

$$\frac{T^{(c)}_{[rt]} - [rt] M \eta_r}{\sqrt{r M \sigma(\theta_0)}} \xrightarrow{f_{\theta_0}} \mathcal{W}_{+d/2}(t),$$

(4.17)

where $f_{\theta_r}$ (resp. $f_{\theta_0}$) denotes weak convergence to a random process [52], under $f_{\theta_r}(x)$ (resp. $f_{\theta_0}(x)$) as $r$ diverges. As previously discussed for the FSS test, the convergence under $\theta_0$ is expected by application of the (functional version of the) CLT, also known as Donsker’s theorem [52]. For the same reasons explained for the FSS test, proving the convergence under $\theta_r$, which is usually met in many practical cases, is more tricky.
In light of (4.17), the results on testing two continuous Wiener processes $W_{\pm d/2}(t)$ summarized at the beginning of this section, can be exploited in our discrete-time setup. Indeed, a threshold comparison of the form

$$T_{c,0}^{(c)} [rt] M \eta_r \in (\alpha, \beta),$$

yields explicit expression of the thresholds to be used in (4.12):

$$a_r = \sqrt{r M} \sigma(\theta_0) \alpha, \quad b_r = \sqrt{r M} \sigma(\theta_0) \beta. \tag{4.18}$$

It is possible to show that the detection and false alarm probabilities used for setting the thresholds $\alpha$ and $\beta$, are asymptotically attained by the designed detector [51]. Namely,

$$\lim_{r \to \infty} p_d r = p_d, \quad \lim_{r \to \infty} p_f r = p_f. \tag{4.19}$$

Furthermore, in view of the relationship $n = [rt]$, one would expect that $N_r \sim [r \tau]$. Indeed, it turns out that, in the light of eq. (4.10), the expected sample size scales as [51]

$$\lim_{r \to \infty} E_{\theta_r}[N_r] = 2D_b(p_f, p_d) d^2,$$

$$\lim_{r \to \infty} E_{\theta_r}[N_r] = 2D_b(p_d, p_f) d^2.$$

By defining the asymptotic relative efficiency as the ratio of the expected sample sizes, and accordingly taking the limit for large $r$, it is immediate to see that these two detectors are simply compared in terms of their efficacies, just as happens in the FSS case. It is now natural to use, for the sequential framework, the same definitions of asymptotic equivalence and asymptotic optimality adopted in the FSS case.

In addition, the unbeatable efficacy in the sequential case is the same of $d_{\text{max}}$ defined in the FSS context. This result can be obtained by using the expressions of the average stopping time for the optimal Sequential Probability Ratio Test (SPRT [57, 61]), in the limit of large $r$, see [50, 51].

The above reasoning has been focused on the ideal centralized statistic $T_n^{(c)}$. As done before, we switch now on the decentralized...
statistic available at the $j^{th}$ node $T_{n,j}$ and, as for the FSS test, we consider a sequential detector using the same thresholds $a_r$ and $b_r$ which are used for the ideal system. The main results about the sequential scenario is that the sequential test

$$
\begin{cases}
T_{n,j} - n M \eta_r \geq b_r, & \text{declare } H_1, \\
T_{n,j} - n M \eta_r \leq a_r, & \text{declare } H_0, \\
T_{n,j} - n M \eta_r \in (a_r, b_r), & \text{take another sample,}
\end{cases}
$$

(4.20)

behaves asymptotically as one in which the statistic $T_{n,j}$ is replaced by the ideal centralized statistic $T_{n}^{(c)}$. This statement is made precise in the following claim.

**Theorem 3.** Under mild technical conditions, see Subsection 2.4.2 if the network graph is connected then the decentralized detection statistic $T_{n,j}$ in eq. (4.3) is asymptotically equivalent to the centralized detection statistic $T_{n}^{(c)}$.

Before ending this section we would like to make a remark. Since now, we have considered a family of alternative hypotheses with parameter $\theta_r = \theta_0 + 1/\sqrt{r}$, and accordingly designed and characterized the asymptotic performances of tests (4.12) and (4.20), with thresholds (4.18), as $r$ diverges. It is also of interest to consider $\theta_r = \theta_0 + \chi/\sqrt{r}$, with $\chi$ being unknown. The lack of knowledge about $\chi$ prevents us from setting the thresholds on the actual value of $\theta_r$.

One solution is to design tests (4.12) and (4.20) with thresholds (4.18), i.e., assuming a nominal value of $\chi = 1$. Clearly, the form of the test being unchanged, the error probability and average sample number under the null hypothesis will not depend upon $\chi$. Moreover, it is shown in [51] that, provided the technical conditions compactly called uniform invariance principles, this test exhibits an asymptotic detection probability and an expected sample number under the alternative hypothesis still expressible in terms of the performances of a Wiener process, but with drift parameter $(\chi - 1/2) d$. As a result, the comparison between two different test statistics can be still made by essentially looking at their efficacies.
4.4 Examples and Numerical Experiments

The previous analysis is now corroborated by computer experiments with the twofold goal of (i) providing a sanity check for the developed theory, and (ii) investigating the behavior of the running consensus detection in practical, i.e., non-asymptotic, scenarios.

As to the running consensus scheme, among the many possible choices we consider a simple pairwise exchange protocol, wherein a pair of sensors \((j,k)\) is randomly and uniformly selected among all the possible pairs taken from the set \(\{1, 2, \ldots, M\}\), see previous chapters for details. The implicit assumption of a fully connected network is also made. If the pair of sensors is \((j,k)\), the consensus matrix \(W_{jk}\) would take the following form:

\[
W_{jk} = I - \frac{(e_k - e_j)(e_k - e_j)^T}{2}, \quad (4.21)
\]

where \(e_k\) denotes a vector of zeros, but for the \(k^{th}\) entry which equals to 1, and where \(I\) is the identity matrix. Note that, when the above \(W_{jk}\) is multiplied by a vector, its effect is to replace the vector entries \(j\) and \(k\) by their arithmetic mean. Actually we assume that, in a single consensus step, many pairs of sensors can average their data. This amounts to a connection matrix given by the product of \(v \geq 1\) pairwise matrices of the form (4.21). Clearly, such a consensus matrix operates by averaging the states of \(v\) pairs of randomly selected nodes.

We first address, both in the FSS case and in sequential framework, a Gaussian example which is particularly interesting since it naturally leads to optimal detectors. Then, a non-Gaussian test is considered, focusing on the sequential case.
Figure 4.2: FSS test for the Gaussian example in a network made of $M = 10$ sensors. We show the detection probability $p_{dn}$ of the running consensus versus $n$, for $v = 1, 10$, pairwise exchanges per epoch. According to the asymptotic framework here we set $\theta_n = 1/\sqrt{n}$. Also shown is the performance of the ideal centralized system that, in this example, is constant with $n$ and represents the asymptotic performance. The number of Monte Carlo trials is $10^4$. 
4.4.1 Neyman-Pearson FSS Tests

Let us consider the following Gaussian shift-in-mean hypothesis test: for \( i = 1, 2, \ldots, n \), and \( j = 1, 2, \ldots, M \),

\[
\begin{align*}
H_0 & : x_{i,j} \sim \mathcal{N}(0, \sigma^2), \\
H_1 & : x_{i,j} \sim \mathcal{N}(\theta, \sigma^2),
\end{align*}
\]

where recall that the \( x_{i,j} \) are iid, and \( \mathcal{N}(\theta, \sigma^2) \) is our shortcut for a Gaussian distribution with mean \( \theta \) and variance \( \sigma^2 \). For this scheme, clearly \( \theta_0 = 0 \), and

\[
\mu(\theta) = \theta, \quad \sigma^2(\theta) = \sigma^2, \quad d = \sqrt{\frac{M}{\sigma^2}} = \sqrt{MI(0)},
\]

where \( I(0) = 1/\sigma^2 \) is the Fisher information. In this particular case, the optimal detection statistic (i.e., the log-likelihood) is just in the additive form of (4.3), with \( t(x_{i,j}) \propto x_{i,j} \). As a consequence, the ROC (Receiver Operating Characteristic) of the centralized statistic \( T_{n}^{(c)} \) can be computed in closed form as

\[
Q \left( Q^{-1}(p_f) - \sqrt{\frac{\theta^2}{\sigma^2} nM} \right).
\]

Note that, imposing \( \theta = \theta_n = \gamma/\sqrt{n} \), as prescribed by the asymptotic theory, straightforwardly yields \( p_d = Q \left( Q^{-1}(p_f) - \gamma d \right) \).

We are now ready to comment on the simulation results pertaining to the considered Gaussian shift-in-mean detection problem, with and without the consensus stage. From the theory, we know that, in the asymptotic setting specified in the previous sections, with the hypotheses coming closer and closer as \( n \to \infty \), the decentralized detector must approach the same performance as that of the ideal centralized one. Figure 4.2 accordingly shows the detection and false alarm probabilities of a test based on the decentralized statistic \( T_{n,j} \) (for a generic \( j \)) of the running consensus scheme, as function of \( n \), by assuming \( \theta_n = \gamma/\sqrt{n} \) with \( \gamma = 1 \). Therefore, Figure 4.2 provides a sanity-check for the asymptotic
results and is useful to show the rate of convergence toward the ideal system.

Figure 4.2 also emphasizes the role that different kinds of consensus schemes may have. With reference to the pairwise algorithm it shows that, by increasing the number $v$ of pairwise averaging per single time slot, the convergence become faster, as one may expect. In addition, gossip protocols with different number of sensors participating to the single-slot average are considered. Intriguingly, it is not easy to anticipate the relative merits of the different communication schemes. For instance, in Figure 4.2, a pairwise with $v = 3$, involving 6 sensors per single time slot, is outperformed by the 4-wise scheme.

A possible explanation of this behavior can be that, in order to achieve 4-wise averaging, we need at least 4 pairwise exchanges. For example, with four sensors, we can first take average of $x_1$ and $x_2$, and then of $x_3$ and $x_4$. Then, sensors 1 and 3, and sensors 2 and 4 can do pairwise averaging, which then gives each sensor the 4-wise average $(x_1 + x_2 + x_3 + x_4)/4$.

### 4.4.2 Sequential Tests

**Gaussian example**

For the same shift-in-mean Gaussian problem studied in the previous section, we have implemented our sequential detection strategy with running consensus. Assuming $\theta_r = 1/\sqrt{r}$, and using the detection statistic in eq. (4.3), with decision rule and thresholds given by eq. (4.12) and (4.18), the centralized system will compare the statistic

$$\sum_{i=1}^{n} \sum_{j=1}^{M} (x_{i,j} - \frac{nM}{2\sqrt{r}}),$$

with thresholds

$$a_r = \sqrt{r} \sigma^2 \log \frac{1 - p_d}{1 - p_f}, \quad b_r = \sqrt{r} \sigma^2 \log \frac{p_d}{p_f}.$$  

As in the FSS case, for this simple Gaussian problem, it is straightforward to show that the above (ideal, centralized) sequential test
Figure 4.3: Sequential tests for the Gaussian example, with $M = 10$ sensors and $v = 5$ pairwise exchanges per each time instant. The average sample number $E[N]$ multiplied by the SNR is displayed as function of this latter, for three different values of the (nominal, asymptotic) error probability $p_e \approx [0.01, 0.05, 0.1]$. The three different curves refer to the ideal centralized strategy, the decentralized one with running consensus, and the asymptotic value. The number of Monte Carlo trials is $10^4$. 
Figure 4.4: This figure complements Figure 4.3 by showing the actual error probability, as function of the SNR, for the same case study of the previous figure.
is nothing but the optimal SPRT. In the computer simulations, and according to our theoretical findings, we use the above thresholds also for the running consensus scheme. Furthermore, for simplicity, we work under the assumption that $p_f = 1 - p_d$, that we call $p_e$, yielding symmetric thresholds and equality of the average sample numbers, that we accordingly denote by $E[N]$.

As discussed in Section 4.3.2, the asymptotic performance of the SPRT (and, in view of Theorem 3, also of the running consensus) is just ruled by $d = \sqrt{M/\sigma^2}$, such that, following eq. (4.20),

$$E[N] \sim 2 r \frac{D_b(1 - p_e, p_e)}{d^2} = \frac{2 M}{M} \frac{D_b(1 - p_e, p_e)}{\text{SNR}},$$

where we used $\text{SNR} = \theta^2 / \sigma^2$.

In Figure 4.3 the (scaled) expected stopping time is displayed, as function of the signal-to-noise ratio, for three different values of the nominal (asymptotic) error probabilities. It can be seen that, as the SNR goes to zero, the product $E[N] \cdot \text{SNR}$ of the sequential test with running consensus approaches that of the ideal centralized system, and both converge toward the asymptotic constant value $2D_b(1 - p_e, p_e) / M$. Figure 4.3 also reveals that the expected sample number of the decentralized scheme is smaller than that of the optimal SPRT. This actually makes sense, in view of Figure 4.4 where the other relevant performance index, that is, the error probability, is displayed. Figure 4.4 shows that the error probability enforced by the decentralized scheme is always greater than that of the ideal system, thus explaining the decrease of $E[N]$ for the running scheme. The simulation results also show that both the running consensus scheme and the ideal centralized entity exhibit error probabilities that approach their nominal asymptotic values, as predicted by (4.19). The above considerations suggest making a further comparison between the two strategies in the following way: For each value of the error probability actually achieved by the decentralized scheme (see Figure 4.4), we evaluate the correspondent average sample sizes obtained with the ideal centralized system (i.e., an SPRT). The ratio between these numbers, namely, the relative efficiency, is displayed in Figure 4.5.
Figure 4.5: A summary of the evidences in Figures 4.3 and 4.4. The (non asymptotic) relative efficiency of the decentralized versus an optimal SPRT using the same (actual) error probabilities reached by the decentralized strategy as in Figure 4.4 see main text for details.
as function of the SNR, for the same three different values of error probability as in the previous figures. As it should be, being the comparison made for the same error probabilities, the SPRT outperforms our strategy, while being asymptotically equivalent (relative efficiency equal to one), when the SNR goes to zero. Furthermore, it is seen that, the lower the error probability, the faster is the convergence toward the asymptotic value. This may be explained because smaller error probabilities mean larger thresholds and, for fixed SNR, the number of samples required to end the test grows. In these conditions, the relative impact of the consensus error is less predominant.

A Non-Gaussian Example

We switch now to a non-Gaussian shift-in-mean detection problem:

\[ H_0 : x_{i,j} \sim p N(0, \sigma_1^2) + (1 - p) N(0, \sigma_2^2) , \]
\[ H_1 : x_{i,j} \sim p N(\theta, \sigma_1^2) + (1 - p) N(\theta, \sigma_2^2) , \]

that is, a *p*-weighted mixture of two Gaussian random variables having different variances.

We shall use, as nonlinearity characterizing the statistic in (4.3), the well-known score function [11]:

\[ t(x) = \frac{\partial}{\partial \theta} \log f_\theta(x) \bigg|_{\theta=0} = x \frac{p e^{-\frac{x^2}{2\sigma_1^2}} + (1-p) e^{-\frac{x^2}{2\sigma_2^2}}}{pe^{-\frac{x^2}{2\sigma_1^2}} + (1-p) e^{-\frac{x^2}{2\sigma_2^2}}} . \]

For the score detection statistic applied to our non-Gaussian setup, it is no longer easy to write explicit expressions for \( \mu(\theta) \) and \( \sigma(\theta) \), as well as for the Fisher information. As a consequence, we evaluate these quantities by numerical integration, and accordingly use the computed values for setting the thresholds. We would like to stress that the test can be shown to attain the maximum efficacy \( \sqrt{MI(0)} \), see, e.g., [50].

As done before, the simulations are designed to compare the centralized detector and the decentralized structure that implements the running consensus, under the scaling law \( \theta_r = 1/\sqrt{r} \).
Figure 4.6: Sequential tests for the mixture-of-Gaussians example, with $M = 10$ sensors and $v = 5$ pairwise exchanges per each time instant. The parameters of the distribution in eq. (4.23) are $p = 0.3$, $\sigma_1 = 1$ and $\sigma_2 = 5$. The arithmetic average $E[N]$ of the sample numbers in eq. (4.24) multiplied by the SNR is displayed as function of this latter, for $p_e \approx 0.1$. The four different curves pertain to the ideal centralized strategy with a score detector, the running consensus with score detector, the optimal SPRT and the asymptotic value. The number of Monte Carlo trials is $10^4$. 
4.5. Some Practical Issues

In addition, an optimal (centralized) Wald’s SPRT is also considered. Our Monte Carlo results are summarized in Figures 4.6 and 4.7. Since in this case it is no longer possible to enforce perfect symmetry in terms of error probabilities and expected sample sizes, we display the arithmetic averages

\[ p_e = \frac{p_f + (1 - p_d)}{2}, \quad E[N] = \frac{E_{\theta_0}[N] + E_{\theta_r}[N]}{2}, \quad (4.24) \]

and further define the signal-to-noise ratio as

\[ \text{SNR} = \frac{\theta_r^2}{p \sigma_1^2 + (1 - p) \sigma_2^2}. \]

As can be seen, for the score detector, the theoretical predictions are verified by the numerical evidence, in that the two statistics are seen to be asymptotically equivalent as the SNR decreases, and further they reach the best attainable performance \( \sqrt{\text{MI}(0)} \), emphasizing how the consensus algorithm is effective for sequential detection. Switching to the comparison with the optimal SPRT, we see that the differences between this latter and the centralized score detector are moderate, while the consensus error has an impact in non-asymptotic regimes.

Similar considerations apply to the error probability \( p_e \), as shown in Figure 4.7. As found in the previous example, the centralized and the decentralized schemes do not achieve the same actual error probabilities. Therefore, we have again evaluated the (non-asymptotic) relative efficiencies of the proposed detectors and similar considerations as in in Figure 4.5 apply, not reported here for sake of brevity.

4.5 Some Practical Issues

Let us consider the sequential paradigm with running consensus, and note that in formulation (4.20) there is no explicit global stopping time for the detection task. In fact, when an individual sensor exceeds a threshold level we say that its individual decision is
Figure 4.7: This figure complements Figure 4.6 by showing the actual error probability, as function of the SNR, for the same case study of the previous figure.
4.5. Some Practical Issues

taken but, nonetheless, the sensor actually continues sensing the environment and participating in the data exchange. This is convenient from an analytic point of view because it preserves the problem symmetry, and allows easier analysis. Indeed, as already noted, the performance of the running consensus scheme can be computed with reference to an arbitrary sensor.

From a more practical perspective, however, a global termination rule is certainly needed. One option is to stop the task of the node not at its threshold crossing but a little later, when it is very likely that all other nodes finished the task as well. Alternatively, there may be an additional flag in the sensors' data exchange that allows counting the number of sensors with individual decision taken. When such number reaches the total number of sensors, the system is turned off. Or, else, one could use broadcast messages to communicate the threshold crossings.

The choice of one specific rule is expected to have a little impact on the system behavior, in the light of the key point shown in the previous sections, that all nodes behaves asymptotically in the same manner. In Figure 4.8 we report an instance of a quantitative analysis carried out on this aspect. We see that the different stopping times at different sensors are close to each other, implying that the additional processing and sensing for a sensor whose statistic has already crossed the threshold is essentially negligible.

A second issue that merits attention concerns the communication burden implied by the running consensus. Here we make specific reference to the sequential framework, but the arguments can be adapted to the FSS case as well. The energy spent for the communication task certainly depends upon a number of factors, including the specific communication protocols, the physical resource available, the presence of feedback, the implementation of error control coding, the presence of synchronism signals, and so forth. To simplify the matter, let us assume roughly that the energy spent in any time slot is proportional to the number of connections. For instance, if $v = 1$, in the pairwise case the data exchange involves two sensors and hence costs 2 energy units, while in the 3-wise protocol the average of the sensors’ state costs 6
Figure 4.8: A realization, under $\mathcal{H}_1$, of the sequential test in (4.20), for the Gaussian shift in mean problem with SNR $= -30$ dB. Here we consider a simple network made of $M = 5$ sensors and the pairwise algorithm with $v = 1$. The detection statistics of each sensor (dashed curves) are close to each other, and all closely follow the behavior of the ideal centralized statistic (solid curve). When the threshold is crossed the process is stopped to show how different stopping times cluster around a single value, as shown in the zoomed plot.
4.5. Some Practical Issues

units, since each sensor must communicate with the remaining 2, and in the 4-wise the cost is 12 units. If \( v > 1 \) these costs must be multiplied for \( v \), the number of exchanges per time slot\(^2\).

As to the ideal centralized system, this corresponds to the case where all the sensors share their state with all other sensors in any time slot: the energy cost amounts to \( M(M - 1) \approx M^2 \) units. Alternatively, in an architecture equipped with a fusion center that implements the sequential test, the \( M \) sensors at any time slot send their data to the fusion center and the communication cost reduces to \( M \) energy units. In any case, we see that the ideal centralized system (ideal only with respect to the detection performance) implies a cost proportional or quadratic with \( M \). Conversely the running consensus strategy is much more parsimonious in terms of energy.

Summarizing, with the running consensus, it is possible to achieve asymptotic optimal performances with a communication cost that does not grow with the network size \( M \). On the other hand, increasing the energy spent for the gossip protocol (that is, increasing the number of sensors sharing data in a single slot), might serve at speeding up the convergence toward the asymptotic performance.

As a final remark, we would like to stress that an implicit assumption made in this work is that the sensors exchange unquantized data, i.e., real numbers. A possible concern is that such working assumption can be only approximately met by practical systems. Furthermore, when the communication resources are particularly scarce, it would be certainly of interest to consider the case where sensors can only exchange severely quantized data and, to one extreme, only binary digits, i.e., local decisions corresponding to one-bit quantization of the local likelihoods. The problem can be conveniently put in the framework of the so-called “quantized consensus”, see e.g., [40, 48]. Addressing the asymptotic performance of systems operating under a running quantized

\(^2\)The classical consensus procedure is obtained by assuming that the number \( v \) of exchanges goes to infinity, and the communication cost must be computed accordingly, see e.g., [42, 62].
consensus paradigm is unlikely to be a simple generalization of the results proposed in this work, and probably requires quite different tools of analysis.
Chapter 5

Quickest Change Detection

The quickest change detection, implemented via running consensus, is addressed in this Chapter, see [5]. The problem is to detect an event (that is modeled by a change in the data distribution) as soon as possible. Approximate performance evaluation is developed, moreover we show that the running consensus performance is close to the ideal centralized system.

The Chapter is organized as follows. In Section 5.1 the problem is formalized, and classical results on Page’s test are summarized. Section 5.2 deals with the proposed decentralized counterparts thereof. In Section 5.3 a bank of parallel Page’s test is provided. In Section 5.4 we give approximate formulas for the performance, in terms of relative efficiencies, of each system. In Section 5.5 the algorithms are tested on a typical change detection problem, and the theoretical formulas are compared with the results of Monte Carlo simulations.

5.1 Problem Formalization: Centralized Page’s Test

We assume that a wireless sensor network is engaged in solving a change detection problem, which can be modeled as follows. Let \( n \geq 1 \) be a discrete time index, and let \( j \in \{1, 2, \ldots, M\} \) denote
5. Quickest Change Detection

the sensor of the network. The \( n^{th} \) observation \( x_{n,j} \) collected by the \( j^{th} \) remote node obey a given distribution \( f_0(x) \) until an unknown time \( n_0 \). From \( n_0 \) (included) on, the distribution at all sensors changes to \( f_1(x) \). The goal of the network is to detect the change as soon as possible, with a constraint on the average time between false alarms. Formally, \( \forall j \), we have

\[
f_0(x) : x_{1,j}, x_{2,j}, \ldots, x_{n_0-1,j} \quad \Rightarrow \quad f_1(x) : x_{n_0,j}, x_{n_0+1,j}, \ldots
\]

(5.1)

All the involved random variables are assumed to be statistically independent. According to this model, at each time slot \( n \), the network globally collects a vector of observations

\[
x_n = [x_{n,1}, x_{n,2}, \ldots, x_{n,M}].
\]

(5.2)

This vector should be actually available to an ideal centralized entity, this latter would operate according to the well-known Page’s test [63], which works as follows. As detection statistic, it exploits the so-called CUSUM log-likelihood of the data

\[
\sum_{i=1}^{n} l(x_i) = \sum_{i=1}^{n} \sum_{j=1}^{M} \log \frac{f_1(x_{i,j})}{f_0(x_{i,j})},
\]

(5.3)

where \( l(x) \) denotes the log-likelihood pertaining to vector \( x \). The standard recursion for implementing Page’s test is

\[
L_n^{(c)} = \max\{0, L_{n-1}^{(c)} + l(x_n)\},
\]

(5.4)

and the associated (random) stopping time is

\[
N = \arg \min_n \{L_n^{(c)} \geq \gamma\}.
\]

(5.5)

This defines a decision rule: we say that a change happened when the first crossing of the threshold \( \gamma \) takes place. Note also that the recursion rule implies that the log-likelihood resets each time it falls below zero, which is thus the point from which it restarts.
5.1. Problem Formalization: Centralized Page’s Test

Figure 5.1: The Page’s test, under $\mathcal{H}_0$, is schematically depicted here. The detection statistic has a negative drift, then it might be expected a large number of resets. The change of distribution is not present so a false alarm occurs when the statistic hits the threshold (red line).
5. Quickest Change Detection

Figure 5.2: The Page’s test is depicted here: initially the statistics often reset to zero while, after the change in distribution, it tends to increase up to eventually declare the change, when the threshold (red line) is crossed. The detection delay is the time required to declare the change, if this occurs.
In Figures 5.1, 5.2 it is depicted schematically the behavior of the Page’s test.

The usual optimality criterion in change detection is to minimize the delay needed for detecting the change, while diluting as much as possible the false alarms. Accordingly, the main performance metric will be the average delay for detection, corresponding to a prescribed average false alarm rate. As to the latter, it is usually defined as the reciprocal of the average sample size under the null hypothesis, that is, $1/E_0[N]$, with $E_{0,1}[\cdot]$ denoting expectation computed under distribution $f_{0,1}(x)$.

As for the delay to detection $E_1[N]$, we as usual work with the upper bound that one obtains by assuming that the CUSUM is zero at time $n_0$ [50]. The exact computation, in fact, would require to account for the exact value of the CUSUM statistic at $n_0$, and is usually intractable.

Approximate formulas for these relevant quantities are available in the literature, see e.g., [50,57,63]. These mainly rely upon neglecting end effects, that is, the excess over the threshold of the test statistic at the stopping time. The relationship between the detection threshold $\gamma$ and the false alarm rate $R_c$ can be written as

$$R_c(\gamma) \approx \frac{M \Delta_{01}}{e^\gamma - \gamma - 1} \approx M \Delta_{01} e^{-\gamma}$$

(5.6)

where $\Delta_{01}$ is the Kullback-Leibler divergence between $f_0(x)$ and $f_1(x)$, see [53]. Inverting the previous relationship defines the function $\gamma_c(R)$, which, for large thresholds, simplifies in

$$\gamma_c(R) \approx \log \left( M \Delta_{01} / R \right).$$

(5.7)

The approximate relationship between the average delay $D_c$ and the threshold can be found as

$$D_c(\gamma) \approx \frac{\gamma + e^{-\gamma} - 1}{M \Delta_{10}} \approx \gamma \frac{e^{-\gamma}}{M \Delta_{10}}$$

(5.8)

where $\Delta_{10}$ is the divergence between $f_1(x)$ and $f_0(x)$. Note that in the previous equations we append a “c” for this centralized detector.
It is seen that the delay essentially depends linearly from the threshold, while the false alarm has an exponential dependence thereof. Furthermore, in the large-$\gamma$ regime the system operating characteristic $D_c(R)$, i.e., the relationship between the detection delay and the false alarm rate, can be expressed in simple closed form:

$$D_c(R) \approx \frac{\gamma_c(R)}{M \Delta_{10}} \approx \frac{\log (M \Delta_{01}/R)}{M \Delta_{10}}. \quad (5.9)$$

We note explicitly that the factor $M$ which multiplies the divergences takes into account the fact that, at each time slot, $M$ independent observations are collected, such that the overall divergence of the single-time observations is $M \Delta$. For later use, let us report also the formulas pertaining to a single sensor, i.e., those of the idealized entity, specialized for $M = 1$:

$$R_s(\gamma) \approx \frac{\Delta_{01}}{e^\gamma - \gamma - 1} \approx \frac{\gamma}{\Delta_{10}} e^{-\gamma}, \quad (5.10)$$

$$D_s(\gamma) \approx \frac{\gamma + e^{-\gamma} - 1}{\Delta_{10}} \approx \frac{\gamma}{\Delta_{10}}. \quad (5.11)$$

For large $\gamma$, these yield

$$D_s(R) \approx \frac{\log (\Delta_{01}/R)}{\Delta_{10}}. \quad (5.12)$$

### 5.2 Decentralized Page’s Test by Running Consensus

Let us now switch to consider a genuinely decentralized scenario, and suppose, in particular, that a wireless sensor network engaged in a change detection has a fully flat architecture without fusion center. A running consensus protocol is implemented to reach agreement about the ideal centralized detection statistic, as the number of consensus steps grows. The running consensus procedure is explained in the Chapter 2 and will not be repeated here; in the following we only recall the basic elements needed to make this chapter self-consistent.
5.2. Decentralized Page’s Test by Running Consensus

With reference to the running consensus strategy, we assume that sensors are able to “simultaneously” acquire, exchange and process data for eventually declaring the change. More in detail, during each time slot \( n \), observations are first collected by the sensors, and then a randomly chosen subset thereof is selected for sharing their data according to a standard gossip algorithm [42]. The exchanged data are not simply the measurements, but rather the current detection statistic available to a sensor, also referred to as the state, say \( L_{n,j} \), that summarizes its state of knowledge.

What one wants is, of course, that \( L_{n,j} \) for all \( j \) would converge to the the centralized detection statistic \( S_n \), as the time index \( n \) diverges. To this aim, the running consensus protocol prescribes using the following update, where \( l(x) = \log f_1(x)/f_0(x) \) represents the log-likelihood pertaining to a single sample \( x \):

\[
\begin{pmatrix}
L_{n,1} \\
L_{n,2} \\
\vdots \\
L_{n,S}
\end{pmatrix} = W_n \begin{pmatrix}
L_{n-1,1} \\
L_{n-1,2} \\
\vdots \\
L_{n-1,S}
\end{pmatrix} + M W_n \begin{pmatrix}
l(x_{n,1}) \\
l(x_{n,2}) \\
\vdots \\
l(x_{n,M})
\end{pmatrix}
\]  

(5.13)

and where the \( M \) by \( M \) matrices \( W_n, n = 1, 2, \ldots \), are iid (independent identically distributed) and doubly stochastic. The equation (5.13) derives from (2.1) with \( \alpha_n = 1, \beta_n = M \) and \( t(x) = l(x) \).

In Theorem 1, by a very simple manipulation, it is possible to show that the squared error of the running consensus is bounded by a term\(^{1}\)

\[
\propto M^2 E_{0,1} \left[ l^2(x) \right] \frac{\lambda_U}{1 - \lambda_U},
\]  

(5.14)

where the bound is independent of the time instant \( n \). In the above expression \( \lambda_U \) is the second largest eigenvalue of the matrix \( E[W_n W_n^T] \) and takes account of the connection properties of the network [2]. In (5.14) the error is bounded by a term that varies with the network topology and with \( M \).

Let us compactly denote the update rule for the \( j^{th} \) sensor by \( L_{n,j} = U(L_{n-1,j}) \). By using update rule (5.13), together with

\(^{1}\)Note that in Proposition 2 a similar bound is present in which the term \( M^2 \) is replaced with \( M^3 \).
recursion rule (5.4), the overall recursion at the \( j^{th} \) sensor can be written as follows

\[
L_{n,j} = \max\{0, \mathcal{U}(L_{n-1,j})\}.
\]  

(5.15)

Note that, while the update rule \( \mathcal{U} \) is linear, the addition of Page’s reset rule introduce a nonlinear effect, which is not present in the classical gossip algorithms.

A key observation should be made at this point. The nature itself of the running consensus scheme implies that all the nodes asymptotically share one and the same state (statistic). Thus, what one expect is that all the nodes report a detected change at approximately the same time instant. Running consensus introduces strong correlations among nodes by continuously propagating information across the system, see the metric behavior \( \rho_{n,ij} \) for \( n \) large in Theorem 1. This correlation is responsible for detecting the change at almost equal times. Therefore, provided that the system evolves for a sufficiently long time, the estimated time at which the distribution change took place can be recovered by any of the \( M \) nodes of the network, and there is no need to implement some “broadcast” form of halting in the system. Moreover, this also implies that the performance of the running consensus scheme can be computed with reference to any sensor.

Since the running consensus protocol is designed to reach agreement about the ideal centralized statistic \( L_n^{(c)} \), (see previous chapters), it is expected that the statistic \( L_{n,j} \) (any \( j \)) tends to track \( L_n^{(c)} \). Figure 5.3 shows the behavior of these two statistics, and we indeed see that they behave similarly. One can hope that also the performances of the running consensus detector stay close to the theoretical limit represented by the performances of the centralized ideal system. To elaborate on the issue, it is convenient to write the detection statistic available at the \( j^{th} \) sensor as

\[
L_{n,j} = L_n^{(c)} + e_{n,j},
\]  

(5.16)

emphasizing the distance of the current state \( L_{n,j} \) from its asymptotic value \( L_n^{(c)} \), measured by the error term \( e_{n,j} \). Note that this
5.2. Decentralized Page’s Test by Running Consensus

Figure 5.3: Behavior of the running consensus statistics for quickest detection, for a simple network made of three sensors (three dashed lines, almost superimposed each other). Initially, the statistics often reset to zero while, after the change in distribution, they tend to increase up to eventually declare the change, when the threshold is crossed. The times at which the threshold is crossed almost coincide for the three running consensus statistics, indeed, as seen in the zoomed plot, the three running consensus statistics always behave quite similarly. The continuous curve shows the statistic of an ideal centralized system. We see that the running consensus statistics closely track the statistic of the ideal system: their detection delays are close to that of the ideal system.
error is different from that defined in Chapter 2 because of the non-lin-
early in the update rule (5.15). The corresponding stopping time at the $j^{th}$ sensor, which implicitly defines the decision rule, becomes

$$N_j = \arg \min_n \{ L_{n,j} \geq \gamma \}.$$  \hspace{1cm} (5.17)

Approximate performance evaluation of the decentralized Page’s test with running consensus is now in order, via the behavior of the error term $e_{n,j}$. To this aim, we shall exploit results from previous chapters, and try to apply them to the nonlinear update (5.15). We would like to stress that the arguments below are only heuristic, and their validity is checked by computer simulations in a later section.

That said, let us first work under the assumption that the absolute error is strictly bounded, that is, $|e_{n,j}| \leq \epsilon$, $\forall n$ and $\forall j$. The system initially collects data distributed according to $f_0(x)$ and, until a threshold crossing occurs (either because a real change happened, or because a false alarm is going to occur), the $j^{th}$ sensor may have made a certain number of resets to zero, depending upon both $L_n^{(c)}$ and $e_{n,j}$. It is reasonable to assume that, for large values of $\gamma$ (i.e., $\gamma \gg \epsilon$), the threshold crossing will be typically determined by the behavior of the centralized statistic $L_n^{(c)}$, rather than by the error term. Otherwise stated, when $L_n^{(c)}$ starts driving the random walk toward the upper threshold (again, either because a real change happened, or because a false alarm is going to occur), the error term that differentiates the states of the various sensors will become less and less influential compared to $L_n^{(c)}$, and the sensors will tend to agree toward declaring a change.

More formally, by noting that $L_n^{(c)} - \epsilon \leq L_n^{(c)} + e_{n,j} \leq L_n + \epsilon$, we can write (assuming $\gamma > \epsilon$):

$$\arg \min_n \{ L_n^{(c)} + \epsilon \geq \gamma \} \leq N_j \leq \arg \min_n \{ L_n^{(c)} - \epsilon \geq \gamma \}. \hspace{1cm} (5.18)$$

We can thus define

$$N = \arg \min_n \{ L_n^{(c)} \geq \gamma - \epsilon \}, \quad \overline{N} = \arg \min_n \{ L_n^{(c)} \geq \gamma + \epsilon \},$$
which are recognized to be the stopping times associated to classical centralized Page’s tests having thresholds $\gamma - \epsilon$ and $\gamma + \epsilon$, respectively. Thanks to eq. (5.18), we can thus write
\[ E_0[N_j] \leq E_0[N_j] \leq E_0[N], \quad \forall j = 1, 2, \ldots M. \] (5.19)

Thanks to the above, we can make an approximate performance evaluation of the running consensus test. As to the false alarm rate $R_j$ at sensor $j$, we use the lower bound $E_0[N]$ that yields
\[ R_j \leq R_c(\gamma - \epsilon) \]
with the function $R_c(\cdot)$ defined in (5.6). Now, let us fix $R$ and consider for the running consensus the threshold value $\gamma = \gamma_c(R) + \epsilon$, namely the threshold of the centralized system plus $\epsilon$. In this way we get $R_j \leq R_c(\gamma - \epsilon) = R$, implying that such threshold setting provides a conservative choice ensuring that the prescribed false alarm $R$ is not exceeded.

Switching to the analysis of the average delay at sensor $j$, say $D_j$, using (5.19), we have
\[ D_c(\gamma - \epsilon) \leq D_j \leq D_c(\gamma + \epsilon), \]
and using again $\gamma = \gamma_c(R) + \epsilon$, yields
\[ D_c(\gamma_c(R)) \leq D_j \leq D_c(\gamma_c(R) + 2 \epsilon). \]

All these delays diverge when the threshold grows. However, since $D_c(\gamma)$ depends essentially linearly upon $\gamma$, we have that, in the regime of small $R$ (that is, large $\gamma_c(R)$), the ratio between $D_j$ and the average delay of a centralized test with false alarm rate $\bar{R}$ approaches unity, for any $j$. Accordingly, the overall performance of the running consensus test are well-described by the same operating characteristic of the centralized system:
\[ D_r(R) \approx D_c(R) \approx \frac{\log (M \Delta_{01}/\bar{R})}{M \Delta_{10}}, \] (5.20)
where the subscript “$r$” stands for “running”, versus “$c$” for “centralized”.

5.2. Decentralized Page’s Test by Running Consensus
The more important fact to be emphasized, however, is that the \( \epsilon \) appearing in (5.18) is unknown and, for this reason, a useful relationship between the threshold and the false alarm rate is actually not available. To avoid the impasse, one might simply neglect the \( \epsilon \) and set the threshold of the running consensus detector equal to that of the centralized scheme, provided by relationship (5.6), or by approximation (5.7); note that this does not ensure that the design is conservative. A better option would be that of adjusting the threshold by adding a term related to (5.14), in order to provide a more reliable protection against exceeding the false alarm constraint. This, however, requires a case-by-case analysis.

At any rate, while the precise relationship to set the threshold is problematic, what remains true is that any choice of the threshold leads to an operating point belonging to the optimal operating characteristic (5.20).

### 5.3 Bank of Parallel Page’s Detectors

In this section we study the case that each sensor runs its own a Page’s test, without exchanging data on-the-fly with other sensors. Such a bank of parallel Page’s processors evolves until the first of the \( M \) sensors detects a change, and when this happens a broadcast message from the “firing” sensor is sent through the network in order to terminate the inference task. This allows energy saving, once that the change of the state of the nature has been observed. On the other hand, it is just this latter broadcast which enforces some form of cooperation by exploiting the diversity among sensors’ states, and might provide performance improvements with respect to a single Page’s test.

The false alarm rate of the bank can be easily characterized. Let \( L_j \) be the duration of the \( j^{th} \) Page’s test, i.e., the time at which the \( j^{th} \) sensor would detect a change. The overall stopping time is defined as

\[
N^{\text{(bank)}} = \min_{j \in \{1, \ldots, M\}} L_j. \tag{5.21}
\]

Assuming \( f_0(\cdot) \), we resort to the following heuristic arguments.
5.3. Bank of Parallel Page’s Detectors

According to [50], for \( \gamma \) large enough, we can suppose that time interval between two successive false alarms is essentially ruled by the number of resets (a reset happens when the value of the CUSUM statistic is zero) times the length of a single path of the statistics evolving between two successive resets; the length of the path from the last reset to the threshold crossing may be neglected in this approximation (see also Figure 5.1-5.2). Computing the time interval between false alarms as the random number of resets multiplied by the average length of a single inter-reset path, we are faced with characterizing such a random number. For a single Page’s test the number of resets can be shown to follow a geometric distribution [50]; let \((1 - p)p^k\) be the probability of \(k\) resets, \(p\) being the characteristic parameter of the geometric. Then, in the case of a bank of \(M\) Page’s tests the false alarm rate is ruled by the minimum of \(M\) such random variables, which is a geometric variable itself, whose parameter changes to \(p^M\). Accordingly, by replacing \(p\) with \(p^M\) in eq. (5.2.22) of [50], we finally have that the relationship between the threshold \(\gamma\) of the bank and the false alarm rate \(R_b\) is exactly the same already obtained with the ideal centralized system and with the running consensus scheme. Roughly speaking, the false alarm of bank made of \(M\) filters is \(M\) times larger than that of a single filter. Therefore, we assume

\[
R_b(\gamma) \approx R_c(\gamma) \approx \frac{M \Delta_{01}}{e^{\gamma} - \gamma - 1} \approx \frac{M \Delta_{01}}{e^{\gamma}}
\] 

(5.22)

Switching our attention to the average delay, note that \(D_b = E_1[N^{(bank)}]\), and the approximate evaluation of this statistical expectation is now in order. Let \(F_1(x)\) be the cumulative distribution function (CDF, hereafter) of \(L_j\) when the observation model is \(f_1(\cdot)\). Such a CDF can be approximated by that of the stopping time arising from a random walk with positive drift and a single barrier \(\gamma > 0\) (see also the discussion in [64]). This approximation involves the so-called Wald or inverse Gaussian distribution, defined as

\[
F_W(x; z) = \left[1 - Q\left(\frac{x - 1}{\sqrt{x}} \sqrt{z}\right)\right] + e^{2z} Q\left(\frac{x + 1}{\sqrt{x}} \sqrt{z}\right),
\]
5. Quickest Change Detection

$Q(\cdot)$ being the standard Gaussian exceedance probability function. Indeed, 

$$F_1(\xi \mathbb{E}_1[L_j]) \approx F_W(\xi; \gamma \delta), \quad \text{with } \delta = \frac{\Delta_{10}}{\text{VAR}_1[l(x)]},$$

where for $\mathbb{E}_1[L_j]$ we use (5.11). Therefore, from (5.21) and from the fact that the expectation of a nonnegative random variable can be computed as the integral of complementary CDF, we have

$$D_b(\gamma) \approx \gamma + e^{-\gamma} - 1 \int_0^\infty \left[1 - F_W(\xi; \gamma \delta)\right]^M d\xi \quad (5.23)$$

that can be easily used for numerical evaluation.

An explicit expression for operating characteristic of the bank can be found by approximating $\gamma + e^{-\gamma} - 1 \approx \gamma$ in (5.23), and exploiting the last expression in (5.22), yielding

$$D_b(R) \approx \log \left(\frac{M \Delta_{01}/R}{g(M, R) \Delta_{10}}\right) \quad (5.24)$$

where the function $1/g(M, R)$ is defined as

$$\int_0^\infty \left[1 - F_W\left(\xi; \frac{\Delta_{10}}{\text{VAR}_1[l(x)]} \log \left(\frac{M \Delta_{01}}{R}\right)\right)\right]^M d\xi$$

In this paper we mainly use the approximations of (5.23) and (5.24). Nonetheless, a brief digression on how a simpler formula can be also obtained for moderately large $M$, is now provided. Given $M$ iid random variables $L_j$, it holds true that

$$\min_{j \in \{1, \ldots, M\}} L_j \overset{d}{=} F_1^{-1}\left(\min_{j \in \{1, \ldots, M\}} U_j\right),$$

where $U_j$ are iid $(0, 1)$-uniform variates, and $X \overset{d}{=} Y$ means that the two random variables $X$ and $Y$ have the same distribution. This implies

$$\mathbb{E}_1[\mathcal{N}^{(bank)}] = \mathbb{E}\left[F_1^{-1}\left(\min_{j \in \{1, \ldots, M\}} U_j\right)\right].$$
5.4. Relative Efficiencies

For large $M$, a legitimate approximation consists in exchanging the expectation operator with the function $F_1^{-1}(x)$ (see, e.g., Castillo [66], p. 76), yielding

$$E_1 \left[ N^{(bank)} \right] \approx F_1^{-1} \left( \frac{1}{M + 1} \right), \quad (5.25)$$

where we used the fact that the expected value of the minimum of $M$ uniform random variables is $1/(M+1)$. Since, using Wald’s distribution, $F_1(x) \approx F_W(x/E_1[L_j]; \gamma \delta)$, the above yields $F_1^{-1}(y) \approx E_1[L_j] F_W^{-1}(y; \gamma \delta)$. By substituting in (5.25), the average delay can be expressed as

$$D_b(\gamma) \approx \frac{\gamma}{\Delta_{10}} F_W^{-1} \left( \frac{1}{M + 1}; \gamma \delta \right), \quad (5.26)$$

which, in addition to (5.23), is a further approximation of the delay, that could be useful in the regime of large $M$.

5.4 Relative Efficiencies

We use as a proxy to compare different detection schemes the so-called relative efficiency of strategy 1 with respect to strategy 2, for a given false alarm rate $R$. This is defined as

$$\eta_{1,2}(R) = \frac{D_2(R)}{D_1(R)},$$

$D_1(R)$ and $D_2(R)$ being the average delays of detectors 1 and 2, respectively, when their false alarm rate is fixed to the same value $R$. With this definition, $\eta_{1,2}(R) > 1$ implies that strategy 1 outperforms strategy 2, in the sense that it exhibits a smaller delay, for the same false alarm rate, and vice versa for $\eta_{1,2}(R) < 1$. As to the notations, recall that we use the subscript “r” to denote the decentralized Page’s test with running consensus, “b” refers to the bank of parallel Page’s processors, while “s” and “c” refer to the single Page’s test and to the ideal centralized entity, respectively. The approximations derived in the previous sections allows us writing the following relationships.
The efficiency of the optimal centralized Page's detector with respect to its running consensus counterpart, thanks to (5.20), is simply:

\[ \eta_{c,r}(R) = \frac{D_r(R)}{D_c(R)} \approx 1, \tag{5.27} \]

meaning that the two strategies are essentially equivalent. This result merits emphasis: within the limits of the approximation, the decentralized scheme is as efficient as the ideal centralized detector, in terms of detection quickness.

Note that, in our context of a flat network architecture, the best would be that each sensor had the same optimal detection performance of the ideal centralized system, and there is no doubt that any sensible implementation of such a system would require transferring a huge amount of data among the sensors. Therefore, the basic message of (5.27) is that the same goal of mimic the ideal centralized system can be approximately achieved, by means of the running consensus protocol based upon gossip algorithms that lead to much more parsimonious energy consumption for communications.

Recalling that the operating characteristic of the running consensus coincides approximatively with that of the centralized system, the efficiency of the single-sensor test with respect to the running consensus strategy can be computed by combining (5.6) with (5.8), and (5.10) with (5.11). When \( \gamma \) is large enough, using the second approximation of (5.9) and (5.12), one gets

\[ \eta_{s,r}(R) = \frac{D_r(R)}{D_s(R)} \approx \frac{1}{M} \left( 1 + \frac{\log M}{\log(\Delta_{01}/R)} \right). \tag{5.28} \]

For relatively small false alarm rate \( R \), we see that using \( M \) units essentially reduces the delay of a factor \( M \), with respect to the use of a single detector. We note explicitly that the gain in terms of detection performances is paid in the coin of the communication expense: the single sensor, by definition, requires no communication at all; the running consensus, whose energy parsimony with respect to the centralized system has been emphasized above, requires a certain amount of communication among the
sensors. Computing the exact communication cost requires specifying, among other things, the exact gossip procedure adopted.

The comparison between the two decentralized architectures, i.e., the bank of Page’s tests and the running consensus, follows by combining (5.6) with (5.8), and (5.22) with (5.23). Assuming \( \gamma \) large, the relative efficiency can be derived from the second of (5.9) and (5.24) in the form:

\[
\eta_{b,r}(R) = \frac{D_r(R)}{D_b(R)} \approx \frac{g(M, R)}{M}.
\] (5.29)

Note that the energy spent for communication by the bank is essentially negligible, amounting to a single broadcast at the end of the detection task. In this respect, it is almost equivalent to a single sensor. Conversely, running consensus does involve sensor’s communication, and thus one may expect that the bank is outperformed by the running consensus in terms of quickness.

It is also interesting to compare the behavior of the bank against the use of a single Page’s test. From approximations (5.12) and (5.24) we have:

\[
\eta_{b,s}(R) = \frac{D_s(R)}{D_b(R)} \approx \frac{g(M, R)}{1 + \log \frac{M}{\log(\Delta_0/R)}},
\] (5.30)

while a more accurate expression would involve eqs. (5.10), (5.11), (5.22), and (5.23).

5.5 Examples and Numerical Experiments

The previous analysis is now corroborated by computer experiments, and to this aim we select a case study commonly used as benchmark in the context of model change detection. Consider hence the detection problem formalized in (5.1) and sup-
pose that two zero-mean Gaussian distributions, featuring different variances, are involved:

\[ f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad f_1(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}. \]

As to the specific running consensus protocol employed in the decentralized Page’s test, we refer to a standard repeated pairwise averaging in which, during each time slot \( n \), the sensors have the chance of making 5 successive pairwise exchanges; details can be found in, e.g., [2].

We report the results from \( 10^4 \) Monte Carlo simulations for \( M = 10 \) sensors, and \( \sigma \approx 1.032 \). This rather small value of \( \sigma \) implies that the two hypotheses are quite close, namely, that the detection of the change in the statistical distribution is really a difficult task; we want to test our systems just in this challenging scenario.

Figure 5.4 reports the comparison between the simulated false alarm rate, and that obtained by exploiting the approximate analytical relationships between the threshold \( \gamma \) and the false alarm \( R \). The theoretical relationships for the ideal centralized, the running consensus and the bank coincide, as seen by (5.6), (5.22), and by the discussion on the threshold setting for the running consensus detector, while the curve of the single sensor is scaled by a factor \( 1/M \). In Figure 5.4 the lines in bold represent the more precise approximations contained in the quoted equations, while the simplifications for large \( \gamma \) are reported as thinner curves.

The simulation points of the studied systems (no simulation results are given for the case of a single Page’s test) show a satisfying accuracy, except for the case of the running consensus. Indeed, we have already pointed out that using the threshold of the centralized scheme is not very accurate, and at least a corrective term should be added. This could be certainly estimated by simulations, but a case-by-case analysis is required to address change detection problems with different distributions.

In absence of such correction, the practical impact is that one is not able to set accurately the threshold, to guarantee a precise
Figure 5.4: False alarm rate $R$ versus the threshold $\gamma$, for the four systems considered in the paper, and with specific reference to the Gaussian change in variance example, with the parameters detailed in the main text. The more accurate approximations are given as bold lines, while the simplifications obtainable with large $\gamma$ are depicted with the same linestyle, but with thinner lines.
5. Quickest Change Detection

$R$. On the other hand, once a threshold has been selected, the corresponding pair of detection delay and false alarm rate stays quite close to the optimal curve. This analysis is provided in Figure 5.5, where we run the different tests over a range of detection thresholds, and report in the plot the values of the false alarm rate and of the detection delay, estimated via simulations.

The curves pertaining to the theoretical formulas in Figure 5.5 have been drawn in the following way. The bold solid curve results by combining the first approximation of (5.6) with that in (5.8), while the thinner solid line is obtained when the corresponding expressions for large $\gamma$ are used. As to the bank detector, the bold curve with diamond markers follows by combining (5.23) with the first (more accurate) approximation in (5.22); the rougher (5.24) is depicted as the thinner curve with diamond markers. Finally, as to the single sensor case, the bold dashed curve follows from the first expressions in (5.10) and in (5.11), while the thinner dashed curve is depicts (5.12).

Comfortably, from Figure 5.5 we see that all the simulation points stay very close to the respective theoretical prediction. Figure 5.5 also shows that the small discrepancies between theory and simulation tend to vanish as $R$ decreases, as the arguments exploited in deriving the approximate formulas anticipated. In addition, we note that the operating characteristic of the bank approaches that of a single sensor, when the false alarm rate increases: with large $R$, using $M$ processors organized in a bank architecture provides negligible improvements with respect to the simple single test. On the other hand, for any value of false alarm rate, it is worth noting how the running consensus provides substantial improvements with respect to the bank, in terms of detection performances. We have already pointed out that the situation is reversed in terms of communication burdens: the bank might represent a viable choice whenever the usage of the running scheme is hampered by the energy costs.

In Figures 5.6 and 5.7, we display the theoretical relative efficiencies $\eta_{b,r}(R)$, $\eta_{b,s}(R)$ and $\eta_{b,s}(R)$, as function of the number of sensors $M$, for several values of the false alarm rate $R$. We
Figure 5.5: Operating characteristics of the four systems considered in the work. The detection delay $D$ is depicted as a function of the false alarm rate $R$, for the same scenario of Figure 5.4. As for this latter, thinner lines represent the rougher approximations valid for large $\gamma$. 
see that $\eta_{s,r}(R)$ is always less than unity as we expected, and the curves approach that of $1/M$, also reported for comparison, for small false alarm rates. As a matter of fact, this behavior can be easily predicted by inspecting directly eq. (5.28).

From $\eta_{b,r}(R)$ we observe that the running scheme always outperforms the bank of filters, as the arguments below (5.29) suggested. It is also worth noting how $\eta_{b,r}(R)$ approaches zero for any choice of $R$, thus revealing how the fusion of the sensors’ data during the detection test significantly improves the performances.

The behavior of $\eta_{b,s}(R)$ in Figure 5.7 reveals that the bank of Page’s processor exhibits an optimum value of $M$ at which the detection performances are maximized, when compared with those of a single sensor. There also exists a maximum value of $M$ beyond which $\eta_{b,s}(R)$ falls below unity: interestingly, for moderately large $M$ there is no benefit at all in using a bank of Page’s detectors with respect to a single unit, in terms of system performances. As an example, with a false alarm rate of $10^{-4}$, it turns out that a bank made of only 5 filters provides the best performance improvements with respect to a single filter. The relative efficiency decreases for $M > 5$ and at about $M = 220$ falls below unity, meaning that a bank of more than 220 filters performs worse than a single detection unit.

However, this limiting value of $M$ seems to increase significantly with decreasing false alarm rate, falling in a region where the number of sensors might be prohibitively large for practical purposes. We see that, at false alarm rates below $10^{-5}$ it is always convenient to use the bank, and the optimal $M$ that maximizes the performances takes values in the hundreds.

Note, finally, that the curves in Figures 5.6 and 5.7 are obtained with the more accurate expressions derived in the previous sections. Actually, using the simpler relationships (5.28) and (5.29) in Figure 5.6 would change nothing since the curves would be practically indistinguishable from those given in the figure. As to $\eta_{b,s}$ in Figures 5.7, eq. (5.30) does give the precise behavior of the efficiency but, to address the quantitative analysis of the limiting values of $M$, the more accurate formulas are certainly preferable.
5.5. Examples and Numerical Experiments

Figure 5.6: Relative efficiencies $\eta_{sr}$ and $\eta_{br}$ versus the number of sensors, for several values of the false alarm rate $R$. 

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5_6.png}
\caption{Relative efficiencies $\eta_{sr}$ and $\eta_{br}$ versus the number of sensors, for several values of the false alarm rate $R$.}
\end{figure}
Figure 5.7: Relative efficiency $\eta_{bs}$ versus the number of sensors, for several values of the false alarm rate $R$. 
Chapter 6

Conclusions

The number of scientific works concerning sensor network problems increased in the last few decades. Many applications with new challenges were born and a general definition of sensor network seems impossible\[1\] it seems that a common design goal does exist: To organize a multitude of simple, tiny, cheap, (singly) unreliable, and low-capability sensors under a global architecture, in such a way that --through cooperation-- the aggregate system would behave better than the sum of its constituent.

A key aspect of such systems is the global reliability. Despite of the fact that single sensors can be, and usually are, quite unreliable, the global system is required to be very robust to faults. The network must work properly even in the case that a large part of the system is impaired or destroyed. Leading to one extreme this concept, we would like that all the relevant information collected by the network (the sufficient statistic, in the jargon of inference-oriented sensor networks that we consider in the thesis) would be stored at any single sensor.

To make a concrete example, in a hypothesis testing problem we would like that all the sensors would store the same decision that would be made if all the observations collected by the sensors were available, without any distortion, at a single site. This would

\[1\] In [67] Sadler affirms: “What is a sensor network? We postulate that, given any definition of a sensor network, there exists a counter example”.

85
6. Conclusions

ensure that the “global” decision can be recovered by any single sensor of the system.

In this work we consider networks with a completely flat architecture, in which the nodes cooperate with each other without any hierarchical structure. As a consequence the network can be seen as a single robust entity, where a node failure does not compromise the system functionality. The so-called consensus algorithms, that are key to our development here, were indeed introduced in the scientific community exactly with these goals in mind.

The idea of a distributed algorithm that organizes some agents to reach the consensus originally arises from the social-economic community, where the study was focused on the human behavior. Consensus strategies are applied in many engineering fields [37]:

- computer science;
- systems and control theory;
- distributed signal processing;

in particular there are subjects such as collective behavior of flocks and swarms, sensor fusion, random networks, synchronization of coupled oscillators, algebraic connectivity of complex networks, asynchronous distributed algorithms, formation control for multi-robot systems, optimization-based cooperative control, dynamic graphs, complexity of coordinated tasks, and consensus-based belief propagation in Bayesian networks.

Within the context briefly described above, thesis presents a new paradigm for sensor networks that we call running consensus. It can be thought as a sort of time-varying consensus, in which the remote nodes sense the environment and, at the same time, implement agreement algorithms to corroborate measurements made at different sites. The state of each node acts as a detection (or estimation) statistic to infer the unknown state of the nature.

Typical inference problem, such as estimation, detection and change point detection are considered. When compared with an ideal, centrally computed, detection statistic, the local states are affected by an error term that, as we prove here, converges to zero.
This motivates the investigation of whether the running consensus attains the same performances of such idealized entity, or not. To answer properly a precise asymptotic framework is needed and we develop that in due detail and rigor. Mathematical properties are proved showing the asymptotic equivalence of the running consensus with the idealized system, in terms of detection/estimation capabilities. For what concerns the change detection problem, an asymptotic equivalence is only argued and the formal mathematical study left for future work.

When the running consensus is applied to parameter estimation, the performances are computed in terms of mean square error, and the asymptotic optimality is proved by the bounds on the running consensus statistic. In the fixed-sample size (FSS) detection case, the theory based on Pitman Asymptotic Relative Efficiency (ARE) and on the asymptotic normality of the detection statistics is used. On the other hand, in the sequential case more advanced tools, concerning the asymptotic behavior of sequential decision rules, are used. The convergence toward Wiener processes is required. These theoretical results are the main findings of the thesis, and provide additional practical significance to the running consensus paradigm. In simple terms they emphasize how the decentralized algorithms based on running consensus reach the same performances of the centralized schemes, as long as the system evolves for a sufficiently large amount of time.

A distinct feature of the running consensus scheme is the speed of convergence of the algorithm. In fact, this is substantially different from the exponential law that governs the classic consensus algorithms [12]. Furthermore, in our setup, the specific network topology/connectivity (which rules the system eigenvalues) is less crucial with respect to the classical consensus. The scaling law of running consensus is $n^{-1}$, and only the value of the rate coefficient can be tuned by the eigenvalues. This phenomenon seems in fact related to the rate of data acquisition, which is just $n^{-1}$.

Summarizing, this thesis proposes a new paradigm of consensus, called running consensus, that is particularly suited to applications where the system operates in dangerous environments.
6. Conclusions

The basic idea is to combine the two stages of sensing and of data fusion into a single step. This overcomes design troubles related to the necessity of extending as much as possible the sensing stages, to collect more measurements.

The introduction of the running consensus with the above characteristics represents the main contribution of this work, from a conceptual viewpoint. Technically, the main contribution is in providing a rigorous assessment of the asymptotic convergences of the running consensus, for typical inference applications: binary detection and nonrandom parameter estimation.
Appendix A

A.1 Proof of Proposition 2

Recall that $\mu$ and $\sigma^2$ are the first two moments of $t(x)$, and

$$\xi^3 = E \left[ \|t(x) - \mu 1\|^3 \right]$$  \hspace{1cm} (A.1)

$$e_n = M \sum_{i=1}^{n} \tilde{\Phi}_{n,i} t(x_i) \overset{\text{def}}{=} \sum_{i=1}^{n} y_i(n)$$  \hspace{1cm} (A.2)

where $\tilde{W}_n = W_n - 11^T/M, \forall n \geq 1$ and $\tilde{\Phi}_{n,i} = \tilde{W}_n \tilde{W}_{n-1} \cdots \tilde{W}_i$.

Let $\lambda_U$ be the the maximum eigenvalue of the matrix $E[\tilde{W}^T \tilde{W}]$, and assume that $\lambda_U < 1$. From the Proposition 1 the error term has zero mean, i.e. $E[e_{n,s}] = 0$ for $s = 1, \ldots, M$.

In (A.2) we defined

$$y_i(n) = \tilde{\Phi}_{n,i} M t(x_i)$$

$$= M \tilde{\Phi}_{n,i} [t(x_i) - \mu 1] \overset{\text{def}}{=} M \tilde{\Phi}_{n,i} \tilde{t}(x_i),$$

where, for ease of notation, the dependence of $\tilde{t}(x)$ upon $\mu$ has been skipped. We now prove some properties of $y_i(n)$, making use of the relationship [12]

$$E \left[ \|\tilde{\Phi}_{n,i} z\|^2 \right] \leq \lambda_U^{(n-i+1)} \|z\|^2.$$  \hspace{1cm} (A.3)

for any vector $z$. Thanks to the independence between the random
connection matrices and the observations \( \mathbf{x}_i \), the above yields

\[
E \left[ \| \mathbf{y}_i(n) \| ^2 \right] \leq M^2 \lambda_U^{(n-i+1)} E \left[ \| \mathbf{t}(\mathbf{x}) \| ^2 \right]. \tag{A.4}
\]

Since \( \| \mathbf{y}_i(n) \| ^2 \geq y_{i,s}^2(n) \), where \( y_{i,s}(n) \) is the \( s \)th entry of vector \( \mathbf{y}_i(n) \), we further have

\[
E \left[ | y_{i,s}(n) | ^2 \right] \leq M^3 \lambda_U^{(n-i+1)} \sigma^2, \tag{A.5}
\]

where we explicited the definition of \( E[\| \tilde{\mathbf{t}}(\mathbf{x}) \| ^2] \).

Now, the maximum eigenvalue of the product \( \Phi_{n,i}^T \Phi_{n,i} \) is equal to unity, \( \Phi_{n,i} \) being doubly stochastic by construction. This implies (see, e.g., [2]) that the maximum eigenvalue of the product matrix \( \Phi_{n,i}^T \Phi_{n,i} \) is less than or equal to unity. Hermitianity of this latter matrix yields, for all vectors \( \mathbf{z} \):

\[
z^T \Phi_{n,i}^T \Phi_{n,i} \mathbf{z} = \| \Phi_{n,i} \mathbf{z} \| ^2 \leq \| \mathbf{z} \| ^2. \tag{A.6}
\]

From the above inequality, and using the definition of \( \mathbf{y}_i(n) \), we can write

\[
E \left[ \| \mathbf{y}_i(n) \| ^3 \right] = M^3 E \left[ \| \Phi_{n,i} \mathbf{t}(\mathbf{x}_i) \| ^3 \right] \\
\leq E \left[ \| \Phi_{n,i} \mathbf{t}(\mathbf{x}_i) \| ^2 \| \mathbf{t}(\mathbf{x}_i) \| \right].
\]

Using the bound in eq. (A.3), we have

\[
E \left[ | y_{i,s}(n) | ^3 \right] \leq M^3 \lambda_U^{n-i+1} E \left[ | \mathbf{t}(\mathbf{x}) | ^3 \right] = M^3 \lambda_U^{n-i+1} \xi^3,
\]

where we further referred to the vector entry \( y_{i,s} \).

Before proving eqs. (2.15) and (2.16), we need the further properties

\[
E [ y_{j,s}(n) y_{k,s}(n) ] = 0, \tag{A.7}
\]
\[
E [ | y_{i,s}(n) | \times y_{j,s}(n) y_{k,s}(n) ] = 0, \tag{A.8}
\]
A.1. Proof of Proposition 2

for all $j \neq k$. We address eq. (A.8), since eq. (A.7) can be proved similarly. Without loss of generality, suppose $k \neq i$. Conditioning on the connection matrices, we have

$$
E \left[ \left| y_{i,s}(n) \right| \times \left| y_{j,s}(n) \right| \times \left| y_{k,s}(n) \right| \left| \tilde{W}_1, \ldots, \tilde{W}_n \right| \right] 
\begin{align*}
= & \quad \left| y_{i,s}(n) \right| \times \left| y_{j,s} \right| \left| \tilde{W}_1, \ldots, \tilde{W}_n \right| \\
\times & \quad M \Phi_{n,k} E \left[ \tilde{t}(x_k) \right] = 0,
\end{align*}
$$

where the matrix $\tilde{\Phi}_{n,k}$ should be intended as the one corresponding to the conditioning matrices $\tilde{W}_1, \ldots, \tilde{W}_n$. Removing the conditioning clearly proves eq. (A.8).

We are now ready to prove eqs. (2.15) and (2.16). Indeed, from eq. (A.2), we know that $e_{n,s} = \sum_{i=1}^{n} y_{i,s}(n)$, yielding

$$
E \left[ e_{n,j}^2 \right] = \sum_{i=1}^{n} E \left[ y_{i,j}^2(n) \right] 
\leq M^3 \sum_{i=1}^{n} \lambda_U^{n-i+1} \sigma^2 
\leq M^3 \frac{\lambda_U}{1 - \lambda_U} \sigma^2,
$$

where we neglected the cross-products thanks to (A.7), and used the bound (A.5). Therefore, eq. (2.15) follows.

Let us switch to the third-order bound (2.16). It is expedient to write

$$
|e_{n,s}|^3 = \left( \sum_{i=1}^{n} y_{i,s}(n) \right)^2 \left| \sum_{i=1}^{n} y_{i,s}(n) \right|.
$$

Application of the triangle inequality to the last factor yields

$$
|e_{n,s}|^3 \leq \left( \sum_{i=1}^{n} y_{i,s}(n) \right)^2 \sum_{i=1}^{n} |y_{i,s}(n)|
\leq \sum_{i,j,k} |y_{i,s}(n)| |y_{j,s}(n)| |y_{k,s}(n)|. \quad (A.9)
$$
Taking expectation yields

\[ E[|e_{n,s}|^3] \leq \sum_{i=1}^{n} E[|y_{i,s}(n)|^3] + \sum_{i=1}^{n} \sum_{j \neq i} E[|y_{i,s}(n)| y_{j,s}(n)^2], \tag{A.10} \]

where the cross-terms corresponding to \( j \neq k \) have been deleted by virtue of eq. (A.8). Suppose now \( i < j \). We have

\[ |y_{i,s}(n)| \leq \|y_i(n)\| = M \|\Phi_{n,j}\tilde{\Phi}_{j-1,i}\tilde{t}(x_i)\| \leq M \|\Phi_{j-1,i}\tilde{t}(x_i)\| = \|y_i(j-1)\| \]

in view of eq. (A.6). This allows writing

\[
\begin{cases}
|y_{i,s}(n)| y_{j,s}^2(n) \leq \|y_i(j-1)\| y_{j,s}^2(n), & \text{for } i < j, \\
|y_{i,s}(n)| y_{j,s}^2(n) \leq |y_{i,s}(n)| \|y_j(i-1)\|^2, & \text{for } i > j.
\end{cases}
\]

Taking expectations yields

\[
\begin{cases}
E[|y_{i,s}(n)| y_{j,s}^2(n)] \leq E[\|y_i(j-1)\|] E[y_{j,s}^2(n)], i < j, \\
E[|y_{i,s}(n)| y_{j,s}^2(n)] \leq E[|y_{i,s}(n)|] E[\|y_j(i-1)\|^2], i > j,
\end{cases}
\tag{A.11}
\]

where we used the fact that \( y_i(j-1) \) and \( y_j(n) \) involve different observations as well as different connection matrices, whence they are statistically independent. The same holds for \( y_i(n) \) and \( y_j(i-1) \).

The bounds in eqs. (A.4) and (A.5), by Jensen’s inequality (being \( \sqrt{x} \) a convex-concave function), give

\[ E[|y_{i,s}(n)|] \leq E[\|y_i(n)\|] \leq M^2 \lambda_U^{(n-i+1)} \sigma. \tag{A.12} \]

Putting now eqs. (A.4), (A.5) and (A.12) into eqs. (A.11) gives

\[
\begin{cases}
E[|y_{i,s}(n)| y_{j,s}^2(n)] \leq M^2 \sigma^3 \lambda_{U,j}^{n-j+1} \lambda_{i,j}^{n-j}, & \text{for } i < j, \\
E[|y_{i,s}(n)| y_{j,s}^2(n)] \leq M^2 \sigma^3 \lambda_{U,j}^{n-j+1} \lambda_{i,j}^{n-j}, & \text{for } i > j.
\end{cases}
\]
A.2. Proof of Theorem 1

Using these inequalities into eq. (A.10) finally yields

\[ \sum_{i=1}^{n} \sum_{j \neq i} E[|y_{i,s}(n)||y_{j,s}^2(n)|] \leq M^2 \sigma^3 \]

\[ \times \left( \lambda_U \sum_{i=1}^{n} \lambda_U^{i-1} \sum_{j=i+1}^{n} \lambda_U^{n-j} + \sum_{i=1}^{n} \lambda_U^{n-i} \sum_{j=1}^{n-i} \lambda_U^{n-j} \right) \]

\[ \leq M^2 \sigma^3 \left( \frac{\lambda_U}{(1 - \sqrt{\lambda_U})^2} + \frac{1}{(1 - \lambda_U)(1 - \sqrt{\lambda_U})} \right), \quad (A.13) \]

where the summations have been bounded with the pertinent geometric series, although they could be evaluated explicitly. The first term at RHS of (A.10) can be easily bounded similarly, and (2.16) follows.

A.2 Proof of Theorem 1

The covariance matrix can be expressed as

\[ C_n = E \left[ \left( e_n + \frac{11^T}{n} s_n \right) \left( e_n^T + s_n^T \frac{11^T}{n} \right) \right] = E[e_n e_n^T] + \sigma^2_{\text{ctr}}(t) 11^T, \quad (A.14) \]

where the last equality follows from straightforward algebra. Now we need to compute the term \( E[e_n e_n^T] \). According to eq. (2.8) and recalling that the matrices \( \tilde{W}_n \)'s are iid and independent of the sensors' observations, we have

\[ E[e_n e_n^T] = \frac{1}{n^2} \sum_{h=1}^{n} E[\tilde{\Phi}_{n,h} E[\tilde{x}_h \tilde{x}_h^T] \tilde{\Phi}_{n,h}^T] \]

\[ = \frac{\sigma^2}{n^2} \sum_{h=1}^{n} E[\tilde{\Phi}_{n,h} \tilde{\Phi}_{n,h}^T] \overset{\text{def}}{=} \frac{\sigma^2}{n^2} \sum_{h=1}^{n} \tilde{\Phi}_{n-h} \overset{\text{def}}{=} \frac{\sigma^2}{n^2} \sum_{h=1}^{n} \tilde{\Phi}_{n-h}, \quad (A.15) \]

having exploited \( E[\tilde{x}_h \tilde{x}_h^T] = (I - 11^T/M) \sigma^2 \).
To elaborate, we need the following results. First, recalling the definition of $\Phi_{n,h},$ we have

$$\Phi_{n-h} = E[\tilde{W}_n \ldots \tilde{W}_{h+1} \tilde{W}_h \tilde{W}^T_h
\tilde{W}^T_{h+1} \ldots \tilde{W}^T_n]$$

$$\begin{aligned}
&\overset{(i)}{=} E \left[ \tilde{W}_1 \ldots \tilde{W}_{n-h} \tilde{W}_{n-h+1} \tilde{W}^T_{n-h+1} \right. \\
&\left. \tilde{W}^T(t-h) \ldots \tilde{W}^T(1) \right]
\end{aligned}$$

$$\begin{aligned}
&\overset{(ii)}{=} E \left[ \tilde{W}_1 \ldots \tilde{W}_{n-h} E[\tilde{W}_{n-h+1} \tilde{W}^T_{n-h+1}] \right. \\
&\left. \tilde{W}^T(n-h) \ldots \tilde{W}^T_{1} \right]
\end{aligned}$$

$$\begin{aligned}
&\overset{(iii)}{=} E \left[ \tilde{W}_1 \ldots \tilde{W}_{n-h} E[\tilde{W} \tilde{W}^T] \right. \\
&\left. \tilde{W}^T_{n-h} \ldots \tilde{W}^T_{1} \right]
\end{aligned}$$

(A.16)

In the above: $(i)$ follows from the independence of the matrices $W_n$’s: for any selection of $n-h$ different time indices, the statistical expectation is the same; $(ii)$ is a straightforward application of the iteration property of conditional expectation, that also exploits again the independence of the $\tilde{W}_n$’s; in $(iii)$ we only simplify the notation since the inner expectation is independent upon the time index.

Furthermore, let us consider an arbitrary (deterministic) vector, say $z_0,$ orthogonal to $1,$ and define $z_1 \overset{def}{=} \tilde{W}^T_1 z_0,$ $z_2 \overset{def}{=} \tilde{W}^T_2 z_1,$ $\ldots,$ $z_{n-h+1} \overset{def}{=} \tilde{W}^T_{n-h+1} z_{n-h}.$ All these vectors still remain orthogonal to $1,$ and we have the following properties: for $k = 1, \ldots, n-h+1$,

$$E \left[ ||z_k||^2 \right] = E \left[ z_{k-1}^T E[\tilde{W} \tilde{W}^T] z_{k-1} \right], \quad (A.17)$$

and, for $k = 1, \ldots, n-h+1,$

$$\lambda_L \ E[||z_k||^2] \leq E \left[ z_k^T E[\tilde{W} \tilde{W}^T] z_k \right] \leq \lambda_U \ E[||z_k||^2]. \quad (A.18)$$
A.2. Proof of Theorem

Equation (A.17) is obvious. As to eq. (A.18), note first that the matrix $E[\tilde{W}\tilde{W}^T]$ has all its eigenvalues equal to those of $E[WW^T]$, but for the eigenvalue 1, which is replaced by 0. Thus, the maximum eigenvalue of $E[\tilde{W}\tilde{W}^T]$ is $\lambda_U$, whence the upper bound in eq. (A.18). Then, since $E[\tilde{W}\tilde{W}^T]$ has a zero eigenvalue with eigenvector $1$ and recalling that $z_k \perp 1$, we can lower bound the quadratic form with the successive eigenvalue of $E[\tilde{W}\tilde{W}^T]$, see e.g., [54, Th. 4.2.2].

Armed with the above results, we can write

$$z_0^T \Phi_{n-h} z_0$$

\begin{align*}
\text{\, (i)} & \quad z_0^T E \left[ \tilde{W}_1 \ldots \tilde{W}_{n-h} E[\tilde{W}\tilde{W}^T] \tilde{W}_{n-h} \ldots \tilde{W}_1 \right] z_0 \\
\text{\, (ii)} & \quad E \left[ z_1^T E \left[ \tilde{W}_2 \ldots \tilde{W}_{n-h} E[\tilde{W}\tilde{W}^T] \tilde{W}_{n-h} \ldots \tilde{W}_1 \right] z_2 \right] \\
\text{\, \ldots} & \quad \ldots \\
\text{\, (iii)} & \quad \leq \lambda_U E \left[ ||z_{n-h}||^2 \right] \\
\text{\, (iv)} & \quad \lambda_U E \left[ z_{n-h-1}^T E[\tilde{W}\tilde{W}^T] z_{n-h-1} \right] \\
\text{\, \ldots} & \quad \ldots \\
\text{\, (v)} & \quad \geq \lambda_U^{n-h+1} ||z_0||^2.
\end{align*}

In the above: (i) is simply eq. (A.16); (ii) and the successive equalities result form the definition of the sequence $z_k$; the upper bound in eq. (A.18) implies the inequality (iii); (iv) is a consequence of
A similar result is obtained by using the lower bound in eq. (A.18), instead of the upper bound. Thus, denoting $z_1$ simply by $z$, one obtains:

$$\lambda_{L}^{n-h+1} \||z||^2 \leq z^T \bar{\Phi}_{n-h+1} z \leq \lambda_{U}^{n-h+1} \||z||^2. \quad (A.19)$$

As $z$ is an arbitrary vector orthogonal to $1$, we can set $z = e_i - 1/M$ in eq. (A.19), where $e_i$ denotes a vector of zeros with only the $i^{th}$ entry equal to 1. This yields

$$\lambda_{L}^{n-h+1}(M - 1)/M \leq (\bar{\Phi}_{n-h})_{ii} \leq \lambda_{U}^{n-h+1}(M - 1)/M. \quad (A.20)$$

Now, as seen in eq. (A.15) the diagonal entries $(C_n)_{ii}$ of the covariance matrix in eq. (A.14) involve just the terms $(\bar{\Phi}_{n-h})_{ii}$, which can be bounded as above. In this way, after straightforward algebra, we finally get

$$\sigma_{ct}^2(n) \left[1 + (M - 1)\psi_L^n\right] \leq (C_n)_{ii} \leq \sigma_{ct}^2(n) \left[1 + (M - 1)\psi_U^n\right] \quad (A.20)$$

and the claim in eq. (2.22) immediately follows.

Similarly, setting $z = e_i - e_j (i \neq j)$ in eq. (A.19), yields

$$\lambda_{L}^{n-h+1} \leq \frac{(\bar{\Phi}_{n-h})_{ii} + (\bar{\Phi}_{n-h})_{jj}}{2} - (\bar{\Phi}_{n-h})_{ij} \leq \lambda_{U}^{n-h+1},$$

from which we obtain

$$M \sigma_{ct}^2(n) \psi_L^n \leq \frac{(C_n)_{ii} + (C_n)_{jj}}{2} - (C_n)_{ij} \leq M \sigma_{ct}^2(n) \psi_U^n.$$ 

Recalling the definition of $\rho_n$, and further using eq. (A.20), the above immediately gives eq. (2.23). The proof is now complete. △
Appendix B

B.1 Proof of Theorem 2

The proof is simple and relies on establishing that $s_{n,j}$ is asymptotically normal:

$$\frac{s_{n,j} - n M \mu(\theta_0)}{\sqrt{n M \sigma(\theta_0)}} \xrightarrow{f_{\theta_0}} N(0, 1),$$

$$\frac{s_{n,j} - n M \mu(\theta_n)}{\sqrt{n M \sigma(\theta_n)}} \xrightarrow{f_{\theta_n}} N(0, 1),$$

with exactly the same normalizing functions as for $s^{(c)}_n$, implying that the two statistics share the same efficacy. We work under $\theta = \theta_0$, the proof being exactly the same for $\theta = \theta_n$. From eq. (2.11) we can write

$$\frac{s_{n,j} - n M \mu(\theta_0)}{\sqrt{n M \sigma(\theta_0)}} = \frac{s^{(c)}_n - n M \mu(\theta_0)}{\sqrt{n M \sigma(\theta_0)}} + \frac{e_{n,j}}{\sqrt{n M \sigma(\theta_0)}}. \tag{B.1}$$

It is easy to show that the second moment of $e_{n,j}$ is uniformly bounded with respect to $n$, see eq. (2.15). Then, direct application of Chebyshev’s inequality ensures that $e_{n,j}/\sqrt{n}$ goes to zero in probability, as $n$ diverges. By observing that, thanks to eqs. (2.29), the centralized statistic is asymptotically normal, and invoking Slutsky’s theorem [68], the claim of the theorem immediately follows.

▽

97
B.2 Proof of Theorem 3

The technical conditions assumed in Theorem 2 essentially cover conditions 1-5 in [51] and are therefore sufficient to conclude that a sequential test based on (2.32), for $T_n = s_n^{(c)}$, with thresholds (2.40), attains asymptotic performances given by eqs. (2.37) and (2.43). In order to prove that the same holds true for $s_{n,j}$, it suffices to prove that the convergence to Wiener in (2.41), as well as the additional condition (2.42), hold true with $s_n^{(c)}$ replaced by $s_{n,j}$. We have

$$\frac{s_{[r]j} - [rt] M \eta_r}{\sqrt{r M \sigma(\theta_0)}} = \frac{s_{[r]j} - [rt] M \eta_r}{\sqrt{r M \sigma(\theta_0)}} + \frac{e_{[r]j}}{\sqrt{r M \sigma(\theta_0)}}.$$  

Was the involved convergence that between random variables, it would suffice to prove that the last term converges to zero in probability. However, here we work with convergence of random process, so that the additional requirement of tightness [52] is needed. This condition (which is reminiscent of an uniform convergence with respect to $t$) can be expressed in a simple way in our case (see lemma A1 in [69]):

$$\max_{k \leq [rT]} \frac{e_{k,j}}{\sqrt{r M \sigma(\theta_0)}} \rightarrow 0 \quad \text{for any} \; T > 0,$$

where the convergence is now in probability, under the pertinent distribution. Working for instance under $\theta_r$ (the case $\theta_0$ is managed similarly), this condition is verified, in that, by joint appli-
B.2. Proof of Theorem 3

cation of the union bound and Markov inequality:

\[
P_{\theta_0}\left[ \max_{k \leq [rT]} \frac{e_{k,j}}{\sqrt{rM} \sigma(\theta_0)} > \epsilon \right]
\leq \sum_{k=1}^{[rT]} P_{\theta_0}\left[ \left| \frac{e_{k,j}}{\sqrt{rM} \sigma(\theta_0)} \right|^3 > \epsilon^3 \right]
\leq \frac{1}{\epsilon^3 \sqrt{M} \sigma^3(\theta_0)} \sum_{k=1}^{[rT]} E_{\theta_0}\left[ |e_{k,j}|^3 \right]
\leq \frac{C_1(M, \lambda_U) \xi^3(\theta_r) + C_2(M, \lambda_U) \sigma^3(\theta_r) \left[ rT \right]}{\epsilon^3 \sqrt{M} \sigma^3(\theta_0)}
\]

where the last inequality follows by eq. (2.16) proved in Appendix A.1

such that convergence to zero is proved.

Let us switch now to the additional conditions (2.42). We address only the first one, the other being treated similarly. We can write

\[
P_{\theta_0}\left[ \frac{s_{[r \ell],j} - [r \ell] M \mu(\theta_0)}{\sqrt{rM} \sigma(\theta_0)} > \epsilon \ell \right]
\leq P_{\theta_0}\left[ \frac{s_{[r \ell]} - [r \ell] M \mu(\theta_0)}{\sqrt{rM} \sigma(\theta_0)} > \frac{\epsilon \ell}{2} \right]
+ P_{\theta_0}\left[ \frac{e_{[r \ell],j}}{\sqrt{rM} \sigma(\theta_0)} > \frac{\epsilon \ell}{2} \right]
\leq g_c(t/2) + \frac{4}{[\epsilon t M \sigma(\theta_0)]^2} E_{\theta_0}\left[ |e_{[r \ell],j}|^2 \right]
\leq g_c(t/2) + \frac{4C_1(M, \lambda_U)}{[\epsilon t M]^2},
\]

where the last bound follows by jointly applying Chebyshev inequality and eq. (2.15). The integrability of the RHS in the last equation trivially follows by that of \( g_c(t) \) and of the function \( 1/t^2 \).

\( \nabla \)
Bibliography

[1] P. Braca, S. Marano, and V. Matta, “Running consensus in wireless sensor networks,” in Proc. of the Intern. Conf. on Information Fusion (FUSION 2008), Cologne, Germany, June 30 - July 3, 2008, pp. 152–157.

[2] ——, “Enforcing consensus while monitoring the environment in wireless sensor networks,” IEEE Trans. Signal Process., vol. 56, no. 7, pp. 3375–3380, 2008.

[3] P. Braca, S. Marano, V. Matta, and P. Willett, “Decentralized asymptotic detection by running consensus,” in Signal Processing Advances in Wireless Communications (SPAWC 2009), Perugia, Italy, June, 21-24 2009, pp. 41–45.

[4] ——, “Asymptotic optimality of running consensus in testing statistical hypotheses,” IEEE Trans. Signal Process., vol. 58, no. 2, pp. 814–825, 2010.

[5] ——, “Consensus-based Page’s test in sensor networks,” Signal Processing, vol. 91, no. 4, pp. 919–930, Apr. 2011.

[6] ——, “Quickest distributed detection via running consensus,” in Proc. of the 19th European Signal Processing Conference 2011 (EUSIPCO 2011), Barcelona, Sep. 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, 2011.
[8] F. Cattivelli and A. Sayed, “Distributed detection over adaptive networks using diffusion adaptation,” *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 1917–1932, 2011.

[9] S. Kar, R. Tandon, H. Poor, and S. Cui, “Distributed detection in noisy sensor networks,” in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2011, pp. 2856–2860.

[10] 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,” *Signal Processing, IEEE Transactions on*, vol. 60, no. 11, pp. 5987–6002, 2012.

[11] E. Lehmann, *Testing Statistical Hypotheses*. Springer, 2005.

[12] S. A. Kassam, *Signal Detection in Non-Gaussian Noise*. Springer-Verlag, 1987.

[13] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. New York: Springer-Verlag, 1998.

[14] D. Jakovetic, J. M. F. Moura, and J. Xavier, “Distributed detection over noisy networks: Large deviations analysis,” *Signal Processing, IEEE Transactions on*, vol. 60, no. 8, pp. 4306–4320, 2012.

[15] S. Kar and J. Moura, “Consensus + innovations distributed inference over networks: cooperation and sensing in networked systems,” *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 99–109, 2013.

[16] G. Battistelli, L. Chisci, C. Fantacci, A. Farina, and A. Graziano, “Consensus CPHD filter for distributed multitarget tracking,” *IEEE J. Sel. Topics Signal Process.*, 2013.

[17] R. Olfati-Saber, “Distributed Kalman filter with embedded consensus filters,” in *Proc. of the 44th IEEE Conference on Decision & Control*, 2005, pp. 8179–8184.
[18] P. Braca, S. Marano, V. Matta, and P. Willett, “Multitarget-multisensor ML and PHD: Some asymptotics,” in Proc. of the 15th Intern. Conf. on Inform. Fusion (FUSION), Singapore, 2012.

[19] ——, “Asymptotic efficiency of the PHD in multitarget/multisensor estimation,” IEEE J. Sel. Topics Signal Process., 2013.

[20] L. Tong, Q. Zhao, and S. Adireddy, “Sensor networks with mobile agents,” in Proceedings of MILCOM 2003, vol. 1, Boston MA, Oct. 2003, pp. 688–693.

[21] Z. Yang and L. Tong, “On the error exponent and the use of ldpc codes for cooperative sensor networks with misinformed nodes,” IEEE Trans. Inf. Theory, vol. 53, no. 9, pp. 3265–3274, Sep. 2007.

[22] S. Marano, V. Matta, P. Willett, and L. Tong, “DOA estimation via a network of dumb sensors under the SENMA paradigm,” IEEE Signal Process. Lett., vol. 12, no. 10, pp. 709–712, Oct. 2005.

[23] ——, “Support-based and ML approaches to DOA estimation in a dumb sensor network,” IEEE Trans. Signal Process., vol. 54, no. 4, pp. 1563–1567, Apr. 2006.

[24] A. Gersho, “Asymptotically optimal block quantization,” IEEE Trans. Inf. Theory, vol. 25, no. 4, pp. 373–380, Jul. 1979.

[25] H. Gish and J. Pierce, “Asymptotically efficient quantizing,” IEEE Trans. Inf. Theory, vol. 14, no. 5, pp. 676–683, Sep. 1968.

[26] J. A. Gubner, “Distributed estimation and quantization,” IEEE Trans. Inf. Theory, vol. 39, no. 4, pp. 1456–1459, Jul. 1993.
[27] P. Ishwar, R. Puri, S. Pradhan, and K. Ramchandran, “On compression for robust estimation in sensor networks,” in Proceedings of IEEE International Symposium on Information Theory (ISIT 2003), 2003, p. 193.

[28] S. S. Pradhan and K. Ramchandran, “Distributed source coding using syndromes (discus): design and construction,” IEEE Trans. Inf. Theory, vol. 49, no. 3, pp. 626–643, Mar. 2003.

[29] A. Rao, D. Miller, K. Rose, and A. Gersho, “Generalized vector quantization: Jointly optimal quantization and estimation,” in Proceedings of the IEEE International Symposium on Information Theory (ISIT 1995), 1995, p. 423.

[30] P. Venkitasubramaniam, L. Tong, and A. Swami, “Quantization for maximin are in distributed estimation,” IEEE Trans. Signal Process., vol. 55, no. 7, pp. 3596–3605, 2007.

[31] T. Berger, Z. Zhang, and H. Viswanathan, “The ceo problem,” IEEE Trans. Inf. Theory, vol. 42, no. 3, pp. 887–902, May.

[32] H. Zhu, I. Schizas, and G. Giannakis, “Power-efficient dimensionality reduction for distributed channel-aware kalman tracking using wsns,” IEEE Trans. Signal Process., vol. 57, no. 8, pp. 3193–3207, Aug. 2009.

[33] S. Marano, V. Matta, and P. Willett, “Some approaches to quantization for distributed estimation with data association,” IEEE Trans. Signal Process., vol. 53, no. 3, pp. 885–895, Mar. 2005.

[34] P. Braca, M. Guerriero, S. Marano, V. Matta, and P. Willett, “Distributed estimation with data association: Is the nearest neighbor the most informative?” in Proceedings of Intern. Conf. on Information Fusion (FUSION 2009), Seattle WA, USA, Jul. 2009, pp. 780–785.
[35] ——, “Selective measurement transmission in distributed estimation with data association,” *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4311–4321, Aug. 2010.

[36] M. DeGroot, “Reaching a consensus,” *J. American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.

[37] R. Olfati-Saber, A. Fax, and R. M. Murray, “Consensus and cooperation in networked multi-agent systems,” *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[38] W. Ren, R. W. Beard, and E. M. Atkins, “A survey of consensus problems in multi-agent coordination,” in *American Control Conference*, Portland, OR, USA, Jun. 2005, pp. 1859–1864.

[39] N. A. Lynch, *Distributed Algorithms*. San Mateo, CA, USA: Morgan Kaufmann Publishers, 1997.

[40] P. F. Swaszek and P. Willett, “Parley as an approach to distributed detection,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 1, pp. 447–457, Jan. 1995.

[41] J. Tsitsiklis and M. Athans, “Convergence and asymptotic agreement in distributed decision problems,” *IEEE Trans. Autom. Control*, vol. 29, pp. 42–50, Jan. 1984.

[42] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, “Randomized gossip algorithms,” *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.

[43] S. S. Boyd, A. Ghosh, and B. S. D. Prabhakar, “Gossip algorithms: Design, analysis and applications,” in *Proc. of INFOCOM*, Miami, USA, March, 13-17 2005, pp. 1653–1664.

[44] Y.-W. Hong, A. Scaglione, and P. K. Varshney, “A communication architecture for reaching consensus in decision for a large network,” in *Proc. of the IEEE/SP 13th Workshop on Statistical Signal Processing*, July, 17-20, 2005, pp. 1220–1225.
[45] Y.-W. Hong, L. F. Cheow, and A. Scaglione, “A simple method to reach detection consensus in massively distributed sensor networks,” in Proc. of the International Symposium on Information Theory (ISIT), Chicago, IL, USA, June 27 - July 2, 2004.

[46] Z. Jin and R. M. Murray, “Multi-hop relay protocols for fast consensus seeking,” in Proc. of the 45th IEEE Conference on Decision & Control, vol. 1, San Diego, CA, USA, Dec., 13-15, 2006, pp. 1001–1006.

[47] S. Barbarossa and G. Scutari, “Bio-inspired sensor network design,” IEEE Signal Process. Mag., vol. 24, no. 3, pp. 26–35, May 2007.

[48] A. Kashyap, T. Basar, and R. Srikant, “Quantized consensus,” in Proc. of the International Symposium on Information Theory (ISIT), Seattle, USA, July, 9-14, 2006, pp. 635–639.

[49] C. C. Moallemi and B. Van Roy, “Consensus propagation,” IEEE Trans. Inf. Theory, vol. 52, no. 11, pp. 4753–4766, Nov. 2006.

[50] M. Basseville and I. V. Nikiforov, Detection of Abrupt Changes: Theory and Application. Englewood Cliffs, N.J: Prentice-Hall, 1993.

[51] T. L. Lai, “Pitman efficiencies of sequential tests and uniform limit theorems in nonparametric statistics,” The Annals of Statistics, vol. 6, no. 5, pp. 1027–1047, 1978.

[52] P. Billingsley, Convergence of Probability Measures, 2nd ed. New York: Wiley-Interscience, 1999.

[53] T. Cover and J. Thomas, Elements of Information Theory. New York: John Wiley & Sons, 1991.

[54] R. Horn and C. Johnson, Matrix Analysis. Cambridge, UK: Cambridge University Press, 1985.
[55] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*. Englewood Cliffs, New Jersey: Prentice Hall, 1998.

[56] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988.

[57] A. Wald, *Sequential Analysis*. New York: Dover, 1947.

[58] E. J. G. Pitman, “Lecture notes on nonparametric inference,” *Columbia Univ.*, 1949.

[59] G. E. Noether, “On a theorem of Pitman,” *Ann. Math. Statist.*, vol. 26, pp. 64–68, 1955.

[60] H. Van Trees, *Detection, Estimation, and Modulation Theory. Part I*. New York: John Wiley & Sons, Inc., 1968 (reprinted, 2001).

[61] A. Wald and J. Wolfowitz, “Optimum character of the sequential probability ratio test,” *The Annals of Statistics*, vol. 19, no. 3, pp. 326–339, Sep. 1948.

[62] T. Aysal, M. Yildiz, and A. Scaglione, “Broadcast gossip algorithms: Design and analysis for consensus,” in *47th IEEE Conference on Decision and Control, 2008 (CDC 2008)*, 9-11 Dec. 2008, pp. 4843–4848.

[63] E. Page, “Continuous inspection schemes,” *Biometrika*, vol. 41, pp. 100–115, Jan. 1954.

[64] S. Marano, V. Matta, and P. Willett, “Sequential detection of almost-harmonic signals,” *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 395–406, Feb. 2003.

[65] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Volume 1*. New York: John Wiley & Sons, 1994.
[66] E. Castillo, *Extreme Value Theory in Engineering*. Academic Press, 1988.

[67] B. M. Sadler, “Fundamentals of energy-constrained sensor network systems,” *IEEE Aerosp. Electron. Syst. Mag.*, vol. 20, no. 8, Aug. 2005.

[68] E. Lehmann, *Elements of Large Sample Theory*. Springer, 2004.

[69] W. J. Hall and R. M. Loynes, “Weak convergence of process related to likelihood ratio,” *The Annals of Statistics*, vol. 5, no. 2, pp. 330–341, 1977.
