In this dissertation we focus on decentralized signal processing in Sensor Networks (SN). Four topics are studied: (i) Direction of Arrival (DOA) estimation using a Wireless Sensor network (WSN), (ii) multiple target tracking in large SN, (iii) decentralized target detection in SN and (iii) decentralized sequential detection in SN with communication constraints. The first topic of this thesis addresses the problem of estimating the DOA of an acoustic wavefront using a WSN made of isotropic (hence individually useless) sensors. The WSN was designed according to the SENMA (SEnsor Network with Mobile Agents) architecture with a mobile agent (MA) that successively queries the sensors lying inside its field of view. We propose both fast/simple and optimal DOA-estimation schemes, and an optimization of the MAs observation management is also carried out, with the surprising finding that the MA ought to orient itself at an oblique angle to the expected DOA, rather than directly toward it. We also consider the extension to multiple sources; intriguingly, per-source DOA accuracy is higher when
there is more than one source. In all cases, performance is investigated by sim-
ulation and compared, when appropriate, with asymptotic bounds; these latter
are usually met after a moderate number of MA dwells. In the second topic, we
study the problem of tracking multiple targets in large SN. While these networks
hold significant potential for surveillance, it is of interest to address fundamental
limitations in large-scale implementations. We first introduce a simple analyti-
cal tracker performance model. Analysis of this model suggests that scan-based
tracking performance improves with increasing numbers of sensors, but only to a
certain point beyond which degradation is observed. Correspondingly, we address
model-based optimization of the local sensor detection threshold and the number
of sensors. Next, we propose a two-stage tracking approach (fuse-before-track)
as a possible approach to overcoming the difficulties in large-sensor surveillance,
and we illustrate promising performance results with simulated surveillance data.
The third topic of this dissertation deals with distributed target detection in SN
using Scan Statistics. We introduce a sequential procedure to detect a target
with distributed sensors in a two dimensional region. The detection is carried
out in a mobile fusion center which successively counts the number of binary de-
cisions reported by local sensors lying inside its moving field of view. This is
a two-dimensional scan statistic an emerging tool from the statistics field that
has been applied to a variety of anomaly detection problems such as of epidemics
or computer intrusion, but that seems to be unfamiliar to the signal processing community. We show that an optimal size of the field of view exists. We compare the sequential two-dimensional scan statistic test and two other tests. We also present results for system level detection. In the last topic we study a Repeated Significance Test (RST) with applications to sequential detection in SN. We introduce a \textit{randomly} truncated sequential hypothesis test. Using the framework of a RST, we study a sequential test with truncation time based on a random stopping time. Using the Functional Central Limit Theorem (FCLT) for a sequence of statistics, we derive a general result that can be employed in developing a repeated significance test with random sample size. We present effective methods for evaluating accurate approximations for the probability of type I error and the power function. Numerical results are presented to evaluate the accuracy of these approximations. We apply the proposed test to a decentralized sequential detection in sensor networks (SN) with communication constraints. Finally a sequential detection problem with measurements at random times is investigated.
Statistical Signal Processing in Sensor Networks

Marco Guerriero

B. Sc. in Electrical Engineering, University of Salerno, Italy, 2003
M. Sc. in Electrical Engineering, University of Salerno, Italy, 2005

A Dissertation
Submitted in Partial Fulfilment of the
Requirements for the Degree of
Doctor of Philosophy
at the
University of Connecticut

2009
Statistical Signal Processing in Sensor Networks

Presented by

Marco Guerriero,

Major Advisor

______________________________
Peter K. Willett

Associate Advisor

______________________________
Stefano Coraluppi

Associate Advisor

______________________________
Shengli Zhou

Associate Advisor

______________________________
Lennart Svensson
To Elisabetta
ACKNOWLEDGEMENTS

First and foremost I want to thank my advisor, Professor Peter Willett. With his ideas and fundings, he made my PhD experience productive and unforgettable. He contaminated me with his enthusiasm for research. Without his help it would not have been possible for me to pursue and to complete my Ph.D. His intuition has made him a constant oasis of ideas. His passions for science has inspired and enriched my growth as a student, a researcher and especially as man. I am indebted to him.

I would also like to express my sincere gratitude to Dr. Stefano Coraluppi, who was my supervisor for six months during my visit at NURC. He provided me with many helpful suggestions and important advice. The way Dr. Coraluppi thinks of his research is driven by the Einstein’s quote: "Everything should be made as simple as possible, but not simpler". It was a pleasure to learn this from him. A little thought goes to Stefania for her valuable advice about life.

I would also like to thank Prof. Shengli Zhou, Prof. Y. Bar-Shalom, Prof. Joseph Glaz, Prof. V. Pozdnyakov and Prof. Stefano Marano for many helpful discussions.

I would like to record my gratitude to Prof. Enzo Matta for his supervision, adv-
vice, and guidance from the very early stage of my research. I gratefully acknowledge my friend, Prof. Lennart Svensson. I’m fascinated by his ingenious creativity which made my visit in Sweden very challenging and exciting. I would like to thank my best friend in USA, Dr. Vishal Ravindra. Those long discussion about life and research on his sofa will not be forgotten. I will never forget the Australia Open 2008. We didn’t sleep for two weeks to watch tennis while thinking about life. He amazed me with his sweetness and pleasantness and I really hope that our friendship will never end.

It was great to work and enjoy with my friends and colleagues Özgür Erdinç, Javier Areta, Carolina Bagnato, Bill Donat, Pasquale Di Rubbo, Stefano Baldi and Dave Morizio.

I will never forget the barbecue that Javier prepared in Italy last summer. My heart was warm like the meat on the grill. My regret is that I got to know a brilliant researcher, David Crouse during the end of my stay at Uconn. However we had very productive discussions. German thanks are extended to “Prof. Berger” for many stimulating discussions.

My time at UConn was made enjoyable in large part due to the many friends who are a part of my life. I am grateful for time spent with roommates and friends. Special gratitude goes to Tina for her constant kindness during the course of this study.
I would like to thank my family for all their love and encouragement.

*Un pensiero particolare vola in alto a mio nonno Otello. Grazie per tutto nonno, ti voglio bene.*

I am also indebted to my *italian brothers* Michelangelo, Luigi, Alfonso, Raffaele and Zebbrone.

Last but not least, I would like to thank my loving, supportive, encouraging, and patient Elisabetta. She illuminates me with the bright light of her eyes. Thank you Amore.
## TABLE OF CONTENTS

1. Introduction .................................................. 4
   1.1 Motivation ................................................. 4
   1.2 Publications ............................................. 8

2. Some aspects of DOA estimation using a network of blind sensors 14
   2.1 Introduction ............................................... 14
   2.2 Shaded field of view ....................................... 19
     2.2.1 Simple DOA estimation based on empirical spread ... 23
     2.2.2 Optimal DOA estimation and Fisher information ....... 25
     2.2.3 Performances ........................................... 27
     2.2.4 Optimum rover orientation ............................. 29
     2.2.5 Impact of timing errors ................................. 32
   2.3 Multiple sources .......................................... 33
     2.3.1 Decoupling approach .................................. 33
     2.3.2 Joint Estimation of Multiple DOAs .................... 34
   2.4 Conclusions ................................................ 40
   2.5 Appendix ................................................... 41
     2.5.1 Uncoupled ML estimation ............................... 41
     2.5.2 Joint ML estimation .................................... 42
3. **Fuse-before-Track in Large Sensor Networks**

3.1 Introduction

3.2 Tracker performance modeling

3.2.1 Tracker Model

3.2.2 Tracker Performance Analysis

3.2.3 Optimal Detection Threshold and Number of Sensors

3.2.4 Tuning the Track Management Parameters

3.3 The Fuse-Before-Track Architecture

3.4 Static Fusion

3.5 Fuse-Before-Track Performance Study

3.6 Summary

4. **Distributed Target Detection in Sensor Networks Using Scan Statistics**

4.1 Introduction

4.2 Two-Dimensional Scan Statistics

4.2.1 Definitions

4.2.2 Bernoulli Model
4.2.3 Poisson Model ................................................. 92
4.3 A Two-Dimensional Bernoulli Grid of Sensors .................. 94
4.3.1 Observations Model ............................................ 94
4.3.2 The Scan Statistic ............................................ 98
4.3.3 False Alarm Probability at the MA .......................... 100
4.3.4 Optimized MA Window Size ................................. 102
4.4 A Two-Dimensional Poisson Field of Sensors ................... 104
4.4.1 The Scan Statistic ............................................ 104
4.4.2 The Disjoint-Window Test ................................... 112
4.4.3 The Counting Rule Test ..................................... 113
4.5 Conclusions and Future Directions ............................... 115
4.6 Appendices ..................................................... 116

5. A Repeated Significance Test With Applications To Sequential Detection In Sensor Networks ......................... 123
5.1 Introduction .................................................... 123
5.2 Classical Repeated Significance Test (RST) ..................... 126
5.3 SPRT vs RST in testing for a normal mean ..................... 132
5.4 A Randomly Truncated Test ..................................... 134
5.4.1 Evaluation of $\alpha(t_0, b)$ ................................. 138
5.4.2 Power Function Calculation ................................ 139
5.5 Application: Decentralized Sequential Detection in Sensor Networks

With Communication Constraints . . . . . . . . . . . . . . . . . . . . 142

5.5.1 Problem Formalization . . . . . . . . . . . . . . . . . . . . . . . . . . 142

5.5.2 Energy Constraint by Random Truncation . . . . . . . . . . . . . 143

5.5.3 Minimum Amount of System Resources $M$ . . . . . . . . . . . . . 146

5.5.4 Optimal Censoring Threshold $\tau$ . . . . . . . . . . . . . . . . . . . . 148

5.6 Application: Sequential detection with observations at random times 151

5.7 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 154

Bibliography 156
2.1 The reference scenario. The MA is depicted as a rover and its field
of view as a family of ellipses where the probability that a sensor
is seen stays constant. (Farther sensors have smaller probability
of being seen.) In this example two sources with different DOAs
impinge the network at arbitrary times prior of the MA visit. Note
how the ordering and the relative delays of the impinging times de-
pend upon the DOA. The separate box introduces some definitions:
\( \alpha \) and \( \beta \) measure the ellipses’ major and minor axis, respectively;
\( \theta_i, i = 1, 2 \) represent the two DOAs; and \( \phi_s \) is the MA orientation
at current snapshot \( s \).

2.2 MSE \( E[(\hat{\theta} - \theta)^2] \) versus the total number \( M \) of snapshots. The per-
formance of the estimators proposed in Sects. 2.2.1 and 2.2.2 are
labeled as ‘Rough, \( \phi \) uniform’ and ‘ML, \( \phi \) uniform’, respectively;
the pertinent Fisher proxy is shown as a solid curve (‘CRLB’). The
ML estimator (‘ML, \( \phi \) optimum’) and the Fisher proxy referred to
the optimized rover orientation (‘Optimum CRLB’). For illustrative
purposes, it is also shown the case of \( \phi_s = \hat{\theta} \).
2.3 Notional figure to explain the reason that a MA FOV that is obliquely-angled with respect to true (or best current estimate of true) DOA is best. Oscillating lines represent the relative spread of beep times within a poll and the concentration, for a FOV with a typical number of sensors enclosed. FOV’s are as given in lower part of plot. Note that the derivative of both curves is zero for when FOV is pointed toward true DOA.  

2.4 Uncoupled versus joint ML estimation procedures. Here we have $v = 1\text{ m/s}, \alpha = 1\text{ m}, \beta = 0.3\text{ m}, \lambda = 3\text{ m}^{-2}, \bar{N} \approx 6, \theta = (\theta_1, \theta_2) = (\pi/3, \pi/6)$. The MSE of $\hat{\theta}_1$ is compared to its lower bound and to the performances of the uncoupled approach. Simulations are based on standard Monte Carlo counting process, involving 1000 runs for each point. 

3.1 Markov chain model for tracker logical state, in the target-present case. (A similar Markov chain applies to the target-absent case.) 

3.2 Performance curves for several network-size assumptions. Solid lines characterize sensor performance, while dotted lines characterize tracker performance as characterized by the analytical model in Section 3.2.1. 

3.3 Performance curve obtained solving the optimization problem (3.2.19).
3.4 Performance curves for several network-size assumptions, with adaptive track-management parameters. Solid lines characterize sensor performance, while dotted lines characterize tracker performance as characterized by the analytical model in Section 3.2.1.

3.5 Candidate fusion and tracking architectures.

3.6 FbT includes static fusion and scan-based tracking.

3.7 Illustration of one realization in the simulation-based analysis of single-sensor, (single-stage) multi-sensor and fuse-before-track processing. False contacts are black dots, target-originated contacts are magenta dots, target trajectories are magenta, single-sensor tracks are red, multi-sensor tracks are blue, and FbT tracks are cyan; the (intermediate stage) contact fusion output in FbT processing are cyan dots.

3.8 Close-up view on one target in Figure 3.7. False contacts are black dots, target-originated contacts are magenta dots, target trajectories are magenta, single-sensor tracks are red, multi-sensor tracks are blue, and FbT tracks are cyan; the (intermediate stage) contact fusion output in FbT processing are cyan dots.
3.9 Another example, with same color-coding as above. Note that no single-sensor track is formed, and the multi-sensor track has significantly less track hold ($P_D^T$) than the FbT result.

3.10 Close-up view on one target: a case where larger $n$-scan helps slightly. False contacts are black dots, target-originated contacts are magenta dots, target trajectories are magenta, single-sensor tracks are red, multi-sensor tracks are blue (dotted in case of larger $n$-scan), and FbT tracks are cyan; the (intermediate stage) contact fusion output in FbT processing are cyan dots.

4.1 Scanning the unit time interval with a window of length $w = 0.2$. The * represent times of occurrence of $N = 19$ events, $S_2 = 8$ The bar represents the window that contains the maximal number of events; that is, the scan statistic.

4.2 The MA is depicted with its field of view which we consider to be a square of size $F_{ov}$. Only the sensors which detect the target are depicted. The different level curves correspond to a different target power. The MA traveling across the sensor network simply counts how many sensors are inside its field of view. We assume that the target (the black bold square) is a-priori uniformly distributed within the sensor network.
4.3 A two-dimensional Bernoulli grid of sensors is depicted. \( M = 625 \) sensors are deployed in a grid pattern in the sensor field, which we consider to be a square of area \( b^2 \). The sensors which detect the target (the red bold circle) tend to form a cluster around the target (the black bold square). Here we used \( b = 5, \sigma^2 = 1, \sigma^2_w = 4, p_{fa} = 0.05 \).

4.4 Probability of false alarm for the MA \( P_{fa} \) versus probability of false alarm for the local sensor \( p_{fa} \). Here we have \( b = 5, \sigma^2 = 1, \sigma^2_w = 4, M = 625, k = 7, f_{ov} = 6 \) and \( F_{ov} = 1.2 \).

4.5 The probability of detection \( P_d \) for the MA versus the field-of-view \( F_{ov} \) is plotted for a fixed value of \( \alpha \). Here we have \( b = 5, \sigma^2 = 1, \sigma^2_w = 4, p_{fa} = 0.04, M = 625, \) and \( \alpha = 0.1 \) Simulations are based on 10000 runs.

4.6 The optimal \( F_{ov}^{opt} \) versus \( \sigma \) (\( \sigma^2 \) is the power of the signal that is emitted by the target at distance \( d_s = 1m \)) is plotted for a fixed value of \( \alpha \). Here we have \( b = 5, \sigma^2_w = 4, p_{fa} = 0.04, M = 625, \) and \( \alpha = 0.1 \) Simulations are based on 10000 runs.
4.7 A two-dimensional Poisson field of sensors is depicted. The sensors which
detect the target (the red bold circle) tend to form a cluster around the
target (the black bold square). Here we used $\lambda = 20$, $b = 5, \sigma^2 = 1,$
$\sigma^2_w = 4, \ pfa = 0.05.$ .................................................. 107

4.8 Probability of false alarm for the MA $P_{fa}$ o versus probability of false alarm
for the local sensor $p_{fa}$. Here we have $b = 5, \sigma^2 = 1, \sigma^2_w = 4, \lambda = 20,$
$N = 10, k = 6, f_{ov} = 2$ and $F_{ow} = 1.$ .................................................. 109

4.9 The MA travels across the Poisson sensor field. The overlapping win-
dows (the two-dimensional scan statistic test) and the non-overlapping
windows (non-overlapping test) are depicted. The non-overlapping
test is sensitive to the relative target position with respect to the
scanning window, which turns out to affect the power of the test. 111

4.10 Probability of detection for the MA $P_d$ versus the target location in
the $x$ direction ($x_t$) is plotted for the scan statistic test and the
disjoint-window test respectively. Here we have $b = 5, \sigma^2 = 1,$
$\sigma^2_w = 4, \lambda = 20, N = 25, f_{ov} = 5, F_{ov} = 1, p_{fa} = 0.05,$ and
$\alpha = 0.1$ for both tests (in order to get the prescribed value $\alpha = 0.1$
the use of the randomized test as previous described is needed).
Simulation are based on standard Monte Carlo counting process,
involving 10000 runs. .................................................. 118
4.11 ROC curves obtained by simulations for the scan statistic test and the disjoint-window test. Here we have $b = 5$, $\sigma^2 = 1$, $\sigma_{w}^2 = 4$, $\lambda = 20$, $N = 25$, $f_{ov} = 5$, $F_{ov} = 1$, $p_{fa} = 0.02$. Simulation are based on standard Monte Carlo counting process, involving 10000 runs.

4.12 ROC curves in semilogarithmic scale obtained by simulations for the scan statistic test and the disjoint-window test. Here we have $b = 5$, $\sigma^2 = 1$, $\sigma_{w}^2 = 4$, $\lambda = 20$, $N = 25$, $f_{ov} = 5$, $F_{ov} = 1$, $p_{fa} = 0.02$. Simulation are based on standard Monte Carlo counting process, involving 10000 runs.

4.13 ROC curves in semilogarithmic scale obtained by simulations for the scan statistic test and the counting rule test. Here we have $b = 5$, $M = 625$, $\sigma^2 = 1$, $\sigma_{w}^2 = 4$, $N = 25$, $p_{fa} = 0.05$. Simulation are based on standard Monte Carlo counting process, involving 10000 runs.

4.14 Number of sensors $N_s$ which report their decision to the MA (scan statistic) and to the fusion center (counting rule). Here we have $b = 5$, $M = 625$, $\sigma^2 = 1$, $\sigma_{w}^2 = 4$, $N = 25$, $p_{fa} = 0.05$. Simulation are based on standard Monte Carlo counting process, involving 10000 runs.

5.1 Pictorial sketch of the RST with $\sigma = 1$. 

xviii
5.2 ASN as function of $\mu_1$ is plotted for the SPRT and the RST, respectively. Here we used $n_0 = 1$, $\sigma = 1$. The values of $b$ and $N$ for the RST and the values of $\gamma_1$ and $\gamma_0$ for the SPRT, were chosen such that $\alpha$ and $\beta$ were kept constant to the prescribed nominal value of 0.05.

5.3 Two-level quantization and cost function $g(\cdot)$.

5.4 Sensor network for decentralized sequential detection. Due to bandwidth constraints the observations have to be quantized and then transmitted to a FC which implements a truncated RST. In the unlikely case that the RST crosses the boundaries at the same sensor’s index $i$ for which $\sum_{i \geq 1} g(x_i) \geq M$ the FC decides in favor of $H_1$.

5.5 In $(M, b)$ plane the curve above is a plot of an increasing function.

5.6 Simulated probability of Type II error $\beta^{\text{sim}}$ as function of $M$ is plotted. Simulations are based on 10,000 runs.

5.7 Censoring function when (a) $\mu_1 >> \mu_0$: highly distinguishable hypotheses.

5.8 Censoring function when (b) $\mu_1 \rightarrow \mu_0$: hypotheses are close to each other.
5.9 Arrival process of the observations is described by a one-dimensional Poisson process. The interarrival times are independent and obey the exponential distribution with parameter $\lambda$.  

5.10 The probability of type I error and the power of the test obtained by simulation and by theory, respectively, are plotted as function of $T$. Here, we used, $\lambda = 7$, $n_0 = 40$, $\sigma = 1$, $b = 2.5$ and $\mu_1 = 0.3$. Simulations are based on 10,000 runs.
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Model simulation parameter</td>
<td>56</td>
</tr>
<tr>
<td>3.2</td>
<td>Adaptive track management for the results in Figure 3.2</td>
<td>63</td>
</tr>
<tr>
<td>3.3</td>
<td>Parameters for single-sensor tracking, multi-sensor tracking, and FbT</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>simulation-based performance evaluation.</td>
<td></td>
</tr>
<tr>
<td>3.4</td>
<td>Tracker performance results</td>
<td>72</td>
</tr>
<tr>
<td>3.5</td>
<td>Tracker performance results with two centralized settings.</td>
<td>72</td>
</tr>
<tr>
<td>5.1</td>
<td>Approximation and exact values of $\alpha(t_0, b)$. The performance of the</td>
<td>140</td>
</tr>
<tr>
<td></td>
<td>approximation increases as $b$ becomes larger.</td>
<td></td>
</tr>
<tr>
<td>5.2</td>
<td>Optimization Parameters</td>
<td>148</td>
</tr>
<tr>
<td>5.3</td>
<td>Results</td>
<td>148</td>
</tr>
<tr>
<td>5.4</td>
<td>Optimization Parameters 10,000 MonteCarlo simulations</td>
<td>150</td>
</tr>
<tr>
<td>5.5</td>
<td>Results</td>
<td>150</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Motivation

There has been an explosive growth in the amount of effort devoted to the field of sensor networks (SN) in recent years covering a wide range of areas, from sensor fabrication to protocol development to understanding theoretical issues. The surge in interest is due substantially to the fact that technological development in sensors and communication has made the practical realization of such networks possible. Ultra small sensing devices based on micro-electromechanical or MEMS processes, wireless multi-access technology and GPS-based geolocation are some examples of technologies that have fueled the current interest in the field. SN typically consist of miniature battery-operated sensor nodes with wireless communication capabilities deployed in an ad hoc manner. One of the requirements in such deployments is that the amount of data transmitted be limited both to conserve the typically low energy pool of the autonomous sensors and to conserve bandwidth. Signal processing plays an important part in this process through
application of techniques such as data aggregation, compression, sensor fusion, distributed estimation/detection and target tracking. Such methods allow a certain amount of front-end processing that reduces the amount of transmitted data, thus limiting channel usage, while ensuring that application requirements are still met.

Nowadays sensor networks offer exciting applications in a wide range of fields such as defense, medicine, law enforcement, and space exploration. While significant effort over the last decade in sensor development, physical layer transmission and networking infrastructure has laid the initial groundwork for practical deployment, the full potential for networked sensing systems can only be realized through a fundamental understanding of statistical signal processing in networked and uncertain environments.

The focus of this thesis is on decentralized signal processing in SN. The first topic of this dissertation addresses the problem of estimating the Direction of Arrival (DOA) of an acoustic wavefront using a wireless sensor network (WSN) made of isotropic (hence individually useless) sensors. The WSN was designed according to the SENMA (SEnsor Network with Mobile Agents) architecture with a mobile agent (MA) that successively queries the sensors lying inside its field of view. We propose both fast/simple and optimal DOA-estimation schemes, and an optimization of the MAs observation management is also carried out, with the
surprising finding that the MA ought to orient itself at an oblique angle to the expected DOA, rather than directly toward it. We also consider the extension to multiple sources; intriguingly, per-source DOA accuracy is higher when there is more than one source. In all cases, performance is investigated by simulation and compared, when appropriate, with asymptotic bounds; these latter are usually met after a moderate number of MA dwells.

In the second topic, we study the problem of tracking a single target in large SN. While these networks hold significant potential for surveillance, it is of interest to address fundamental limitations in large-scale implementations. We first introduce a simple analytical tracker performance model. Analysis of this model suggests that scan-based tracking performance improves with increasing numbers of sensors, but only to a certain point beyond which degradation is observed. Correspondingly, we address model-based optimization of the local sensor detection threshold and the number of sensors. Next, we propose a two-stage tracking approach (fuse-before-track) as a possible approach to overcoming the difficulties in large-sensor surveillance, and we illustrate promising performance results with simulated surveillance data.

The third topic of this dissertation deals with distributed target detection in SN using Scan Statistics. We introduce a sequential procedure to detect a target with distributed sensors in a two dimensional region. The detection is carried
out in a mobile fusion center which successively counts the number of binary
decisions reported by local sensors lying inside its moving field of view. This is
a two-dimensional scan statistic an emerging tool from the statistics field that
has been applied to a variety of anomaly detection problems such as of epidemics
or computer intrusion, but that seems to be unfamiliar to the signal processing
community. We show that an optimal size of the field of view exists. We compare
the sequential two-dimensional scan statistic test and two other tests. We also
present results for system level detection.

In the last topic we study a Repeated Significance Test (RST) with applications
to sequential detection in SN. We introduce a randomly truncated sequential hy-
pothesis test. Using the framework of a RST, we study a sequential test with
truncation time based on a random stopping time. Using the Functional Central
Limit Theorem (FCLT) for a sequence of statistics, we derive a general result that
can be employed in developing a repeated significance test with random sample
size. We present effective methods for evaluating accurate approximations for the
probability of type I error and the power function. Numerical results are presented
to evaluate the accuracy of these approximations. We apply the proposed test to a
decentralized sequential detection problem in sensor networks (SN) with commu-
nication constraints. Finally a sequential detection problem with measurements
at random times is investigated.
1.2 Publications

Publications directly related to the dissertation research:

A. Journal Papers


B. Conference Papers


Other publications relevant to the Ph.D. research, but not otherwise reported in this dissertation include:

A. Journal Papers


B. Conference Papers

NATO Undersea Research Centre (NURC), La Spezia, Italy, 15-17 September, 2009.


Chapter 2

Some aspects of DOA estimation using a network of blind sensors

2.1 Introduction

The Direction Of Arrival (DOA) estimation is a key topic in statistical signal processing, with relevant practical impact and well established approaches and methodologies, see e.g., [1–6] and therein references for useful entry points to that literature. Some recent papers [7–10], however, addressed the issue in a specific and relatively new scenario whose main features are: (i) the estimation system consists of a Wireless Sensor Network (WSN) [11–18] with a SENMA (SEnsor Network with Mobile Agents) architecture [19–22]; (ii) sensors (also referred to as nodes) are isotropic, hence completely blind to DOA, and they are only capable of sensing the incoming signal and of recording the impinging time instant; (iii) the Mobile Agent (MA) does not know the number nor the position of the nodes, i.e., sensors are unlabeled and disseminated at random; (iv) nodes have very limited
communication capabilities, since they can only recognize a wake-up signal to switch from sleep to working mode and, when operating in this latter modality, they can transmit short beeps; (v) the sought DOA refers to an acoustic (plane) wavefront, while beeps are electromagnetic. We can summarize this setting by saying that the acoustic-wave DOA estimation must be accomplished by a network of *dumb beepers*.

As to the choice of the SENMA, we stress that the recent growing interests in this paradigm is mainly motivated by some key advantages over alternative architectures (such as multi-hop protocols) in terms of energy savings and scalability [19–22]. In addition, the SENMA scheme is particularly suited to applications where it may be not safe or not convenient to deliver informations towards a fusion center (or exchanging informations among sensors) while the system is sensing the environment. In these cases, separating the environmental monitoring from the estimation stage can be in fact a necessity. In the SENMA, the MA (think of a wheeled rover or an aircraft, for instance) travels across the surveyed area and repeatedly polls the sensors lying inside its field of view. As shown in [7–10], a viable DOA estimation strategy may be accordingly conceived: when polled, each sensor transmits a periodic train of beeps and the aggregate of these signals is what the MA observes. Based on this information, after an appropriate number of network polls, the MA can estimate the sought DOA.
Fig. 2.1: The reference scenario. The MA is depicted as a rover and its field of view as a family of ellipses where the probability that a sensor is seen stays constant. (Farther sensors have smaller probability of being seen.) In this example two sources with different DOAs impinge the network at arbitrary times prior of the MA visit. Note how the ordering and the relative delays of the impinging times depend upon the DOA. The separate box introduces some definitions: $\alpha$ and $\beta$ measure the ellipses’ major and minor axis, respectively; $\theta_i$, $i = 1, 2$ represent the two DOAs; and $\phi_s$ is the MA orientation at current snapshot $s$. 
While we borrow the scenario and the basic ideas from [7–10], the main contributions of this paper is to relax some of the over-idealized assumptions made in these earlier works, and to address some related new issues not considered before:

- A more physically appealing model for the FOV (field of view) is introduced (shaded FOV). Using that, we first propose a very simple DOA estimation method based on the empirical spread of the data. Then, the Maximum Likelihood (ML) estimator is investigated (we loosely call it the optimal estimator) and its asymptotic performances (Fisher information) are analytically derived. We are able to find a closed-form result based on a Gaussian FOV model; despite the relaxation of assumptions, the expressions are simpler than those previously reported.

- We formulate and solve the problem of optimizing the MA’s path (actually, orientation) while polling. We find significant performance enhancement from a rover field of view that is squinted near, but not at, the current DOA estimate.

- The problem of possible inaccuracies in determining the wavefront hitting time is also briefly discussed. Under the assumption of Gaussian timing errors, we still derive the ML-estimator and the Fisher information. The pertinent formulas admit a straightforward interpretation, in the light of those found without accounting for timing inaccuracies.
• We then consider the multiple-source problem in which the correlations arising from the interaction of the planar wavefronts from different sources with the network’s sensors are properly exploited. This requires a slightly more sophisticated communication protocol with respect to the previous approach, but the estimation performances can be significantly improved. An appealing result here is that the accuracy is actually enhanced: the DOA of each of two sources is more closely estimated than it would be if it were the only source.

The reference scenario is depicted in Fig. 2.1, where some symbolisms and notations are also introduced: \( \theta_i \in (0, \pi), \ i = 1, 2, \) are the DOAs of the acoustic waves that impinge at different and arbitrary times on the sensor network; later on, a MA polls the sensors within its field of view. The model of this latter implies that sensors closer to the MA have larger probability of being seen, with respect to farther nodes. At each snapshot the MA’s field of view has a different and arbitrary orientation \( \phi_s \in (0, 2\pi), \) with \( s = 1, 2, \ldots, \) the snapshot index. It is convenient to model such a sequence \( \{\phi_s\}_{s=1}^{\infty} \) as independent realizations of a random variable uniformly distributed in \((0, 2\pi)\), an assumption that we relax in Section 2.2.4. In any case we assume that successive dwells are independent of each other, in the sense that they always involve different sensors\(^1\). Note finally

\(^1\) The independence assumption seems quite reasonable at least for large networks where the successive snapshots can be taken far enough each other. Such assumption, however, seems not
that we assume $\theta_i \in (0, \pi)$ because the methods presented in this paper, as well as those considered in similar studies, are unable to distinguish between the two DOAs $\theta$ and $\theta + \pi$. Such a $\pi$-ambiguity must be resolved by other means.

The remainder of the paper is organized as follows. Section 2.2 introduces the probabilistic field of view and suggests two new inference procedures for DOA estimation: one simple and one optimal. The rover’s optimized path design is also addressed. Section 2.3 considers the multiple DOAs problem, first by decoupling it in its constituent single-DOA estimation problems, and then by addressing the genuinely coupled approach. Conclusions are summarized in Sect. 2.4, while many mathematical details are postponed to the appendices. Part of this work has been presented in [23].

2.2 Shaded field of view

In earlier works, the rover’s field of view was identified as a precisely delimited region surrounding the MA such that all the sensors lying inside are certainly seen by the rover, while sensors lying outside are surely invisible [7–10]. Clearly, accounting for the noisy communication environment, for the possible presence of fading and interferences, for the inhomogeneity of the electromagnetic characteristics of the polled area, etc., the event that a sensor be visible or not is better modeled as having a certain probability $\in (0, 1)$, rather than being on/off. Let very critical: we expect that slight violations will not have a dramatic impact.
us introduce the probability that a sensor lying in the position \((x, y)\) (in a two-dimensional coordinate system with the origin centered on the MA) is visible\(^2\) by the MA. Our choice is\(^3\):

\[
p(x, y) = \exp \left\{ -\frac{1}{2} \left[ \frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} \right] \right\},
\]

where the positive constants \(\alpha\) and \(\beta\) (we assume \(\alpha \geq \beta\)) rule the eccentricity of this probabilistic field of view and its rate of decaying to zero. The level curves of the above Gaussian-shaped probability are ellipses, see Fig. 2.1: this is a natural generalization of the on/off model earlier considered.

The displacement of the sensors obeys a homogeneous Poisson distribution with parameter \(\lambda\) that represents the average number of sensors per unit area. The distribution of the visible sensors in a single snapshot, however, results in an inhomogeneous process, as consequence of the shape \(p(x, y)\) of the MA’s field of view. Specifically, the distribution of the sensors seen (i.e., polled and heard) by the travelling rover forms an inhomogeneous spatial Poisson process such that the probability that one such sensor lies within an elementary area of size \(dx dy\) is \(\lambda p(x, y) dx dy\). Let \(N_s\) be the number of sensors seen at snapshot \(s\). We have [24]:

---

\(^2\) This means that the node receives the polling (wake-up) signal emitted by the MA and that the MA receives the consequent sensor’s signaling.

\(^3\) We stress that such choice is only for concreteness and ease of mathematical computation: other assumptions would yield different analytical expressions but would otherwise have little impact on our mode of analysis.
\( P \{ N_s = k \} = \bar{N}^k e^{-\bar{N}}/k! \), where \( \bar{N} = 2\lambda\pi\alpha\beta \) is the average number of sensors seen by the MA in any snapshot. On accounting for eq. (2.2.1), given \( N_s \), the location of each visible sensor is a Gaussian-distributed two-dimensional random vector, see e.g., [24].

Let us assume that a single acoustic wave impinges on the network, and let \( \tau_i \) be the time instant that the \( i^{th} \) sensor is hit by the wavefront. At the \( s^{th} \) snapshot corresponds the vector of hitting times \( \tau_s = (\tau_1, \tau_2, \ldots, \tau_{N_s}) \). In parallel to the arguments in [7, 8], one can imagine that a continuum of sensors lies inside the surveyed area. Then, the pdf of the vector \( \tau_s \) can be computed as the line integral of \( p(x, y) \) in eq. (2.2.1) along the acoustic wavefront lines, see Fig. 2.1. According to the chosen random sensor displacement and visibility, the measured times \( \tau_i \)s are independent samples from that density. This latter is computed in Appendix 2.5.1, and results in the following Gaussian density:

\[
f_{\tau_i}(\tau) = \frac{1}{\sigma_s \sqrt{2\pi}} e^{-\frac{\tau^2}{2\sigma_s^2}}, \quad \sigma_s^2 = \frac{a_s^2 \alpha^2 + \beta^2}{(1 + a_s^2)v^2}, \tag{2.2.2}
\]

where \( v \) is the velocity of sound, \( a_s = \tan(\theta - \phi_s) \), and \( \phi_s \) is the orientation of the MA at snapshot \( s \).

To collect information from the sensors, the MA observes a certain region of the surveyed area for a time \( T \). A reasonable choice for this parameter is to set it as a multiple of \( 2\alpha/v \) (recall that \( \alpha \) is the largest ‘standard deviation’ of the Gaussian-like shape \( p(x, y) \), and \( v \) is the velocity of the acoustic wave). To fix ideas, let
$T = 12\alpha/v$ so that we are neglecting the possibility that a sensor farther than $6\alpha$
is seen by the MA: this happens, indeed, with negligible probability.

Let $\tau_i$ be the time instant that the $i^{th}$ sensor is impinged by the acoustic wavefront.

At time $\tau_i$, the $i^{th}$ sensor starts an internal clock of period $T$. We say that the
node *virtually* emits a periodic signal, say $z_i(t) = \sum_k d(t - \tau_i - kT)$, where $d(t)$ is a convenient basic pulse (say, delta-like). For obvious reasons of energy saving, this emission is only virtual and is only aimed at recording the time instant $\tau_i$ of the hit on sensor $i$.

Only when polled by the rover (that may visit the sensor much later than the acoustic wavefront), the signal $z_i(t)$ is really emitted as an electromagnetic waveform. As detailed in [9], upon observing $z_i(t)$, the MA is able to recover the vector of the hitting times, but for a time reference. That is to say, the observables $\tau_i$s are known but for the time origin or, otherwise stated, the MA can only measure $\tau_i - \tau_0$, with $\tau_0$ arbitrary. In this respect, it should be stressed that expression (2.2.2) follows from assuming that $\tau_i$s be measured by setting the time origin in correspondence of the MA. The lack of knowledge about the time origin implies that the actual pdf would have unknown average value (arbitrarily set to 0 in eq. (2.2.2)). Due to the lack of time reference, at each snapshot the MA can only extract the $N_s - 1$ *relative times* taken with respect to (say) $\tau_1$: $t_s = (t_2, t_3, \ldots, t_{N_s})$, where we define $t_i = \tau_i - \tau_1$, $i \in (2, \ldots, N_s)$. 
We would like to stress, finally, that the communication between sensors and MA is electromagnetic, while the impinging wave is acoustic. In fact, the estimation of an electromagnetic source seems a far more complicated task, see also [8].

2.2.1 Simple DOA estimation based on empirical spread

The estimate of $\sigma_s^2$ in eq. (2.2.2) can be computed using the $t_s$. Indeed:

$$\hat{\sigma}_s^2 = \frac{1}{N_s} \sum_{i=1}^{N_s} \left( \tau_i - \frac{1}{N_s} \sum_{k=1}^{N_s} \tau_k \right)^2 = \frac{1}{N_s} \sum_{i=1}^{N_s} \left( t_i - \frac{1}{N_s} \sum_{k=1}^{N_s} t_k \right)^2$$

which is nothing but an ML estimator and, by definition, $t_1 = 0$. It should be also stressed that this simple estimator can be used for a large class of fields of view, other than that in eq. (2.2.1), as long as the sample variance is a reliable measure of the sample spread.

Once, at snapshot $s$, $\hat{\sigma}_s^2$ has been obtained, a simple rough estimator can be conceived as follows. Following the method of moments (see, e.g., [25]), from eq. (2.2.2) we get an estimate of $a_s$ in the form

$$\hat{a}_s = \pm \sqrt{\frac{v^2 \hat{\sigma}_s^2 - \beta^2}{\alpha^2 - v^2 \hat{\sigma}_s^2}}$$

provided that the admissibility condition $\beta^2 \leq v^2 \hat{\sigma}_s^2 < \alpha^2$ is fulfilled. From the

---

4 Formally, it is so provided that, in estimating $\sigma_s^2$ from $\tau$, the unknown mean value of the Gaussian distribution (2.2.2) is modelled as a deterministic nuisance parameter [25]. A similar form of the estimator is also obtained by solving the ML equation using $t_s$ as observables.
definition of \( a_s \), two values of DOA result:

\[
\hat{\theta}_{\text{rough},s,1} = [\phi_s + \arctan(|\tilde{a}_s|)] \mod \pi, \\
\hat{\theta}_{\text{rough},s,2} = [\phi_s - \arctan(|\tilde{a}_s|)] \mod \pi.
\]

This ambiguity (i.e., the presence of two candidate DOAs) is implicit in the geometry of the system (see Fig. 2.1) in that two symmetric DOAs with respect to the MA orientation \( \phi_s \) cannot be distinguished in a single snapshot: we have a spurious mirror DOA along with the true \( \theta \). Successive snapshots \( s = 1, 2, \ldots, M \), result in \( 2M \) partial estimates, to be properly fused for the final estimator, say \( \hat{\theta}_{\text{rough}} \), and the fusion rule should hopefully remove the ambiguity. Note that an inherent sub-optimality of the method is just the post-estimation fusion.

As a fusion rule, we simply divide the \((0, \pi)\) interval in a sufficiently large number of bins, and compute the relative amount (out of \( 2M \)) of partial estimates that belong to each bin. The bin containing the largest number of samples is selected, and the arithmetic mean of the partial estimates that belong to it is defined as \( \hat{\theta}_{\text{rough}} \). The idea is that the \( M \) mirror DOAs are uniformly distributed in \((0, \pi)\), as a consequence of the uniform orientation of the MA. The remaining \( M \) partial estimates tend to crowd the bin containing the true DOA\(^5\).

\(^5\) Clearly, the \( 2M \) partial estimates are unlabelled: we cannot simply separate the \( M \) true values from the \( M \) mirror ones. Also, to avoid misunderstanding, we stress that the focus here is on the simplicity of the method and accordingly, we make no effort to optimize the described...
2.2.2 Optimal DOA estimation and Fisher information

Recall that $\tau_s$ is made of independent entries $\tau_i$: we have $\tau_s \sim f_{\tau_s}(\tau_s) = \prod_{i=1}^{N_s} f_{\tau_i}(\tau_i)$, with $f_{\tau_i}(\tau_i)$ given by eq. (2.2.2). Times collected in successive snapshots are independent, but not identically distributed, as eq. (2.2.2) depends on $s$ through $\sigma_s$.

Since we cannot observe $\tau_s$, but only $t_s$, we need the statistical characterization of the latter. Consider hence the $N_s$ random variables $(\tau_1, t_2, t_3, \ldots, t_{N_s}) = [\tau_1, t_s]$. Their joint density is $F(\tau_1) = f_{[\tau_1, t_s]}(\tau_1, t_2, t_3, \ldots, t_{N_s}) = f_{\tau_s}(\tau_1, t_2 + \tau_1, t_3 + \tau_1, \ldots, t_{N_s} + \tau_1)$. We have

$$F(\tau_1) = \prod_{i=1}^{N_s} \frac{1}{\sigma_s \sqrt{2\pi}} \exp\left\{ -\frac{(t_i + \tau_1)^2}{2\sigma_s^2} \right\}$$

(2.2.3)

with $-\infty < \tau_1 < \infty$, and $f_{t_s}(t_s) = \int_{-\infty}^{+\infty} F(\tau_1) d\tau_1$.

Integrating eq. (2.2.3) with respect to $\tau_1$ amounts to marginalize an $N_s$-dimensional multivariate Gaussian density. The result, detailed in Appendix 2.5.1, is

$$f_{t_s}(t_s) = \exp\left( -\frac{1}{2} t_s^T C^{-1} t_s \right) \frac{1}{(2\pi)^{N_s/2} (\det C)^{1/2}}$$

(2.2.4)

where $C = \sigma_s^2 Z$ and where we have defined the $(N_s - 1) \times (N_s - 1)$ covariance matrix $Z = I + O$, with $I$ the identity matrix and $O$ the matrix with all entries equal 1. Clearly, $C$ depends upon the DOA $\theta$ only through $\sigma_s^2$.

estimation procedure. Furthermore, a nontrivial issue is the selection of the bin width, even though the finite support of allowable $\theta$ mitigates the problem. We have found that about 50 bins can be adequate for the examples addressed in this paper.
A closed form for the ML estimator of $\theta$ is not so immediate to derive as it may appear at a first glance, because the functional dependence of $\sigma_s^2$ from $\theta$ changes snapshot by snapshot, as a consequence of the different orientation $\phi_s$, see eq. (2.2.2). The numerical maximization of the likelihood is simpler, and we use that. The likelihood corresponding to $M$ independent snapshots results in the product of the individual likelihoods and the ML DOA estimator is

$$\hat{\theta}_{ML} = \arg \max_{\theta} \prod_{s=1}^{M} f_{t_s}(t_s). \quad (2.2.5)$$

Assuming that mild technical conditions are met, as $M$ grows, $\hat{\theta}_{ML}$ attains its asymptotic properties of unbiasedness and efficiency [26]: $E[\hat{\theta}_{ML}] \to \theta$, and $\text{VAR}[\hat{\theta}_{ML}] \to I_M^{-1}(\theta)$. Here $I_M(\theta) = \sum_{s=1}^{M} J_s(\theta)$ is the $M$-snapshots Fisher information with respect to $\theta$, $J_s(\theta)$ is the Fisher information at snapshot $s$, and we add these informations in view of the independence of the dwells. In Appendix 2.6.1 we find

$$I_M(\theta) = M(\bar{N} - 1 + e^{-\bar{N}}) \left( \frac{\sqrt{1 - \psi^2} - 1)^2}{\sqrt{1 - \psi^2}} \right), \quad (2.2.6)$$

where $M$ is the total number of snapshots, $\bar{N}$ is the average number of sensors per single snapshot, and $\psi = \sqrt{1 - (\beta/\alpha)^2}$ represents the eccentricity of the ellipses’ family, namely the eccentricity of the MA’s field of view.

The functional dependence of the estimation from the relevant model parameters is instructively summarized by eq. (2.2.6). We see that $I_M(\theta)$ stays constant with
\( \theta \): all the DOAs are equally difficult to estimate. Also, \( I_M(\theta) \) grows linearly with \( M \) and approximately linearly with \( \bar{N} \): for \( \bar{N} \) larger than few units, the Fisher information simply depends upon the product \( M(\bar{N} - 1) \), which is the effective total number of timing informations collected by the MA (recall that only the time differences are informative, whence the \( \bar{N} - 1 \)). Finally, not unexpectedly, \( I_M(\theta) \) is a monotonically increasing function of the eccentricity parameter \( \psi \).

2.2.3 Performances

Using standard Monte Carlo simulations, we now evaluate the performances of the proposed estimators, to be compared with the inverse of the Fisher information (see eq. (2.2.6)), the Cramér-Rao Lower Bound (CRLB) [27]. The results of extensive simulations and different case studies can be conveniently summarized by one specific example, as reported in Fig. 2.2, where we set \( v = 1 \) m/s, \( \alpha = 1 \) m, \( \beta = 0.3 \) m, \( \lambda = 10 \) m\(^{-2} \), resulting in \( \bar{N} \approx 19 \). Each simulated point involves 1000 program runs. The general comments are that the ML estimator approaches very quickly the CRLB as the number of snapshots increases. Remarkably, the MSE pertaining to the rough estimator seems to scale almost properly with the number of snapshots, but with a suboptimal scaling coefficient.
Fig. 2.2: MSE $E[(\hat{\theta} - \theta)^2]$ versus the total number $M$ of snapshots. The performance of the estimators proposed in Sects. 2.2.1 and 2.2.2 are labeled as ‘Rough, $\phi$ uniform’ and ‘ML, $\phi$ uniform’, respectively; the pertinent Fisher proxy is shown as a solid curve (‘CRLB’). The ML estimator (‘ML, $\phi$ optimum’) and the Fisher proxy referred to the optimized rover orientation (‘Optimum CRLB’). For illustrative purposes, it is also shown the case of $\phi_s = \hat{\theta}$. 
2.2.4 **Optimum rover orientation**

In the above it is assumed that the rover travels across the network with orientations \( \{ \phi_s \}_{s=1}^{\infty} \) modeled as independent realizations of a \((0, 2\pi)\)-uniformly distributed random variable. We now investigate the performance improvement that can be achieved by optimizing the rover orientation \( \phi_s \) at each snapshot \( s \).

The optimization criterion is based on maximizing the Fisher information \( J_s(\theta, \phi_s) \), see Appendix 2.6.1, with respect to \( \phi_s \). We thus enforce the condition \( \partial J_s(\theta, \phi_s)/\partial \phi_s = 0 \) and select the points of maximum by evaluating the second derivative, thus getting \( \phi_s^{opt} = \theta \pm \arctan(\beta/\alpha) \).

Clearly, in order to select the optimal orientation \( \phi_s \) one should know \( \theta \). In practice we can use its current estimate \( \hat{\theta}_s \) instead. Thus, we can choose the following orientation strategy:

\[
\hat{\phi}_s^{opt} = \hat{\theta}_{s-1} \pm \arctan \left( \frac{\beta}{\alpha} \right), \quad s > 1
\]  

(2.2.7)

for snapshot \( s > 1 \), where \( \hat{\theta}_{s-1} \) is the estimate of the DOA based on the prior observed data. As to the indeterminateness of the signs, for \( s = 1 \) the choice is arbitrary, and for \( s > 1 \) we alternatively switch between the two. Numerical investigation shows that this alternate jumping yields better performance with respect to the choice of a constant sign.
Replacing $\phi_s$ with $\phi_s^{opt}$, the Fisher information (2.2.6) modifies in\textsuperscript{6}

$$ I_{M}^{opt}(\theta) = \frac{M}{2}(\bar{N} - 1 + e^{-\bar{N}}) \left( \frac{\psi^4}{1 - \psi^2} \right), $$

(2.2.8)

which verifies $I_{M}^{opt}(\theta) > I_{M}(\theta), \forall \psi$.

Fig. 2.2 shows the MSE improvement with the optimized rover orientation, compared to choosing uniformly $\phi_s$. Also shown is the MSE pertaining to a rover always pointed in some (non-optimized) direction (e.g., $\phi_s = \hat{\theta}$, shown in the figure, or $\phi_s = \hat{\theta} + \pi/2$, not shown) and it is worth noting that the performance worsens, even with respect to the case of uniformly chosen orientations.

An intuitive explanation of this behavior is available. In fact, $J_s(\theta, \phi_s)$ has two \textit{minima} when the rover is pointing the source (either with the major or minor axis), while two points of maximum are symmetrically located with respect to $\theta$. See also Fig. 2.3: when the FOV is pointed directly \textit{towards} the true DOA the derivative of both the spread and the concentration of “beeps” is zero; this derivative represents the sensitivity of the estimate to the measurement, and the nadir of each when either the major or minor axis is toward the true DOA. The best orientation is between these, as represented notionally in Fig. 2.3.

\textsuperscript{6} Actually, in practice we shall use $\hat{\phi}_s^{opt}$ in place of $\phi_s^{opt}$, and the tightness of the bound (2.2.8) will be checked numerically.
Fig. 2.3: Notional figure to explain the reason that a MA FOV that is obliquely-angled with respect to true (or best current estimate of true) DOA is best. Oscillating lines represent the relative spread of beep times within a poll and the concentration, for a FOV with a typical number of sensors enclosed. FOV’s are as given in lower part of plot. Note that the derivative of both curves is zero for when FOV is pointed toward true DOA.
2.2.5 Impact of timing errors

The measurement errors represent a relevant practical issue in many DOA estimation contexts, so that it is of interest to extend the previous results to a scenario in which one accounts for possible errors in the measurements of the hitting times $\tau_i$. Here we limit the analysis to the case in which the time differences $t_i$ are corrupted by Gaussian noise $\epsilon_i$ with zero mean and variance $\sigma^2_\epsilon$, under the assumption that $\epsilon_i$ are independent and identically distributed random variables.

Also, for simplicity, let us address basic case without optimized rover path.

By paralleling the earlier derivation, whose details are omitted for the sake of brevity, the final results are that the pdf of the noisy time differences is as in eq. (2.2.4) with the covariance matrix $C = \sigma^2_x Z$ simply replaced by $(\sigma^2_x + \sigma^2_\epsilon)Z$, and that the pertinent Fisher information becomes

$$I_{err}^M(\theta) = M(\bar{N} - 1 + e^{-\bar{N}}) \left[ \frac{(\gamma^2_x + \gamma^2_y + 2)}{\sqrt{(\gamma^2_x + 1)(\gamma^2_y + 1)}} - 2 \right],$$

(2.2.9)

with $\gamma^2_x = \alpha^2/(\sigma^2_\epsilon v^2)$ and $\gamma^2_y = \beta^2/(\sigma^2_\epsilon v^2)$. As a sanity check, note that in the limit of $\sigma^2_\epsilon \to 0$, $I_{err}^M(\theta)$ reduces exactly to that in eq. (2.2.6), whereas $\sigma^2_\epsilon \to \infty$ implies a vanishingly small $I_{err}^M$. 
2.3 Multiple sources

2.3.1 Decoupling approach

A first method to account for the presence of multiple acoustic sources is to decouple the multiple DOA estimation problem in its constituent, single-DOA, parts. In other words, we can formulate a different estimation problem for each DOA, thus not exploiting the inherent correlations imposed by the geometry of the system, see Fig. 2.1. This approach has the advantage of being a straightforward extension of the single-DOA estimation problem and is quite easy to implement. At the same time, it is largely sub-optimal. Later, we approach the problem in different way.

The decoupling approach relies upon a minor modification of the communication scheme between sensors and MA proposed described earlier. At the time instant, say $\tau_{ij}$, that the $i^{th}$ sensor is impinged by the wavefront of the $j^{th}$ source, an internal clock of period $jT$ is activated and the emission (only virtual, we recall) of the periodic signal $z_{ij}(t) = \sum_k d(t - \tau_{ij} - k j T)$ starts. The aggregate $z_i(t) = \sum_j z_{ij}(t)$ represents the total signal virtually emitted by the $i^{th}$ sensor. Let us denote by $\mathcal{I}_s$ the ensemble of indices of the sensors seen by the MA at snapshot $s$. Upon receiving the periodic waveform $\sum_{i \in \mathcal{I}_s} z_i(t)$, a little thought reveals that the MA is able to separate the contributions of the different impinging wavefronts.

For example, with reference to Fig. 2.1 where $|\mathcal{I}_s| = 3$ and 2 DOAs $\theta_1$ and $\theta_2$ are
considered, the MA is able to separate the contributions of the first impinging wavefront \( \sum_{i \in I_s} z_{i1}(t) \) from those of the second \( \sum_{i \in I_s} z_{i2}(t) \), provided that the observation interval is not smaller than the period \( 2T \) of the aggregate signal. In fact, in this way, the MA may in principle reconstruct the entire (infinite) periodic waveform and, from that, it is simple to discriminate those pulses that repeat with period \( T \) from those that repeat with period \( 2T \). Contributions arising from different DOAs are thus exactly separable and the complicated multiple-DOA estimation task may be approached as that of many different single-DOA estimation problems. Clearly, the approach illustrated for 2 DOAs can be straightforwardly generalized to an arbitrary number of sources.

Note that we have implicitly assumed that the successive wavefronts impinge on the network separately, in the sense that the \( j^{th} \) wavefront has already hit all the nodes of the network when the \( (j + 1)^{th} \) wavefront arrives. This seems a rather mild working hypothesis, and will be maintained in the following.

### 2.3.2 Joint Estimation of Multiple DOAs

We now consider the joint estimation of multiple DOAs and, again, for simplicity, we only refer to the case of 2 DOAs. The computation of the joint pdf \( f(\tau_{i1}, \tau_{i2}) \) of the time instants \( \tau_{i1} \) and \( \tau_{i2} \) is outlined in Appendix 2.5.2 and results in the
following Gaussian density

\[
f(\tau_1, \tau_2) = \exp\left\{-\frac{\tau_1^2}{\sigma_{s1}^2} - \frac{\tau_1\tau_2}{\sigma_{s1}\sigma_{s2}} + \frac{\tau_2^2}{\sigma_{s2}^2} \right\} \quad \frac{2\pi \sigma_{s1}\sigma_{s2}(1 - \rho_s^2)}{2(1 - \rho_s^2)}
\]

(2.3.1)

where

\[
\sigma_{s1}^2 = a^2\alpha^2 + b^2\beta^2, \quad \sigma_{s2}^2 = c^2\alpha^2 + d^2\beta^2,
\]

\[
a = \sin(\theta_1 - \phi_s)/v, \quad b = \cos(\theta_1 - \phi_s)/v,
\]

\[
c = \sin(\theta_2 - \phi_s)/v, \quad d = \cos(\theta_2 - \phi_s)/v,
\]

\[
\rho_s = \frac{ac\alpha^2 + bd\beta^2}{\sigma_{s1}\sigma_{s2}}.
\]

The key point is that the joint estimation of the two DOAs exploits the dependence between \(\tau_{i1}\) and \(\tau_{i2}\) (where the subscripts 1 and 2 denote \(\theta_1\) and \(\theta_2\) respectively) that we have deliberately ignored previously. Recall that \((\tau_{s1}, \tau_{s2})\) is made up of independent entries \(\tau_{i1}\) and \(\tau_{i2}\) so that the complete statistical characterization of the joint vector \((\tau_{s1}, \tau_{s2})\) is \(f(\tau_{s1}, \tau_{s2}) = \prod_{i=1}^{N_s} f(\tau_{i1}, \tau_{i2})\).

To actually exploit the dependence imposed by the sensors’ locations, however, the communication protocol between the MA and the sensors inside its FOV must be modified. It is necessary for the MA to recognize which signals come from which sensors: sensors must be identifiable. In practice this can be achieved by imposing some transmission diversity among different sensors. A typical example is a diversity in frequency. For instance, each sensor could be designed to transmit signals at a given frequency and the ensemble of possible frequencies is very large.
compared to the average number of sensors in each MA’s dwell (but much smaller than the overall number of sensors, of course, so that frequency reuse is allowed). When deployed over the surveyed area, sensors are chosen at random so that, with large probability, the sensors seen by the MA in each snapshot will transmit at different frequencies. The MA, equipped with a multi-frequency receiver, can easily discern the transmissions from the different sensors.

The derivation of the optimal (ML) estimator and of the pertinent Fisher information is now in order. Recall that, due to the time-origin uncertainty (see discussion in Sect. 2.2.1), at each snapshot the MA can only observe $2(N_s - 1)$ relative times. Taking for instance as reference $\tau_{11}$ and $\tau_{12}$, the observables become $(t_{s1}, t_{s2}) = (t_{21}, t_{31}, \ldots, t_{N_s, 1}, t_{22}, t_{32}, \ldots, t_{N_s, 2})$. In Appendix 2.5.2 the distribution of the joint vector $t_{s12} = (t_{s1}, t_{s2})$ is computed, resulting in a $2(N_s - 1)$-dimensional multivariate Gaussian density

$$f_{t_{s12}}(t_{s12}) = \exp \left(-\frac{1}{2} t_{s12}^T \Sigma^{-1} t_{s12}\right) \left(\frac{1}{2\pi}\right)^{2(N_s - 1)/2} \left(\det \Sigma\right)^{-1/2}, \quad (2.3.3)$$

having the $2(N_s - 1) \times 2(N_s - 1)$ covariance matrix of the form

$$\Sigma = \begin{pmatrix} C_1 & : & C_\rho \\ \vdots & \vdots & \vdots & \vdots \\ C_\rho & : & C_2 \end{pmatrix}. \quad (2.3.4)$$

In the above we have introduced the $(N_s - 1) \times (N_s - 1)$ matrices $C_1 = \sigma_{s1}^2 Z$, $C_2 = \sigma_{s2}^2 Z$, $C_\rho = \rho_s \sigma_{s1} \sigma_{s2} Z$. 
The matrix $\Sigma$ depends upon the parameters $\theta_1$ and $\theta_2$ of the different DOAs through $\sigma_{s1}$, $\sigma_{s2}$ and $\rho_s$, see eq. (2.3.2). The likelihood corresponding to $M$ independent snapshots results in the product of the individual likelihoods, and the ML estimation $\hat{\theta}_{ML} = (\hat{\theta}_{1ML}, \hat{\theta}_{2ML})$ of the 2 DOAs is

$$\hat{\theta}_{ML} = \arg \max_{\theta} \prod_{s=1}^{M} f(t_{s12}(t_{s12}). \quad (2.3.5)$$

In the case of vector estimates, the concept of Fisher information generalizes to the Fisher Information Matrix (FIM) [27]. As it is well known, the difference between the error covariance matrix of any unbiased vector estimator and the inverse FIM is nonegative definite. The FIM is computed in Appendix 2.6.2 and results in the symmetric form

$$I_M(\theta) = \frac{M}{4}(\bar{N} - 1 + e^{-N}) \begin{pmatrix} \Lambda(\theta, \psi) & \Omega(\theta, \psi) \\ \Omega(\theta, \psi) & \Lambda(\theta, \psi) \end{pmatrix}, \quad (2.3.6)$$

where $\theta = (\theta_1, \theta_2)$, and the expressions of $\Lambda(\theta, \psi)$ and $\Omega(\theta, \psi)$ are given in Appendix 2.6, see eqs. (2.6.1) and (2.6.2).

We know that the variances of unbiased estimators are lower bounded by the diagonal entries of the inverse FIM, and that the ML estimator attains these bound asymptotically with the number of independent snapshots $M$ [27]. Accordingly, the trace of the symmetric matrix $I_M(\theta)$ can be taken as a measure of the estimation accuracy for large values of $M$. We have

$$\text{Tr} (I_M(\theta)) = \frac{M}{2}(\bar{N} - 1 + e^{-N})\Lambda(\theta, \psi), \quad (2.3.7)$$
that is an increasing function of the eccentricity parameter $\psi$ and basically depends on the effective total number of timing informations collected by the MA, namely $M(\bar{N} - 1)$. These functional dependencies are similar to those of eq. (2.2.6).

As opposed to that case we now have that the true DOA values $\theta_1$ and $\theta_2$, play a role in determining the accuracy of the estimates as $\text{Tr}(I_M(\theta))$ depends upon $\theta$. Furthermore, due to the symmetric nature of the FIM in eq. (2.3.6), we note that, for a given $\theta$, the MSE lower bounds of the two estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ are the same. Thus, the performances of these two estimators approach the first diagonal entry of the inverse FIM. Using eq. (2.6.1) and (2.6.2) in eq. (2.3.6) yields

$$I_{M(1,1)}^{-1}(\theta) = \frac{1}{4M(\bar{N} - 1 + e^{-\bar{N}})} \left( -\frac{\alpha^4 - 6\alpha^2\beta^2 + \beta^4}{(\alpha^2 - \beta^2)^2} \right) + \frac{\alpha^4 + 14\alpha^2\beta^2 + \beta^4}{\alpha^4 + 10\alpha^2\beta^2 + \beta^4 + 4\alpha^2\beta^2 \cos(2(\theta_1 - \theta_2))},$$

that can be referred to as the CRLB of both the estimators.

When compared to the decoupled ML estimator, the performance improvements are remarkable. An example of such comparison is offered in Fig. 2.4 where almost one order of magnitude is gained in terms of MSE after a hundred of snapshots.

The reason is that measurements are coupled by the fixed (although random) positions of the sensors within the MA’s FOV, so that observations from one source carry information about the locations of the individual sensors, that can be exploited in estimating a second source.
Fig. 2.4: Uncoupled versus joint ML estimation procedures. Here we have \( v = 1 \) m/s, \( \alpha = 1 \) m, \( \beta = 0.3 \) m, \( \lambda = 3 \) m\(^{-2}\), \( \bar{N} \approx 6 \), \( \theta = (\theta_1, \theta_2) = (\pi/3, \pi/6) \).

The MSE of \( \hat{\theta}_1 \) is compared to its lower bound and to the performances of the uncoupled approach. Simulations are based on standard Monte Carlo counting process, involving 1000 runs for each point.
2.4 Conclusions

The paper deals with an architecture by which a network of unattended, randomly-placed and DOA-blind sensors can estimate a DOA from the passage of an acoustic waveform that contains a feature such as a sharp peak, onset or statistical change. Information from the sensors is collected by a roving data “mule,” mobile agent (MA). Previous treatments of the idea (see [7–10]) showed the viability of the concept for an idealized MA field of view (FOV) and single sources. In this paper we extend the scheme of [7–10] to a more practical FOV and to multiple sources. To summarize, there are three main results.

1. We have shown that estimation is possible with the non-ideal FOV: in fact, the expressions are, arguably, simpler. We have developed both fast/simple and optimal DOA-estimation schemes.

2. We have proposed a “sensor management” strategy for the MA to orient its FOV. It turns out that the best orientation is neither directly toward the true DOA (or the current best estimate of it) nor away, but squinted at an oblique angle.

3. We have proposed schemes for multiple-source DOA estimation. One scheme simply assumes separate estimation, and proposes a protocol to accomplish it. The second scheme requires that sensors be identifiable, and is hence more complex to implement; however, the estimation of the multiple DOAs is performed jointly.
It is interesting that the DOA estimation performance can be better for multiple than for single sources.

The intuition behind the latter issue is that the measurements are coupled by the positions of the sensors within the MA’s FOV. Different DOAs impinge on the same (although random) sensor configuration, implying that observations from one source reduce the uncertainty in the locations of the individual sensors, and this can help with estimation of another source’s DOA.

2.5 Appendix

2.5.1 Uncoupled ML estimation

Let \((x_i, y_i)\) be the coordinates of the \(i^{th}\) sensor in a reference system centered on the rover and with the abscissa taken along the major axis of its field of view. Let \(\xi_i\) be the abscissa of the \(i^{th}\) sensor, after an axis rotation such that the new abscissa of the system reference is aligned to the direction of arrival of the wavefront. We have \(\xi_i = x_i \sin(\theta - \phi_s) - y_i \cos(\theta - \phi_s)\).

Since \((x_i, y_i)\) are zero-mean independent Gaussians with variance \(\alpha^2\) and \(\beta^2\), and accounting for \(\tau_i = \xi_i/v\), we have

\[
    f_{\tau_i}(\tau) = \frac{1}{\sigma_s \sqrt{2\pi}} e^{-\frac{\tau^2}{2\sigma_s^2}}, \quad \text{with} \quad \sigma_s^2 = \frac{a^2 \alpha^2 + \beta^2}{(1 + a^2)v^2}. \tag{2.5.1}
\]

In order to get the ML estimator, we must compute \(f_{t_s}(t_s) = \int_{-\infty}^{+\infty} F(\tau_1) d\tau_1\).
where

\[ F(\tau_1) = \prod_{i=1}^{N_s} \frac{1}{\sigma_s \sqrt{2\pi}} \exp \left\{ -\frac{(t_i + \tau_1)^2}{2\sigma_s^2} \right\}. \]

Namely, we have

\[ f_{t_s}(t_s) = \left( \frac{1}{\sigma_s \sqrt{2\pi}} \right)^{N_s} \int_{-\infty}^{+\infty} \exp \left[ -\sum_{i=1}^{N_s} \frac{(t_i + \tau_1)^2}{2\sigma_s^2} \right] d\tau_1. \]

Expanding the square at the exponent and recalling that \( t_1 = 0 \), we get

\[ f_{t_s}(t_s) = \left( \frac{1}{\sigma_s \sqrt{2\pi}} \right)^{N_s} e^{-\frac{N_s t_2^2}{2\sigma_s^2}} \int_{-\infty}^{+\infty} e^{-\frac{N_s t_2^2}{2\sigma_s^2}} \frac{e^{-\tau_1 \sum_{i=2}^{N_s} t_i \sigma_s^2}}{\sigma_s^2} d\tau_1, \]

which finally yields

\[ f_{t_s}(t_s) = \frac{1}{\sqrt{N_s(2\pi\sigma_s^2)^{N_s-1}}} \times \exp \left\{ -\frac{1}{2\sigma_s^2 N_s} \left[ N_s \sum_{i=2}^{N_s} t_i^2 - \left( \sum_{i=2}^{N_s} t_i \right)^2 \right] \right\}. \tag{2.5.2} \]

In matrix form, we exactly get eq. (2.2.4).

### 2.5.2 Joint ML estimation

We only outline the computation, which is a simple generalization of the approach in Appendix 2.5.1. In the presence of two sources with DOAs \( \theta_1 \) and \( \theta_2 \), given that the coordinates \((x_i, y_i)\) of the \( i^{th} \) sensor are zero-mean independent Gaussians with variances \( \alpha^2 \) and \( \beta^2 \), and denoting with \( \xi_{i1} \) and \( \xi_{i2} \) the abscissas after rotating the system reference as done in Appendix 2.5, we have

\[ \xi_{i1} = x_i \sin(\theta_1 - \phi_s) - y_i \cos(\theta_1 - \phi_s), \]

\[ \xi_{i2} = x_i \sin(\theta_2 - \phi_s) - y_i \cos(\theta_2 - \phi_s). \]
Thus $\xi_1$ and $\xi_2$ are linear transformations of $x_i$ and $y_i$ and, accordingly, they are jointly Gaussian so that their joint pdf is easily worked out. Further accounting for $\tau_i = \xi_i/v$, the final result is that in eq. (2.3.1). Note that the generalization to more than two sources is straightforward.

Consider now the case of two sensors, $N_s = 2$, and recall that $t_{21} = \tau_{21} - \tau_{11}$ and $t_{22} = \tau_{22} - \tau_{12}$. Knowing the density of $\tau_{ij}$, where $i = 1, 2$, refers to the sensor and $j = 1, 2$, to the source, we can easily show that the vector $(t_{21}, t_{22})$ is a zero mean Gaussian vector with covariance matrix

$$
2 \begin{pmatrix}
\sigma^2_{s1} & \rho_s \sigma_{s1} \sigma_{s2} \\
\rho_s \sigma_{s1} \sigma_{s2} & \sigma^2_{s2}
\end{pmatrix}
= 2 \begin{pmatrix}
E[t^2_{21}] & E[t_{12}t_{22}] \\
E[t_{12}t_{22}] & E[t^2_{22}]
\end{pmatrix}.
$$

In fact we have, for instance,

$$
E[t_{12}t_{22}] = E[(\tau_{21} - \tau_{11})(\tau_{22} - \tau_{12})]
= E[\tau_{21}\tau_{22}] + E[\tau_{11}\tau_{12}] = 2\rho_s \sigma_{s1} \sigma_{s2}.
$$

Straightforward generalization to a generic number of sensors, gives the result in eqs. (2.3.3) and (2.3.4).

### 2.6 Appendix

#### 2.6.1 Single-DOA Fisher information

Let us assume that $N_s$ is given and let us append the symbol * to quantities (Fisher informations) conditioned on a specific value of $N_s$. We have from [28]
that the scalar Fisher information is given by $J^*_s(\theta, \phi_s) = \text{Tr}\{(C^{-1} \partial C/\partial \theta)^2\}/2$, where $C$ is as in eq. (2.2.4). Thus we get $J^*_s(\theta, \phi_s) = G^2(\theta, \phi_s)/(N_s - 1)/(2\sigma^4_s)$, with

$$G^2(\theta, \phi_s) = \left[ \frac{\partial \sigma^2_s}{\partial \theta} \right]^2 = \frac{4 \tan^2(\theta - \phi_s)(\alpha^2 - \beta^2)^2}{(1 + \tan^2(\theta - \phi_s))^2} \cdot \frac{1}{v^4}.$$ 

Removing the conditioning yields $J_s(\theta, \phi_s) = [\bar{N} - 1 + e^{-\bar{N}}] G^2(\theta, \phi_s)/(2\sigma^4_s)$, where $\bar{N} = 2\pi \lambda \alpha \beta$ is the expected value of the random variable $N_s$. Note that the ratio $G^2(\theta, \phi_s)/\sigma^4_s$ does not depend upon $v$, as we expect.

For large number of snapshots $M$, standard Monte Carlo integration approach allows us to replace the arithmetic mean with statistical expectation with respect to $\phi_s$, assumed uniform in $(0, 2\pi)$:

$$I_M(\theta) = \sum_{s=1}^{M} J_s(\theta, \phi_s) \\
\approx M \left[ \bar{N} - 1 + e^{-\bar{N}} \right] \frac{1}{2\pi} \int_0^{2\pi} \frac{G^2(\theta, \phi_s)}{2\sigma^4_s} d\phi_s \\
= M \left[ \bar{N} - 1 + e^{-\bar{N}} \right] \left( \frac{\alpha}{\beta} + \frac{\beta}{\alpha} - 2 \right),$$

and eq. (2.2.6) follows upon introducing the eccentricity $\psi = \sqrt{1 - \beta^2/\alpha^2}$.

### 2.6.2 Fisher information matrix

Let us assume that $N_s$ is given and let us append the symbol * to quantities (Fisher informations) conditioned on a specific value of $N_s$. We have from [28] that the entries $J^*_{s(i,j)}$ with $i, j = 1, 2$, of the Fisher information matrix are given
by
\[ J^*_s(i,j)(\theta) = \frac{1}{2} \text{Tr} \left\{ \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_j} \right\}. \]

Straightforward but tedious algebra then gives
\[ J^*_{s(1,1)}(\theta) = \frac{N_s - 1}{2} F(\theta, \phi_s), \quad J^*_{s(2,2)}(\theta) = \frac{N_s - 1}{2} H(\theta, \phi_s), \]
\[ J^*_{s(1,2)}(\theta) = J^*_{s(2,1)}(\theta) = \frac{N_s - 1}{2} L(\theta, \phi_s), \]
where the unpleasant explicit expressions of the functions $F$, $H$, and $L$ are omitted for simplicity. Removing the conditioning amounts to computing
\[ J_{s(i,j)}(\theta) = \sum_{k=2}^{+\infty} \text{Pr}(N_s = k) J^*_{s(i,j)}(\theta), \]
and the above average acts only on the factor $(N_s - 1)/2$, yielding $(\bar{N} - 1 + e^{-\bar{N}})/2$ (recall that $N_s$ is Poisson distributed with mean $\bar{N}$).

For a large number of snapshots $M$, a standard Monte Carlo integration approach allows us to replace the arithmetic mean with the statistical expectation, with respect to the uniform variable $\phi_s$. Namely, we have
\[ I_{M(1,1)}(\theta) = I_{M(2,2)}(\theta) = \sum_{s=1}^{M} J_{s(1,1)}(\theta) \approx M \left[ \bar{N}_s - 1 + e^{-\bar{N}_s} \right] \frac{1}{2\pi} \int_0^{2\pi} F(\theta, \phi_s) d\phi_s \]
\[ = \frac{M}{4} (\bar{N}_s - 1 + e^{-\bar{N}_s}) \Lambda(\theta, \psi), \]
and similarly

\[ I_{M(1,2)}(\theta) = I_{M(2,1)}(\theta) \]
\[ = \frac{M}{4}(\bar{N}_s - 1 + e^{-\bar{N}_s})\Omega(\theta, \psi), \]

where the explicit expressions of \( \Lambda \) and \( \Omega \) are

\[ \Lambda(\theta, \psi) = \left( \frac{1}{1 - \psi^2} + 1 - \psi^2 - 6 + \csc^2(\theta_1 - \theta_2) \right) \left( \frac{1}{2(1 - \psi^2)} + \frac{1}{2}(1 - \psi^2) + 7 \right), \quad (2.6.1) \]

and

\[ \Omega(\theta, \psi) = -\left( 6 \frac{1}{2} \left( \frac{1}{\sqrt{1 - \psi^2}} + \frac{1}{\sqrt{1 - \psi^2}} \right)^2 \right) \cos(2(\theta_1 - \theta_2))) \csc^2(\theta_1 - \theta_2).\quad (2.6.2) \]
3.1 Introduction

While multi-sensor systems hold great potential for surveillance performance, the technical challenges are significant, and include the need for effective calibration as well as a statistically-valid characterization of environmental uncertainties and contact measurement errors. Additionally, automatic tracking and fusion processing must contend with high false contact rates and target fading effects. Issues in multi-sensor surveillance and numerous design approaches are discussed in [29–31]. In [32, 33], we present model-based, simulation-based, and sea-trial tracking performance results with a track-oriented, modular multi-hypothesis tracking scheme. Of particular interest is the tradeoff between centralized and multi-stage processing: we have found that, when faced with significant target fading effects and for modest false contact rates, distributed processing can outperform centralized processing. This somewhat surprising result is based on the fundamental sub-optimality of all tracking algorithms that must contend with measurement origin
uncertainty. This explains the seeming contradiction with results in the nonlinear filtering and distributed detection literature, in particular the well-known optimality of centralized processing schemes. Ultimately, for sufficiently low-SNR target scenarios, effective real-time automatic tracking is extremely challenging regardless of the choice of data processing architecture. One approach is to relax the real-time requirement, and to leverage powerful batch processing techniques [34]. However, such schemes are not easily amenable to real-time surveillance requirements, and generally assume non-maneuvering targets. An alternative approach in challenging scenarios is to consider enlarging the surveillance network, possibly through bootstrapping approaches that include sub-band processing techniques [35] whereby a sensor is effectively “replaced” with a number of slightly-degraded sensors. The latter approach (enlarging the surveillance network) implicitly assumes that an increased number of like-performing, calibrated, registered sensors are always to be preferred, i.e. more sensors are always better than fewer. Is this true in general, or are there performance limits as the number of sensors becomes large? This is the issue that we address in this paper. We start by introducing in Section 3.2 a simple analytical model for tracker performance. We study tracker performance as a function of local detection threshold, number of sensors, and track management criteria. The model supports the conclusion that there are performance bounds on achievable performance in large sensor networks. Can
we do better if we consider a more complex, multi-stage processing architecture? In Section 3.3, we describe the fuse-before-track (FbT) architecture and provide motivation for its use in large sensor networks. The key insight that motivates the FbT architecture is that it couples the advantages of batch processing in the fusion step, followed by the advantages of scan-based processing (real-time processing with a maneuvering target model) in the tracking step. It is important to note that we do not argue that FbT will outperform a sufficiently complex centralized processing scheme. Rather, our claim is that multi-stage processing with a relatively simple tracking module can achieve good performance results. Further, as we will see in the Monte Carlo study, this performance is achieved with a significantly lower computational effort than in centralized (single-stage) processing. The first stage in the FbT architecture is a static fusion (or contact fusion) stage [29, 30]. Section 3.5 presents a Monte Carlo study of multi-sensor tracking performance for a representative multi-target surveillance scenario. The results suggest that the FbT architecture has merit and deserves further attention by the target tracking community. Section 3.6 provides conclusions and directions for future work.
3.2 Tracker performance modeling

Tracker performance modelling is addressed at length in [30,36]. Extensions that address target fading effects and distributed tracking architectures are in [32,33] and [37]. For our purposes here, we introduce a simple tracker performance model that identifies a compact relationship between scan rate and performance. Scan rate is directly proportional to the number of sensors, thus the model will support the subsequent analysis on performance as a function of the number of sensors.

3.2.1 Tracker Model

Modeling parameters:

- **Target**: kinematic “nearly constant position” motion model in two dimensions with maneuvering index \( q[m^2s^{-1}] \); fixed target SNR \( d \);

- **Sensor**: each with scan every \( \Delta t \) sec; positional measurements with covariance \( R \); surveillance region of size \( A[m^2] \), detection cells size \( C[m^2] \), detection threshold \( D \);

- **Tracker**: declare track on \( N \) consecutive associated detections, terminate track on \( K \) consecutive coasts (missed updates), association probability gate \( P_G \) and gating parameter \( \gamma \) (with two-dimensional measurements, \( \gamma = 9.2 \) corresponds to \( P_G = 0.99 \); see details in [36]).
The following derived quantities are of interest.

- **Detection probability** $P_D$, where we assume Rayleigh distributed amplitude statistics [29]:
  \[ P_D = \exp \left( -\frac{D}{1 + d} \right). \]  
  \[ (3.2.1) \]

- **False alarm density** [$m^{-2}$], where we again assume Rayleigh distributed amplitude statistics:
  \[ \lambda = \frac{\exp(-D)}{C}. \]  
  \[ (3.2.2) \]

- **False alarm rate** [$hr^{-1}$]:
  \[ \lambda_{FAR} = \frac{3600\lambda A}{\Delta t}. \]  
  \[ (3.2.3) \]

- **Probability of correct association**: we assume that the statistical nearest neighbor is used for track update, that the target-originated contact is one standard deviation in each dimension from the true target location, and that the track has steady-state filter covariance based on consecutive detection events. Thus, letting $S$ be the innovation covariance [30, p. 49] and letting
V be the validation region volume [30, p. 96], we have:

\[ P_{CA} = \exp(\lambda V), \]  
(3.2.4)

\[ V = \pi |S|^{1/2}, \]  
(3.2.5)

\[ S = P(-) + R, \]  
(3.2.6)

\[ P(-) = P(+) + Q, \]  
(3.2.7)

\[ P(+) = P(-) - P(-)(P(-) + R)^{-1}P(-), \]  
(3.2.8)

\[ Q = \begin{bmatrix} q\Delta t & 0 \\ 0 & q\Delta t \end{bmatrix}. \]  
(3.2.9)

Note that \( P(-) \) denotes the filter prediction covariance, while \( P(+) \) denotes the filter update covariance. Further, note that \( P(-) \) is the solution to the (steady-state) algebraic Riccati equation (ARE) [38].

- **Probability of track update and miss** (i.e. track coast), where a track update requires that the current scan include target detection, successful gating to the target track, and correct association:

\[ P_U = P_D P_G P_{CA}, \]  
(3.2.10)

\[ P_M = 1 - P_U. \]  
(3.2.11)

- **Average track confirmation time** (note that the expected value of the geometric distribution with parameter \( p \) is given by \( 1/p \)), where track confirmation
requires $N$ consecutive, associated target detections:

$$\tau_C = \frac{1}{P_{U}^{N-1}} \left( N - 1 + \frac{1}{P_D} \right) \Delta t. \quad (3.2.12)$$

- *Average track hold time*, where track termination is achieved after $K$ consecutive scans without an associated target detection:

$$\tau_H = \frac{1}{P_{M}^{K-1}} \left( K - 1 + \frac{1}{P_M} \right) \Delta t. \quad (3.2.13)$$

Note that equations (3.2.12) and (3.2.13) both rely on nested geometric probability distributions - that is to say, the sojourn time prior to tentative track initiation has a geometric distribution, as does the total track initiation time. It is easy to show (by linearity of the expectation operator) that equations (3.2.12) and (3.2.13) hold.

- *Track detection probability*, given by the fraction of time during which a target has a corresponding confirmed track:

$$P_D^T = \frac{\tau_H}{\tau_C + \tau_H}. \quad (3.2.14)$$

- *Probability of false update*, given by one minus the probability that no false contacts exist in the association gate and again assuming steady-state filter covariance:

$$P_{FU} = 1 - \exp(\lambda V_\gamma), \quad (3.2.15)$$

$$V_\gamma = \gamma \pi |S|^{1/2}. \quad (3.2.16)$$
• **Probability of false track**, given by the probability that a false contact leads to a sequence of associated false contacts:

\[
P_{FT} = P_{FU}^{N-1}.
\]  

(3.2.17)

• False track rate \([hr^{-1}]\):

\[
\lambda_{FTR} = \frac{3600 P_{FT} \lambda A}{\Delta t}.
\]  

(3.2.18)

For simplicity and to minimize the number of modeling parameters, we have assumed track confirmation on \(N\) consecutive detections rather than a more general \(M-of-N\) track initiation criterion. The equations above generalize easily to the \(M-of-N\) criterion. We have invoked several modeling simplifications, including the impact of false updates on the true track formation and maintenance. This effect is estimated empirically in [36, pp. 207-208], as the impact is difficult to capture analytically. Here, we assume for simplicity that the impact of a false update in terms of track degradation is comparable to that of a track miss. An illustration of the Markov chain model that corresponds to the modeling above is given in Figure 3.1. The tracker performance model introduced here shares some commonalities with the system operating characteristics (SOC) curve introduced in [39]. One of the differences is that the metrics of interest differ. In [39], a single, fixed time window in considered, and track detection and false track probabilities are computed. Rather, here the track detection probability is a measure of track
Fig. 3.1: Markov chain model for tracker logical state, in the target-present case.

(A similar Markov chain applies to the target-absent case.)

hold, which answers the following question: “For a given target, what is the probability that there is a corresponding track at any given time?” The false track rate identifies the number of false objects generated by the tracker per unit time.

3.2.2 Tracker Performance Analysis

We are interested to examine input and output performance curves (\( \lambda_{FAR} \) vs. \( P_D \), and \( \lambda_{FTR} \) vs. \( P_D^T \), respectively) as a function of the detection threshold \( D \), and as
Table 3.1: Model simulation parameter

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maneuverability index $q$</td>
<td>$100m^2s^{-1}$</td>
</tr>
<tr>
<td>Target amplitude $d$</td>
<td>$10dB$</td>
</tr>
<tr>
<td>Scan interval $\Delta t'$</td>
<td>60 sec</td>
</tr>
<tr>
<td>Measurement covariance matrix $R$</td>
<td>$\begin{bmatrix} 100 &amp; 0 \ 0 &amp; 100 \end{bmatrix}m^2.$</td>
</tr>
<tr>
<td>Surveillance region $A$</td>
<td>$10^8m^2$</td>
</tr>
<tr>
<td>Detection cell $C$</td>
<td>$100m^2$</td>
</tr>
<tr>
<td>Detection threshold $D$</td>
<td>$5.0 - 9.0dB$</td>
</tr>
<tr>
<td>Track initiation $N$</td>
<td>3</td>
</tr>
<tr>
<td>Track termination $K$</td>
<td>3</td>
</tr>
<tr>
<td>Association gate $\gamma$</td>
<td>9.2</td>
</tr>
<tr>
<td>Gate probability $P_G$</td>
<td>0.99</td>
</tr>
</tbody>
</table>
Fig. 3.2: Performance curves for several network-size assumptions. Solid lines characterize sensor performance, while dotted lines characterize tracker performance as characterized by the analytical model in Section 3.2.1. a function of the number of sensors. The latter can be addressed by setting the scan rate to $\Delta t = \Delta t' / Z$, where $Z$ is the number of equally-performing sensors and $\Delta t'$ is the single-sensor rate. Parameters are set as indicated in Table 3.1, and performance curves are in Figure 3.2. (Note that by object we mean either contact or track.)
Key conclusions are as follows:

- Tracking provides a roughly two-order-of-magnitude reduction in false objects, with comparable object detection performance;
- With a low constraint on false object rate, it is best to use few sensors;
- With a larger constraint on false object rate, it is best to use more sensors;
- For any given number of sensors, unlike the behavior of the (monotonic) input ROC curve, a maximum in track-level detection is achieved for a non-zero SNR detection threshold.

3.2.3 Optimal Detection Threshold and Number of Sensors

An approach to improve centralized tracking performance is to optimize the local sensor detection threshold \((D)\) as well as the number of sensors to be processed \((Z)\), as a function of a constraint on the false track rate. We anticipate that this will lead to a performance curve that is the envelope of the family of curves shown in Figure 3.2. For a given, we wish to optimize the local sensor detection threshold \((D)\) as well as the sensor scan rate \(\Delta t\) (from which we infer the number of sensors). This optimization problem can be recast as the following constrained
**Fig. 3.3:** Performance curve obtained solving the optimization problem (3.2.19).

maximization problem:

$$\max_{\Delta t, D} P_D^T(\Delta t, D),$$

s.t. \quad \lambda_{FTR}(\Delta t, D) = \alpha. \tag{3.2.19}$$

Note that the dependence of $P_D^T$ on $\Delta t$ is complex, since $P_U$ and $P_M$ both depend on $\Delta t$. This optimization problem does not lend itself to an analytical solution.

Using the same parameter settings as in Table 3.1, the solution to equation (3.2.19) leads to the envelope of the family of curves in Figure 3.2, as illustrated in Figure
3.3. (Optimization is performed using the function fmincon in MATLAB.)

3.2.4 Tuning the Track Management Parameters

A possible objection to the results in Sections 3.2.2 and 3.2.3 is that we use the same track initiation and termination criteria throughout. More generally, one might wish to adapt the parameters $N$ and $K$ to the data rate and to the detection threshold. A fully adaptive selection of these parameters is quite complex. We may, however, seek to vary $N$ and $K$ as a function of the data rate only. In particular, neglecting the dependence of $P_{FU}$ on the data rate, we can achieve a comparable false track rate by setting $N(Z)$ in the case of $Z > 1$ sensors as follows, where $N(1) = N_0$. We set the false track rate to be independent of the number of sensors:

$$\lambda_{FTR} = \frac{3600P_{FU}^{N(Z)-1}\lambda A}{\Delta t/Z} = \frac{3600P_{FU}^{N_0-1}\lambda A}{\Delta t}$$

(3.2.20)

Neglecting the dependence of $P_{FU}$ on the scan rate leads to the following simple relationship between the track confirmation window length and the number of sensors:

$$N(Z) = N_0 + \frac{\log Z}{\log (1/P_{FU})}.$$  

(3.2.21)

We scale the parameter $K$ in a comparable manner. As a result of adaptively-selected track-management parameters, the curves in Figure 3.2 are modified to
Fig. 3.4: Performance curves for several network-size assumptions, with adaptive track-management parameters. Solid lines characterize sensor performance, while dotted lines characterize tracker performance as characterized by the analytical model in Section 3.2.1.
those shown in Figure 3.4. The corresponding values of $N$ (and $K$) are given in Table 3.2.

We find that adaptive track management does have an impact on centralized tracking performance. In particular, note that the 100-sensor performance curve is better than the 10-sensor curve, which in turn outperforms the one-sensor curve. Nonetheless, the qualitative findings noted above still hold: significant reduction in false objects with automatic tracking, non-monotonicity in tracker performance curves, and saturation in performance benefits for a large-enough number of sensors.

3.3 The Fuse-Before-Track Architecture

Is it possible to exceed centralized tracking performance? At a conceptual level, the answer would seem to be no: every algorithmic step that is possible in multi-stage or distributed processing can be achieved in a centralized processing configuration. A practical question of interest is whether a particular (sub-optimal) tracking module is better employed in a single-stage processing architecture, or in a two-stage architecture. For the latter approach, we are interested to explore a fuse-before-track (FbT) processing scheme, whereby contact fusion across sensors precedes tracking over time. The two approaches are illustrated in Figure 3.5.

It is important to note that the FbT architecture supports real-time processing
Table 3.2: Adaptive track management for the results in Figure 3.2

<table>
<thead>
<tr>
<th>Number of sensors (Z)</th>
<th>Track initiation (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
</tr>
<tr>
<td>1,000</td>
<td>5</td>
</tr>
<tr>
<td>10,000</td>
<td>6</td>
</tr>
</tbody>
</table>

Fig. 3.5: Candidate fusion and tracking architectures.
just as the centralized architecture does: as fused contacts are produced, they provide input to the second-stage scan-based tracker. For the purposes of this discussion, the particular choice of tracking module (here, a track-oriented multi-hypothesis tracker [32, 33]) is not critical; however, what is critical is that our tracker is representative of a real-time, scan-based algorithm that necessarily discards data-association hypotheses (in the case of non-Bayesian tracking) or that combines hypotheses (in the case of Bayesian tracking). Indeed, while amenable to real-time surveillance requirements, scan-based tracking approaches lack the performance potential of batch processing schemes such as [34]. For simplicity, we assume that the (active) sensors are (nearly) synchronized; that is, we assume that the scans of contact-level data are acquired for the same sequence of times, for all sensors. An alternative, time-series representation of the two architectures in Figure 3.5 is illustrated in Figure 3.6.

The motivation for investigating the FbT architecture is as follows. The static fusion stage is not hindered by the requirement for scan-based processing, since all the sensors scan the surveillance region simultaneously. Thus, for large sensor networks, the two-stage architecture leverages the strength of batch processing in the fusion stage, while maintaining the real-time surveillance requirement with scan-based tracking. Let us return for a moment to the argument that the same processing results obtained with FbT are in principle achievable with single-stage
Fig. 3.6: FbT includes static fusion and scan-based tracking.
or centralized processing. After all, as we saw in the model-based results documented in Section 3.2, there is an advantage to scaling appropriately the track-management parameters with the sensor data rate. However, key data association parameters do not scale well with increasing data rate. For instance, for computational reasons, the \textit{n-scan} track hypothesis depth parameter that is common in multi-hypothesis tracking (MHT) approaches cannot be scaled with the data rate. A similar consideration holds for multiple-model filters. Thus, as we will see in Section 3.5, an adjustment to track-management parameters for centralized processing is insufficient to match the promising performance results exhibited by the FbT architecture.

3.4 Static Fusion

The static fusion problem represents the first stage in the fuse-before-track (FbT) architecture. We adopt the (simple) contact sifting approach and its performance characteristics [40]; subsequently, we will briefly discuss an alternative approach to static fusion. The contact sifting approach relies on a sifting grid in measurement space; in each sifting cell, we sum the number of contacts over all the sensors in the network; sifting cells in which the number of contacts exceeds a preassigned threshold lead to a fused contact, with localization based on appropriate averaging over the location of the single-sensor contacts (accounting for contact uncertain-
ties). Details of this approach can be found in [40]. The contact sifting algorithm is not effective in close-target scenarios. Another approach to the problem is to consider a \textit{multi-sensor probabilistic data association} algorithm. Further details of this approach may be found in [41]. Under this approach, we take a \textit{generalized likelihood ratio test} (GLRT) approach for both the detection and estimation problems. For each hypothesized target, we find its location estimate which maximizes the likelihood function, and choose the hypothesis which has the largest likelihood. This results in a procedure which maximizes the likelihood function with respect to the number of targets and their respective locations. A constraint is imposed on the maximum allowable number of targets present in the surveillance region. A sequential search over the number of targets is used for a computationally feasible solution. The technique provides location estimates as part of the detection process. The location estimates can always be further refined by an estimation process. This approach in a different context is developed in [42, 43]. In [43], comparisons are made between the proposed method and the unstructured and structured techniques based on Akaike information theoretic criteria (AIC) [44], minimum description length (MDL) [45], and Bayesian predictive density [46]. Compared to these approaches to contact fusion, the contact-sifting algorithm is simple and handles situations where we have large number of targets, albeit not closely spaced.
3.5 Fuse-Before-Track Performance Study

We have argued in Section 3.3 that the fuse-before-track architecture holds potential for target tracking in large sensor networks. Here, we provide results of Monte Carlo experimental validation. For both the centralized and FbT architectures, we use a track-oriented multi-hypothesis tracker [32, 33]. Simulation and algorithmic parameters are in Table 3.3. We have stochastically-generated target ground truth based on a nearly constant velocity motion model, for which positional measurements are obtained from a number of like-performing sensors that are synchronized in time and with a fixed sensor revisit time. All target trajectories are initiated at scenario initiation, and target death results if a target exits the scenario area. Initial target location is uniformly distributed in the surveillance region. The tracker is assumed to have knowledge of target motion and sensor parameters; by this we mean that the statistical characteristics of target motion and of the sensor measurement error are known. Performance evaluation relies on track classification whereby those tracks with sufficiently large average localization error from all target trajectories are classified as false; otherwise, the closest trajectory is identified. For tracks that extend in time beyond a given target death, the last target location is used for positional comparison. Note that the track-initiation setting for the multi-sensor (centralized) configuration is more stringent, so as to have a comparable track rate at the processing output. The
Table 3.3: Parameters for single-sensor tracking, multi-sensor tracking, and FbT simulation-based performance evaluation.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monte Carlo realizations</td>
<td>500</td>
</tr>
<tr>
<td>Number of targets</td>
<td>10</td>
</tr>
<tr>
<td>Target SNR</td>
<td>13 dB</td>
</tr>
<tr>
<td>Target maneuverability index</td>
<td>0.01m²s⁻¹</td>
</tr>
<tr>
<td>Initial velocity std. dev.</td>
<td>1ms⁻¹</td>
</tr>
<tr>
<td>Sensor threshold</td>
<td>10dB</td>
</tr>
<tr>
<td>Contact measurement error std. dev.</td>
<td>10m</td>
</tr>
<tr>
<td>Number of sensors</td>
<td>10</td>
</tr>
<tr>
<td>Sensor revisit time</td>
<td>3min</td>
</tr>
<tr>
<td>Surveillance region</td>
<td>(1.5km)²</td>
</tr>
<tr>
<td>Detection cell size</td>
<td>(1m)²</td>
</tr>
<tr>
<td>Sifting cell size</td>
<td>(30m)²</td>
</tr>
<tr>
<td>Sifting threshold</td>
<td>3</td>
</tr>
<tr>
<td>Track initiation (FbT)</td>
<td>4 – of 4</td>
</tr>
<tr>
<td>Track initiation (multi-sensor)</td>
<td>5 – of – 5</td>
</tr>
<tr>
<td>Allowed misses before track termination</td>
<td>3</td>
</tr>
<tr>
<td>Hypothesis tree depth (n-scan)</td>
<td>2</td>
</tr>
<tr>
<td>Track classification distance threshold</td>
<td>14.14m</td>
</tr>
</tbody>
</table>
sensor threshold and target SNR settings above, lead to a target probability of detection ($P_D$) of 0.62. The sensor threshold, detection cell size, and surveillance region sizes lead to a contact false alarm rate ($\lambda_{FAR}$) of 102 contacts per scan. Given the scenario revisit time and scenario duration, for each Monte Carlo realization there are 18 contact files for each sensor, and 180 in total, leading to 18 fused-contact files. Given the sifting cell size and sifting thresholds above, the first stage of FbT processing leads to approximately 40-50 fused contacts per scan. It is worth noting that fused-contact location is given by the mean of the contacts in the sifting cell, and the fused-contact measurement covariance is smaller than that of single-sensor contacts by a factor of 10 (the number of sensors). These statistics follow directly from least squares estimation, and may alternatively be interpreted as the result of Kalman filtering of a sequence of contacts at the same measurement time. Performance results are given in Table 3.4. Note first that the overall false contact rate is of almost 37K contacts per hour, thus target tracking provides a significant data reduction. In the case of FbT processing, the track reduction comes with a track-hold (track detection probability $P_{TD}$) that exceeds the contact-level detection probability. In all cases, track localization accuracy is much lower than contact-level accuracy. As expected, multi-sensor tracking performance is superior to single-sensor performance.

The primary result of interest is the performance increase in FbT processing com-
pared with single-stage multi-sensor tracking. Note also the low execution time [sec] in FbT processing, due to rapid MHT processing with fused-contact files that have less than half the contacts as single-sensor files. Figure 3.7 provides an illustration of a realization of contact-level and track-level data. We next focus on a couple of examples of track formation that illustrate the improved performance that we find in FbT processing.

The first example (Figure 3.8) illustrates the challenge in multi-sensor tracking with a high rate of contact files: this leads to improved track hold over single-sensor tracking, but often with short-duration tracks with a corresponding difficulty to achieve nonlinear filtering convergence. On the other hand, the FbT architecture leads to a contact-file rate that is the same as in single-sensor tracking, and with less than half the number of false contacts. A second example (Figure 3.9) provides a further illustration of improved track hold in FbT processing. Here the single-sensor tracker does not even detect the target, and the multi-sensor tracker is severely challenged. Finally, it is of interest to assess whether an increase in the hypothesis tree depth (or n-scan) in scan-based multi-hypothesis tracking can improve the competitiveness of the centralized multi-sensor tracking architecture with respect to the FbT approach. Table 3.5 summarizes the results of a further simulation-based performance assessment. The scenario description is the same as described in Table 3.3, with the results are based on 50 Monte Carlo scenario
### Table 3.4: Tracker performance results

<table>
<thead>
<tr>
<th>Metric</th>
<th>Single-Sensor</th>
<th>Multi-Sensor</th>
<th>FbT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_D^T$</td>
<td>0.092</td>
<td>0.473</td>
<td>0.659</td>
</tr>
<tr>
<td>$\lambda_{FTR}$</td>
<td>288.7</td>
<td>171.2</td>
<td>65.0</td>
</tr>
<tr>
<td>Exec. time</td>
<td>21.9</td>
<td>171.6</td>
<td>8.5</td>
</tr>
</tbody>
</table>

### Table 3.5: Tracker performance results with two centralized settings.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Single-Sensor</th>
<th>10 sensors n-scan=2</th>
<th>10 sensors n-scan=4</th>
<th>FbT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_D^T$</td>
<td>0.0786</td>
<td>0.449</td>
<td>0.392</td>
<td>0.643</td>
</tr>
<tr>
<td>$\lambda_{FTR}$</td>
<td>308.80</td>
<td>175.2</td>
<td>155.69</td>
<td>56.0</td>
</tr>
<tr>
<td>Exec. time</td>
<td>22.53</td>
<td>174.3</td>
<td>294.6</td>
<td>7.74</td>
</tr>
</tbody>
</table>
**Fig. 3.7:** Illustration of one realization in the simulation-based analysis of single-sensor, (single-stage) multi-sensor and fuse-before-track processing. False contacts are black dots, target-originated contacts are magenta dots, target trajectories are magenta, single-sensor tracks are red, multi-sensor tracks are blue, and FbT tracks are cyan; the (intermediate stage) contact fusion output in FbT processing are cyan dots.
Fig. 3.8: Close-up view on one target in Figure 3.7. False contacts are black dots, target-originated contacts are magenta dots, target trajectories are magenta, single-sensor tracks are red, multi-sensor tracks are blue, and FbT tracks are cyan; the (intermediate stage) contact fusion output in FbT processing are cyan dots.
realizations. For the centralized architecture, note that we have considered both the “standard setting” \( n\text{-scan} = 2 \), as well as an “extended-hypothesis setting” \( n\text{-scan} = 4 \); the latter leads to slower than real-time execution by a factor of approximately 1.64 (given by 294.6/180). An illustration is in Figure 3.10.

Note that the performance numbers are comparable to the previous simulation-based study, with FbT performance exceeding multi-sensor centralized tracking, and the latter exceeding single-sensor tracking. The numbers of particular interest are highlighted in boldface in Table 3.5, and concern the impact of hypothesis tree depth of tracking performance. Indeed, note that, somewhat disappointingly, an increased \( n\text{-scan} \) does not lead to a measurable increase in tracking performance. Centralized tracking with \textit{large} hypothesis tree depth comparable to the effective time depth in the FbT architecture (i.e. \( n\text{-scan} \) in the range of 20-30), should in principle perform as well as the FbT approach. However, the MHT module is inherently a sub-optimal one. Thus, it cannot be guaranteed that similar quantitative results are achieved. Further, due to computational limitations, a direct comparison is prohibitive. We have chosen to take a first step towards an architectural comparison of centralized and FbT tracking with comparable time-depth processing by doubling the \( n\text{-scan} \) parameter in the centralized tracking instantiation of the MHT module. The results indicate that we cannot guarantee a monotonic improvement in performance with increasing \( n\text{-scan} \). Fur-
Fig. 3.9: Another example, with same color-coding as above. Note that no single-sensor track is formed, and the multi-sensor track has significantly less track hold ($P^T_D$) than the FbT result.
**Fig. 3.10:** Close-up view on one target: a case where larger *n-scan* helps slightly.

False contacts are black dots, target-originated contacts are magenta dots, target trajectories are magenta, single-sensor tracks are red, multi-sensor tracks are blue (dotted in case of larger n-scan), and FbT tracks are cyan; the (intermediate stage) contact fusion output in FbT processing are cyan dots.
ther, it appears that a large $n$-scan as noted above would fall short of matching FbT performance. Our MHT hypothesis scoring and pruning logic is predicated on an attempt to partition contact-level data into a minimal number of contact sequences that obey the target kinematic modelling assumptions (i.e. nearly constant velocity). Thus, not surprisingly, an increase in $n$-scan does lead to more effective data partitioning, in the sense that the MHT identifies a smaller number of objects or sets of contacts (i.e. smaller $\lambda_{FTT}$), each of which identifies a confirmed track. Nonetheless, more effective data partitioning as defined here may lead to tracks that, for instance, remain alive longer but deviate from the true target trajectory, thus leading to a track that exceeds the localization threshold to be classified as target-originated, in turn lowering the $P^T_D$ metric. Thus, while in the larger $n$-scan case we expect (and achieve) a decreased $\lambda_{FTT}$, the impact on $P^T_D$ is less predictable. Figure 3.10 illustrates the tracks for one target realization, where we see some performance benefit to large $n$-scan centralized tracking, though the result is still far short of FbT performance. It should be noted that this study has been limited to random target tracks in a fairly wide surveillance region, which seldom leads to dense multi-target instantiations. These would challenge the FbT architecture as implemented here, as the first-stage contact-sifting algorithm is known to be inappropriate for close-target cases. Likewise, dense-target scenarios would challenge the centralized tracker more severely, and would likely lead to a
more appreciable benefit to centralized tracking with a larger \textit{n-scan}.

### 3.6 Summary

While large sensor networks hold great potential for surveillance performance, current scan-based target tracking technology by itself may not offer the best processing paradigm. Conversely, existing batch processing approaches do not provide real-time surveillance outputs. Thus, we believe a two-stage architecture that leverages the strengths of both batch and scan-based processing holds great potential for effective surveillance performance. In particular, contact fusion for a large number of nearly simultaneous sensor scans may be followed quite effectively by scan-based tracking. This paper has addressed these contributions. First, in Section 3.2 we introduced an analytical performance model for scan-based tracking, and studied the performance limitations that the model suggests for increasing date rates (or number of sensors). Next, in Section 3.3 we introduced the fuse-before-track (FbT) architecture for automatic tracking in large sensor networks, which includes contact fusing followed by scan-based tracking; the specific FbT instantiation that we have implemented utilizes the contact sifting algorithm followed by an MHT tracker. We proceeded in Section 3.5 to describe simulation results that compare centralized and FbT processing results. We found that the FbT approach outperforms centralized (single-stage) tracking, provided tar-
gets are not closely spaced. An effective FbT approach for closely spaced targets will require a more effective contact-fusion step to precede scan-based tracking. A number of directions for future work exist. Principally, and in addition to a more effective first stage in the FbT architecture to handle general multi-target scenarios, the future direction includes an analysis of the impact of synchronized vs. staggered sensor sampling times [47,48], for which analysis in the large-sensor case is lacking.
Chapter 4

Distributed Target Detection in Sensor Networks Using

Scan Statistics

4.1 Introduction

Distributed detection using multiple sensors and optimal fusion rules have been extensively investigated (e.g., [15] [11] [12]). The case where the communication links are not ideal but instead are subject to error has been addressed in [49], and that with dependent sensors is studied in [50]. For the fusion scheme, Chair and Varshney [51] showed that the optimal fusion rule is the likelihood ratio test on the decisions from sensors and becomes a threshold detector on the weighted sum of binary sensor decisions and the weight is obtained using the local detection and false alarm probability at each sensor under each hypothesis.

In [52] [53] the authors propose an appealing architecture for sensor networks (SN) in which information is accumulated in a traveling Mobile Agent (MA) (think of a wheeled rover collecting data from a haphazardly deployed dynamic population
of functional ground sensors, or of a surface vessel collecting data from a drifting underwater sensor network), which sequentially queries the sensors that fall in its current (and changing) field of view (FOV). The MA, playing the role of the fusion center, makes the final decision about the presence of a target. There are sequential versions of this (e.g., [54]) that seek to use an MA to decide as quickly as possible which among a pair of hypotheses is true.

Most distributed detection references begin with the assumption that the local sensors’ detection performances – their probabilities of detection and false alarm rate – are known to the fusion center. For a dynamic target and passive sensors, it is difficult to estimate local sensors’ performances via experiment: these performances are time varying as the target moves through the sensor field, and, in any case, it is expensive to transmit them to the fusion center. In [55] a fusion rule that uses the total number of detections (“1”s) transmitted from the local sensors as the test statistic was proposed for the case where the total number of sensors in the SN was known. In [56] the authors studied the case of random deployment of sensors within the SN: even the number of sensors was unknown and was modeled as a random variable, and via this prior information the system level decision at the fusion center compared number of detections made by the local sensors to a threshold.

Most distributed detection references also begin with the assumption that the
hypotheses are homogeneous: either $\mathcal{H}_0$ (target absent) is true at all sensors or $\mathcal{H}_1$ (target present) is uniformly true. For example, while the counting rule is a very reasonable approach, it neither localizes the target nor makes any attempt to use information about the contiguity of sensors that report a target. We believe we can do better, since, more realistically, a target is a local disturbance.

In this chapter of the dissertation we combine the three previous ideas: we introduce a two-dimensional MA-based sequential procedure for detecting and localizing a target, counting only binary detection-level sensor data, and testing for any spatial inhomogeneity of such reports. The tool, of emerging importance in fields such as internet safety (detections denote network intrusion attempts, plotted versus time) or epidemiology (detections represent reported illness, plotted versus location), is the scan statistic [57], [58], [59]. Scan statistics look for events that are clustered amongst a background of those that are sporadic.

In a nutshell, a scan-statistic test slides a window across its observation domain, sums the number of events (detections) within this window, and continually tests the count that it finds against a threshold (see Figure 4.1). Now, assuming that the window matches the shape of the “cluster” that is sought, the scan statistic is clearly a generalized likelihood ratio test (GLRT) [60]; so what is new? The answer is that quite precise expressions have been developed for the exceedance probability for a scan statistic, and it is exactly the lack of ability to set a re-
liable threshold that often makes a GLRT unappetizing. One could, of course, circumvent this by testing the maximum amongst non-overlapping windows (as opposed to a sliding one), since in that case their counts would be independent, and the maximal order-statistic trivial to calculate. But there would be a high price to be paid in detectability if the cluster did not precisely match one of these non-intersecting scans. The scan statistic is appealing both ways: it has excellent detection performance without regard to cluster location, and by accounting for the count dependence between overlapping scans it has reliable probability of false alarm.

In addition, we will show that an optimal window size for the MA’s field of view exists, and we compute it. In light of the above, it also seems reasonable to compare the two-dimensional scan statistic test to:

- a two-dimensional non-overlapping test, and
- a counting rule test [56],

both of which use the total number of detections reported by local sensors.

The informing situation is in Figure 4.2. We test for a “target,” whose presence is suggested by randomly-deployed sensors; for example, these elements may sense a magnetic disturbance, acoustic signature, or even bistatic reflections from a remote radio source. Each sensor can generate a spurious false alarm with probability $P_{fa}$. More interesting, we assume that the sensor’s probability of detection
decreases with increasing distance between it and the target, due to an inverse-
square power law. A MA “drags” a sliding window around the area, and if the
number of reporting sensors that are encompassed exceeds a threshold level that
is determined by scan-statistic theory, a detection is declared.

The remainder of this chapter is organized as follows. A general overview of
scan statistics together with a detailed mathematical formulation of the two-
dimensional scan statistic is given in Section 4.2. Section 4.3 discusses the target
and signal models adopted and analytical/simulation results for the Bernoulli case
where the number of sensors in the SN is known. The optimal field of view of the
MA scenario is dealt with in Section 4.3.4, while Section 4.4 is devoted to consider
the case of Poisson sensor field where the number of sensors in the SN is unknown.
Sect. 4.4.2 provides a comparison with the non-overlapping test. Finally, Section
4.4.3 provides comparative performance of the fixed counting rule test [55] and
the proposed test. Concluding remarks are presented in Section 4.5.

4.2 Two-Dimensional Scan Statistics

4.2.1 Definitions

The detection and analysis of clustering of events is of great importance in many
areas of science and technology including: epidemiology [61], bionformatics [62],
biosurveillance [63], ecology [64], medicine [65], quality control and reliability [66].
Fig. 4.1: Scanning the unit time interval with a window of length $w = 0.2$. The * represent times of occurrence of $N = 19$ events, $S_{0.2} = 8$. The bar represents the window that contains the maximal number of events; that is, the scan statistic.

Theory and applications of scan statistics, as well as recent advances, have been presented in [57], [58] and [67].

The following example serves as a good introduction to scan statistics. Let us consider $N$ independent observations from a uniform distribution on the interval $(0, T)$ (see Figure 4.1). Let $S_w$ be the largest number of observations in any window of length $w$ in the interval $(0, T)$. $S_w$ is called a scan statistic and it has been used in testing the null hypothesis – that the $N$ points in $(0, T)$ are uniformly distributed – against a clustering alternative [68]. Scan statistics have been also formulated for a sequence of independent and identically distributed
(i.i.d.) discrete random variables $X_1, \ldots, X_T$. A *discrete scan statistic* has been defined as follows:

$$S_m = \max \{1 \leq i \leq T - m + 1; X_i + \ldots + X_{i+m-1}\}, \quad (4.2.1)$$

where $m$ is the length of the scanning window. For the special case of iid $0 - 1$ Bernoulli observations, the event $S_m = m$ implies that the longest run of $1$'s is of length $m$. One can view the discrete scan statistic $S_m$ as an extension of the longest run statistic.

Since in this chapter we employ a scan statistic for observed counts of signals in a two dimensional region, we introduce below a two dimensional scan statistic. This scan statistic deals with the following situation. A region $R$ of Euclidean space is tessellated or subdivided into cells (which will be denoted by the symbol $a$, and are presumably much smaller than the window to be used when scanning). Data are available in the form of nonnegative integral counts $Y_a$ on cells $a$. In addition, a “size” value $A_a$ is associated with each cell $a$. The cell sizes $A_a$ are known and fixed, while the cell counts $Y_a$ are independent random variables. Two distributional settings are commonly studied:

- **Bernoulli** - $Y_a \sim Bernoulli(p_a)$, where $p_a$ is a parameter attached to cell $a$ with $0 < p_a < 1$.

- **Poisson** - $A_a$ is a positive real number and $Y_a \sim Poisson(\lambda_a, A_a)$, where $\lambda_a > 0$ is an unknown parameter attached to cell $a$. 
The scan statistic seeks to identify *hot spots* or *clusters* of cells that have an elevated response compared with the rest of the region. The following is a statement of the null and alternative hypotheses in the Bernoulli setting:

- $\mathcal{H}_0$: $p_a$ is the same and known for all cells in region $R$ – that is, there is no *hot spot*.

- $\mathcal{H}_1$: There is a nonempty zone $Z$ (connected union of cells) and parameter values $0 < p_0, p_1 < 1$ such that

$$
p_a = \begin{cases} 
p_1, & \text{for all cells } a \text{ in } Z \\
p_0, & \text{for all cells } a \text{ in } R - Z
\end{cases}
$$

(4.2.2)

with $p_1 > p_0$. The zone $Z$ specified in $H_1$ is an unknown parameter of the model. The null model, $H_0$, is the limit of $H_1$ as $p_1 \to p_0$; however, the parameter $Z$ is not identifiable in the limit.

We now assume that the region $R$ previously introduced is rectangular, which is defined by $[0,T_1] \times [0,T_2]$. Let $h_i = \frac{T_i}{N_i} > 0$, where $N_i$ are positive integers, $i = 1, 2$. In many applications the exact locations of the observed events in the rectangular region are unknown. What is usually available are the counts in small rectangular subregions. For $1 \leq i \leq N_1$ and $1 \leq j \leq N_2$, let $X_{ij}$ be the number of events that have been observed in the rectangular subregion (made up of cells) $[(i-1)h_1, ih_1] \times [(j-1)h_2, jh_2]$. We are interested in detecting unusual clustering of these events under the null hypothesis that $X_{ij}$ are independent and identically
distributed (i.i.d.) nonnegative integer valued random variables from a specified distribution. For $1 \leq i_1 \leq N_1 - m_1 + 1$ and $1 \leq i_2 \leq N_2 - m_2 + 1$, define:

$$Y_{i_1,i_2} = \sum_{j=i_2}^{i_2+m_2-1} \sum_{i=i_1}^{i_1+m_1-1} X_{ij} \quad (4.2.3)$$

to be the number of events in a rectangular region comprising $m_1$ by $m_2$ adjacent rectangular subregions with area $h_1h_2$ and the south-west corner located at the point $((i_1 - 1)h_1, (i_2 - 1)h_2)$. If $Y_{i_1,i_2}$ exceeds a preassigned value of $k$, we will say that $k$ events are clustered within the inspected region. We define a two-dimensional discrete scan statistic as the largest number of events in any $m_1$ by $m_2$ adjacent rectangular subregions with area $h_1h_2$ and the south-west corner located at the point $((i_1 - 1)h_1, (i_2 - 1)h_2)$:

$$S_{m_1 \times m_2; N \times N} = \max \{ Y_{i_1,i_2}; 1 \leq i_1 \leq N_1 - m_1 + 1, 1 \leq i_2 \leq N_2 - m_2 + 1 \} \quad (4.2.4)$$

For simplicity we assume that $N_1 = N_2 = N$ and $m_1 = m_2 = m$. We also abbreviate $S_{m \times m; N \times N}$ to $S_{m \times m}$. The scan statistic $S_{m \times m}$, defined in equation (4.2.4), has been used in [69] to test the null hypothesis of randomness that assumes the $X_{ij}$s are i.i.d. Bernoulli random variables with parameter $P(X_{ij} = 1) = p_0$, and similarly to test a constant-intensity Poisson null hypothesis. For the alternative hypothesis of clustering one can specify the region $Z$ in (4.2.2) to be a rectangular region given by:
\[ R(i_1, i_2) = [(i_1 - 1)h_1, (i_1 + m - 1)h_1] \times [(i_2 - 1)h_2, (i_2 + m - 1)h_2] \quad (4.2.5) \]

such that for any \( i_1 \leq i \leq i_1 + m - 1 \) and \( i_2 \leq j \leq i_2 + m - 1 \), \( X_{ij} \) has a Bernoulli distribution with parameter \( P (X_{ij} = 1) = p_1 \), where \( p_1 > p_0 \) or Poisson distribution with mean \( \theta_1 > \theta_0 \), respectively. For \( (i, j) \notin [i_1, i_1 + m - 1] \times [i_2, i_2 + m - 1] \), \( X_{ij} \) is distributed according to the distribution specified by the null hypothesis. We assume that \( i_1 \) and \( i_2 \) are unknown. If \( m \) is known \( S_{m \times m} \) is a generalized likelihood ratio test (GLRT) for testing the above hypotheses \([70]\). In this case, the null hypothesis is rejected whenever \( S_{m \times m} \geq k \), where \( k \) is determined from \( P (S_{m \times m} \geq k) = \alpha \), the probability of type I error. For this two-dimensional scan statistic to be useful, accurate approximations for \( P (S_{m \times m} \geq k) \) are needed. There are no exact results available for \( P (S_{m \times m} \geq k) \).

### 4.2.2 Bernoulli Model

Let \( X_{ij} \) be i.i.d. Bernoulli random variables with parameter \( P (X_{ij} = 1) = p_0 \), where \( 0 < p_0 < 1 \). For small and moderate values of \( P (S_{m \times m} \geq k) \) a quite accurate approximation has been derived in \([71]\):

\[
P (S_{m \times m} \geq k) \approx 1 - \left[ \frac{[P \{S_{m \times m}(m, m) \leq k-1]\}^{(N-m-1)^2} [P \{S_{m \times m}(m, m+1) \leq k-1]\}^{(N-m-1)(N-m)}}{[P \{S_{m \times m}(m+1, m+1) \leq k-1]\}^{(N-m)^2}} \right].
\]

(4.2.6)
Now, by defining

\[ S_{m \times m}(n_1, n_2) = \max \{Y_{i_1,i_2}; 1 \leq i_j \leq n_j - m + 1, j = 1, 2\} \quad (4.2.7) \]

we have from [71]:

\[ P\{S_{m \times m}(m, m) \leq k - 1\} = F_b(k - 1; m^2, p_0) \quad (4.2.8) \]

\[
P\{S_{m \times m}(m, m + 1) \leq k - 1\} = \sum_{s=0}^{k-1} F_b^2(k - 1 - s; m, p)b(s; m(m - 1), p_0) \quad (4.2.9)
\]

\[
P\{S_{m \times m}(m + 1, m + 1) \leq k - 1\} = \sum_{s_1, s_2=0}^{k-1} \sum_{t_1, t_2=0}^{k-1} \sum_{i_j=0,1 \leq j \leq 4} b_1(s_1)b_1(s_2)b_2(t_1)b_2(t_2)
\]
\[
\times p_0^{\sum_{j=1}^{4} i_j}(1 - p_0)^{4-\sum_{j=1}^{4} i_j}F_b(x; (m - 1)^2, p_0) \quad (4.2.10)
\]

and where for \( i = 1, 2, \)
\begin{align*}
b_1(s_i) &= b(s_i; m - 1, p_0) \\
b_2(t_i) &= b(t_i; m - 1, p_0) \\
b(j; N, w) &= \binom{N}{j} w^j (1 - w)^{N-j} \quad (4.2.11) \\
F_b(i; N, w) &= \sum_{j=0}^{i} b(j; N, w) \quad (4.2.12)
\end{align*}

\[ x = \min(k - 1 - s_1 - t_1 - i_1, \]
\[ k - 1 - s_2 - t_1 - i_2, \]
\[ k - 1 - s_2 - t_2 - i_3, \]
\[ k - 1 - s_2 - t_2 - i_4) \quad (4.2.13) \]

Bounds and approximations for \( P(S_{m \times m} \geq k) \) and are discussed in great detail in [67], [72].

### 4.2.3 Poisson Model

Let \( X_{ij} \) be i.i.d. Poisson random variables with mean \( \theta_0 > 0 \). For \( 1 \leq i_1 \leq N_1 - m_1 + 1 \) and \( 1 \leq i_2 \leq N_2 - m_2 + 1 \) define the events

\[ A_{i_1,i_2} = \sum_{j=i_2}^{i_2+m_2-1} \sum_{i=i_1}^{i_1+m_1-1} X_{ij} \geq k \quad (4.2.14) \]

Then,

\[ P(S_{m_1 \times m_2} \leq k - 1) = P \left( \bigcap_{i_1=1}^{N_1-m_1+1} \bigcap_{i_2=1}^{N_2-m_2+1} A_{i_1,i_2}^c \right). \quad (4.2.15) \]
where we have denoted as $A^c$ the complement event of $A$. To simplify the presentation of the results, we will discuss here only the case of $N_1 = N_2 = N$ and $m_1 = m_2 = m$. For fixed values of $1 \leq i_1 \leq N_1 - m_1 + 1$, one can accurately approximate $P\left(\bigcap_{i_2=1}^{N-m+1} A^c_{i_1,i_2}\right)$ using a product-type approximation discussed in [70]:

$$P\left(\bigcap_{i_2=1}^{N-m+1} A^c_{i_1,i_2}\right)$$

$$= P\left(\bigcap_{i_2=1}^{m+1} A^c_{i_1,i_2}\right) \prod_{i_2=m+2}^{N-m+1} \frac{P\left(\bigcap_{j=1}^{i_2} A^c_{i_1,j}\right)}{P\left(\bigcap_{j=1}^{i_2} A_{i_1,j}\right)}$$

$$\approx P\left(\bigcap_{i_2=1}^{m+1} A^c_{i_1,i_2}\right) \prod_{i_2=m+2}^{N-m+1} \frac{P\left(\bigcap_{j=i_2-m}^{i_2} A^c_{i_1,j}\right)}{P\left(\bigcap_{j=i_2-m}^{i_2-1} A_{i_1,j}\right)}$$

$$\approx q_{m,2m} \left(\frac{q_{m,2m}}{q_{m,2m-1}}\right)^{N-2m} \quad (4.2.16)$$

where

$$q_{m,m+l-1} = P(A^c_{1,1} \cap A^c_{1,2} \cap \cdots \cap A^c_{1,l}) \quad (4.2.17)$$

for $1 \leq l \leq N - m + 1$. The quantities $q_{m,m+l-1}$ are evaluated recursively in [73] for one-dimensional scan statistics. For self-consistency we report in Appendix 4.6 how to compute those quantities for a two-dimensional scan statistics. Two very accurate approximations for $P(S_{m \times m} \geq k)$ for the Poisson model are presented in the literature. Following the product-type approximation, the first of them is
given by [69]:

\[
P(S_{m \times m} \geq k) = 1 - q_{m,2m-1} \left( \frac{q_{m,2m}}{q_{m,2m-1}} \right)^{(N-2m+1)(N-m+1)}
\] (4.2.18)

The second belongs to a class of approximations in [71] and it is given by:

\[
P(S_{m \times m} \geq k) \approx 1 - \frac{[P\{S_{m \times m}(m+1,m+1) \leq k-1\}]^{(N-m)^2}}{[P\{S_{m \times m}(m+1,m) \leq k-1\}]^{(N-m-1)(N-m)}}
\times \frac{(q_{m,2m-1})^{(N-2m)(N-m-1)}}{(q_{m,2m})^{(N-2m+1)(N-m-1)}}
\] (4.2.19)

Based on extensive simulation we have concluded that approximation (4.2.19) is
more accurate than approximation (4.2.18), and therefore it will be used in this
chapter. Upper and lower bounds for these approximations have been developed
in [74].

4.3 A Two-Dimensional Bernoulli Grid of Sensors

4.3.1 Observations Model

Let us consider a scenario as depicted in Figure 4.3, where identical sensors are
deployed in a grid pattern in the sensor field, which we consider to be a square of
area $b^2$. Let us denote with $(x_s, y_s)$ for $s = 1, \ldots, M$ the coordinates of sensor $s$.
Noises at local sensors are i.i.d. and follow the Gaussian distribution with zero
mean and variance $\sigma_w^2$:

\[
w_s \sim \mathcal{N}(0, \sigma_w^2) \quad s = 1, \ldots, M
\] (4.3.1)
Fig. 4.2: The MA is depicted with its field of view which we consider to be a square of size $F_{ov}$. Only the sensors which detect the target are depicted. The different level curves correspond to a different target power. The MA traveling across the sensor network simply counts how many sensors are inside its field of view. We assume that the target (the black bold square) is a-priori uniformly distributed within the sensor network.
Fig. 4.3: A two-dimensional Bernoulli grid of sensors is depicted. $M = 625$ sensors are deployed in a grid pattern in the sensor field, which we consider to be a square of area $b^2$. The sensors which detect the target (the red bold circle) tend to form a cluster around the target (the black bold square). Here we used $b = 5, \sigma^2 = 1, \sigma_w^2 = 4, p_{fa} = 0.05$. 
We assume that sensors make their local decisions independently without collaborating with other sensors. We design each sensor $s$ to decide between the following (composite) hypotheses

\[ \mathcal{H}_0 : r_s = w_s \]

\[ \mathcal{H}_1 : r_s = y_s + w_s \]  

(4.3.2)

where $r_s$ is the received signal at sensor $s$, $y_s = \frac{a_s}{d_s}$, $a_s$ are i.i.d. GRV (Gaussian random variables) with zero-mean and variance $\sigma^2$ ($\sigma^2$ represents the power of the signal that is emitted by the target at distance $d = 1m$) and $d_s$ is the distance between the target and local sensor $s$:

\[ d_s = \sqrt{(x_s - x_t)^2 + (y_s - y_t)^2} \]  

(4.3.3)

and $(x_t, y_t)$ are the unknown coordinates of the target. We further assume that the location of the target also follows a uniform distribution within the sensor field. In (4.3.2) we used a model as in [75], [76], [77].

Assuming all the local sensors use the same threshold $\tau$ to make a decision and according to the Neyman-Pearson lemma [60], we have the local sensor-level false alarm rate and probability of detection given by:

\[ p_{fa} = 2Q \left( \sqrt{\frac{\tau}{\sigma^2_w}} \right) \]

\[ p_{ds} = 2Q \left( \sqrt{\frac{\tau}{\sigma^2_w + \frac{\sigma^2_a}{d_s^2}}} \right) \]  

(4.3.4)

where $Q(x) = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} \, d\xi$ is the unit Gaussian exceedance function. We denote the binary data from local sensor $s$ as $I_s = \{0, 1\}$ ($s = 1, \ldots, M$). $I_s$ takes the
value 1 when there is a detection; otherwise, it takes the value 0.

Remark 1. Note that a different signal power attenuation model can also be used, yielding different expressions for $p_{fa}$ and $p_{ds}$. For example, following the same model as in [56] we have:

$$y^2_s = \frac{P_0}{1 + \gamma d^n_s}, \quad (4.3.5)$$

where $P_0$ is the signal power emitted by the target at distance $d_s = 0$, $n$ is the signal decay exponent and $\gamma$ is an adjustable constant, where a larger $\gamma$ implies faster signal power decay. We will assume the model as in (4.3.2) but note that our results hold more generally.

4.3.2 The Scan Statistic

Assuming we know the total number of sensors $M$, we divide the square into $M$ small sub-squares. The location of the sensor inside each small sub-square (cell) is known\(^1\). Letting $h = \frac{b}{N}$, where $N$ is such that $N^2 = M$, we divide the square of area $b^2$ into $M$ cells such that each cell of area $h^2$ contains only one sensor\(^2\).

Let us denote by $c(i,j)$ the cell $[(i-1)h, ih] \times [(j-1)h, jh]$ (see Figure 4.3). We

\(^1\) Our results hold more generally: we do not need to know the exact location of the sensor inside the cell. We can also assume that the location of the sensor follows a uniform distribution in the cell.

\(^2\) This assumption is related to the theory of two-dimensional scan statistic [59]; further, we will remove this assumption in the case of Poisson model.
define $X_{ij}$ as the binary data from the local sensor $s$ inside $c(i, j)$ with $1 \leq i \leq N$ and $1 \leq j \leq N$. It is easy to verify that

$$\sum_{s=1}^{M} I_s = \sum_{j=1}^{N} \sum_{i=1}^{N} X_{ij} \quad (4.3.6)$$

For $1 \leq i \leq N$ and $1 \leq j \leq N$, let us denote with $X_{ij} = I_s$ be the number of events that have been observed in the cell $c(i, j)$ by the sensor $s$. $X_{ij}$ represents the binary data (“1” or “0”) of the sensor $s$ inside the cell $c(i, j)$. It is straightforward that under $\mathcal{H}_0$ (target absent), the $X_{ij}$ are i.i.d. Bernoulli random variables with $P(X_{ij} = 1) = p_{fa}$, whereas under $\mathcal{H}_1$ (target present), the $X_{ij}$ are discrete conditionally independent random variables with $P(X_{ij} = 1) = p_{d_s(i,j)}$, meaning that they are not identically-distributed$^3$. Now we assume that under $\mathcal{H}_1$, $p_{d_s(i,j)}$ can be approximated with $p_{fa}$ for a sensor $s$ whose distance to the target is such that $d_s^2 \gg \frac{\sigma^2}{\sigma_w^2}$. In fact we have that

$$p_{d_s(i,j)} = 2Q\left( \sqrt{\frac{\tau}{\sigma_w^2 + \frac{\sigma^2}{\sigma_w^2}}} \right) \approx 2Q\left( \frac{\sqrt{\tau}}{\sigma_w^2} \right) = p_{fa} \quad (4.3.7)$$

for $d_s^2 \gg \sigma^2 / \sigma_w^2$. Since the target power attenuates as function of the distance from the target, we expect that there is a cluster of sensors which are stronger (closer to the target), under the hypothesis $\mathcal{H}_1$.

$^3 p_{d_s(i,j)}$, which is the local detection probability for the sensor $s$ which is located in the cell $c(i, j)$; $p_{d_s(i,j)}$ is greater than $p_{fa}$, but is otherwise unknown.
4.3.3 False Alarm Probability at the MA

The MA travels across the sensor network (see Figure 4.2) and sequentially collects the local binary decisions from sensors that lie inside its overlapping FOV’s, which we consider to be squares of size $F_{ov} = f_{ov}h$ ($f_{ov}$ integer). The sequential fusion rule at the MA for $1 \leq i_1 \leq N - f_{ov} + 1$ and $1 \leq i_2 \leq N - f_{ov} + 1$ is given by:

$$
\begin{align*}
  &\begin{cases}
    Y_{i_1, i_2} \geq k & \Rightarrow \text{ decide } \mathcal{H}_1, \\
    \text{otherwise} & \Rightarrow \text{ MA continues to scan.}
  \end{cases} \\
\end{align*}
$$

(4.3.8)

where $Y_{i_1, i_2} = \sum_{j=i_2}^{i_2+f_{ov}-1} \sum_{i=i_1}^{i_1+f_{ov}-1} X_{ij}$. At the MA the probability of false alarm $P_{fa}$ is:

$$
P_{fa} = P(S_{f_{ov} \times f_{ov}} \geq k \mid \mathcal{H}_0) 
$$

(4.3.9)

We note that equation (4.3.9) can be evaluated using the approximation as in equation (4.2.6) after substituting $f_{ov}$ with $m$ and $p_{fa}$ with $p_0$.

Remark 2. The key point is that while the exceedance probability for one location of the FOV (rectangular scan region or window) of (4.2.4) is trivial – and similarly for many non-overlapping windows it is easy – to obtain a powerful GLRT we need to “drag” the window as with the MA. The exceedance probability for (4.2.4) is not at all obvious for overlapping windows, but that is exactly what the scan statistics literature provides. The contribution and importance of scan statistics is their ability to do that, to provide an answer to the question: “How many counts constitute an abnormality?”.
Fig. 4.4: Probability of false alarm for the MA $P_{fa}$ versus probability of false alarm for the local sensor $p_{fa}$. Here we have $b = 5$, $\sigma^2 = 1$, $\sigma_w^2 = 4$, $M = 625$, $k = 7$, $f_{ov} = 6$ and $F_{ov} = 1.2$.

In Figure 4.4 we have plotted the global probability of false alarm $P_{fa}$ for the MA versus the local probability of false alarm $p_{fa}$ for the sensor. In Figure 4.4, the curves obtained by using the scan statistic approximations in equation (4.2.6) and those by simulations (based on 10000 Monte Carlo runs) are plotted. Figure 4.4 shows that the approximation in equation (4.2.6) is very accurate for small values of $P_{fa}$, as expected from the approximations of scan statistic theory [59]. The question of the optimum field-of-view $F_{ov} = f_{ov} h$ (with $f_{ov}$ integer) to use arises
and it will be addressed in the next section.

4.3.4 Optimized MA Window Size

In this section we are interested in finding an answer to the following problem: Does an optimum $F_{ov}$, which maximizes the probability of detection $P_d$ for the MA, while maintaining fixed the probability of false alarm $P_{fa}$, exist? Does the following optimization problem have a solution?

$$
\max_{F_{ov}: P_{fa} = \alpha} P_d
$$

(4.3.10)

where $P_d$ is defined as

$$
P_d = P(S_{fov \times fov} \geq k \mid \mathcal{H}_1)
$$

(4.3.11)

If the “signature” of an anomaly was known and matched a possible MA window the scan statistic (in this case $S_{fov \times fov}$ is the GLRT for testing the two hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ [70]), then the question would scarcely be worth asking since the optimum $F_{ov}$ coincides with the “signature” of the anomaly. However, according to the model of Section 4.3.1 the density of sensors that report “1” is unknown and decreases smoothly as the distance from the source increases, and while it would be appealing to make the window large enough so that all or most of them be encompassed, it is clear that doing so increases the likelihood of false alarms. It will be recalled that for hypothesis testing problems involving discrete distributions, it is usually not possible to choose a critical region consisting of realizable
values of the statistic of size exactly $\alpha$, where $\alpha$ is some prescribed value. The use of a randomized test [78] in order to solve the constrained optimization problem in (4.3.10) is needed. By defining $\alpha_1$ and $\alpha_2$ as follows:

$$
P(S_{fov \times fov} \geq k_1) \approx \alpha_1 < \alpha
$$

$$
P(S_{fov \times fov} \geq k_1 - 1) \approx \alpha_2 > \alpha \tag{4.3.12}
$$

then we use a “coin-flip” decision with probability

$$
p = \frac{\alpha - \alpha_1}{\alpha_2 - \alpha_1} \tag{4.3.13}
$$

whenever $S_{fov \times fov} = k_1$.

As we can see in Figure 4.5, there does exist an optimal field-of-view $F_{ov}$ that maximizes the probability of detection $P_d$ for the MA. By employing this optimum $F_{ov}^{opt}$, a significant improvement in $P_d$ can be achieved. In figure 4.6 the optimal $F_{ov}^{opt}$ versus $\sigma - \sigma^2$ represents the power of the signal that is emitted by the target at distance $d_s = 1m$ – is plotted. Not surprisingly, we observe an approximately linear relationship between the optimal scan-box dimension $F_{ov}^{opt}$ and the target amplitude, and also can see that there is a saturation point beyond that the optimal size of the MA’s field of view stays constant (the power of the signal is so high that all the sensors report “1”s.)

An interesting alternative approach uses Multiple Window Discrete Scan Statistics [79]. Since the shape and the size of the cluster of the alarmed sensors (sensors
that report “1”) is unknown, it might prove to be ineffective to arbitrarily choose a window size for performing the test. The testing procedure is based on several fixed window size scan statistics: $S_{\text{fov}_1 \times \text{fov}_1}, \ldots, S_{\text{fov}_n \times \text{fov}_n}$, where $\text{fov}_1 \leq \ldots \leq \text{fov}_n$. This testing procedure will reject $H_0$ if for some $1 \leq j \leq n$, $S_{\text{fov}_j \times \text{fov}_j} \geq k_j$,

where:

$$P\left\{ \bigcup_{j=1}^{n} \left( S_{\text{fov}_j \times \text{fov}_j} \geq k_j \right) \right\} = \alpha_m \quad (4.3.14)$$

is the desired probability of false alarm for the MA. The MA can be thought of as dragging a family of windows – some large and some small – simultaneously. Approximations for determining the thresholds $k_j$ are given in [79] where numerical results for the power of the multiple window scan statistic test are presented along with a comparison with a multiple tests approach [80].

### 4.4 A Two-Dimensional Poisson Field of Sensors

#### 4.4.1 The Scan Statistic

In Section 4.3.2 we studied the case of a grid of sensors where the number of the sensors was known and was equal to $M$. It is probably more realistic to assume that:

- the sensors are deployed randomly in the SN; and

- some of them are out of communication range of the MA, malfunctioning, or out of battery power.
Fig. 4.5: The probability of detection $P_d$ for the MA versus the field-of-view $F_{ov}$ is plotted for a fixed value of $\alpha$. Here we have $b = 5$, $\sigma^2 = 1$, $\sigma_w^2 = 4$, $p_{fa} = 0.04$, $M = 625$, and $\alpha = 0.1$ Simulations are based on 10000 runs.
Fig. 4.6: The optimal $F_{ov}^{\text{opt}}$ versus $\sigma$ ($\sigma^2$ is the power of the signal that is emitted by the target at distance $d_s = 1m$) is plotted for a fixed value of $\alpha$. Here we have $b = 5$, $\sigma_w^2 = 4$, $p_{f_a} = 0.04$, $M = 625$, and $\alpha = 0.1$. Simulations are based on 10000 runs.
Fig. 4.7: A two-dimensional Poisson field of sensors is depicted. The sensors which detect the target (the red bold circle) tend to form a cluster around the target (the black bold square). Here we used $\lambda = 20$, $b = 5$, $\sigma^2 = 1$, $\sigma_w^2 = 4$, $p_{fa} = 0.05$. 
Therefore, at any given time, the total number of sensors that work properly in
the SN is a random variable. We assume that the initial distribution of sensors
over the space is a homogeneous Poisson process with intensity $\lambda$, therefore the
number of sensors $M$ within the sensor field is a random variable that follows a
Poisson distribution:

$$p(M) = \frac{\lambda^M e^{-\lambda}}{M!}, \quad M = 0, \ldots, +\infty \quad (4.4.1)$$

where $\lambda_b = \lambda b^2$. Given $M$, the locations of the sensors are i.i.d., and follow an
uniform distribution in the sensor field

$$f(x_s, y_s) = \begin{cases} \frac{1}{b^2} & 0 \leq x_s, y_s \leq b, \\ 0 & \text{otherwise} \end{cases}$$

for $s = 1, \ldots, M$, where $(x_s, y_s)$ are the coordinates of sensor $s$. Since each
sensor decision is independent and based on the signal strength at its location,
the local decision-making at each sensor can be viewed as a location-dependent
thinning procedure [81] of the original sensor distribution with probability $p_{fa}$ and
$p_{ds}$, under $\mathcal{H}_0$ and $\mathcal{H}_1$ respectively, and the distribution of the alarmed sensors
(sensors with $I_s = 1$) forms a nonhomogeneous spatial Poisson process. Hence the
distribution of alarmed sensors is a nonhomogeneous Poisson process whose local
intensity is given by $\lambda p_{fa}$ and $\lambda p_{ds}$ under $\mathcal{H}_0$ and $\mathcal{H}_1$ respectively.

We measure the sensor field directly via a scan statistic; that is, the sensors
are Bernoulli with random location. Unfortunately there are no extant results
Fig. 4.8: Probability of false alarm for the MA $P_{fa}$ versus probability of false alarm for the local sensor $p_{fa}$. Here we have $b = 5$, $\sigma^2 = 1$, $\sigma_w^2 = 4$, $\lambda = 20$, $N = 10$, $k = 6$, $f_{ov} = 2$ and $F_{ov} = 1$. 
for scan statistics in this case. However, there are results for fixed numbers of Poisson variates. To exploit these, we model that the sensors still be located on a grid, but now the “gridded” sensor reports being of the number of “1”s observed within each sensor’s hinterland, a rectangle (w.l.o.g. but with nicer notation, a square) of dimension $h_1 \times h_2 = h^2$, as in Section 4.2.3 (see Figure 4.7). Then it is straightforward that under $\mathcal{H}_0$ (target absent), $X_{ij}$ are i.i.d. Poisson random variables with mean $\lambda_0 = \lambda h^2 p_{fa}$, whereas under $\mathcal{H}_1$ (target present) $X_{ij}$ are discrete conditionally independent Poisson random variables but not identically distributed with mean $\lambda_{s(i,j)} = \lambda h^2 p_{d_s(i,j)}$. Now we assume that under $\mathcal{H}_1$, $p_{d_s(i,j)}$ can be approximated with $p_{fa}$ as in (4.3.7) and the same analysis as described for the Bernoulli case can be done. In Figure 4.8 we have plotted the global probability of false alarm for the MA $P_{fa}$ versus the local probability of false alarm $p_{fa}$. In Figure 4.8, the curves obtained by using the scan statistic approximations in (4.2.19) and those by simulations (based on 10000 Monte Carlo runs) are plotted. Figure 4.8 shows that the approximation in (4.2.19) is more accurate. In fact, a detailed simulation analysis confirmed that the approximation in (4.2.19) works better than the approximation in (4.2.6) for the Bernoulli case, especially for small values of $p_{fa}$. 
**Fig. 4.9:** The MA travels across the Poisson sensor field. The overlapping windows (the two-dimensional scan statistic test) and the non-overlapping windows (non-overlapping test) are depicted. The non-overlapping test is sensitive to the relative target position with respect to the scanning window, which turns out to affect the power of the test.
4.4.2 The Disjoint-Window Test

Here we study the performance of a *non-overlapping* (or disjoint-window) test for the Poisson model described in the section 4.4. The size (probability of false alarm) of such a test is trivial to calculate, due to the independence of the windows’ counts, versus the scan statistic expressions. But how much do we lose by constraining the window to “hop” rather than slide?

We now suppose that the MA travels across the sensor sensor network and scans the network using non-overlapping windows. It collects all the detections reported by local sensors which lie in its field of view. Here we assume that the field of view of the MA is a square of size $F_{ov}$. We divide the square of size $b$ in $W$ windows of size $F_{ov}$, with $W = \frac{b^2}{F_{ov}}$ being an integer. We need to evaluate the probability distribution of the highest number of counts detected over the whole set of windows. This calculation can be done exactly. In each window the number of alarmed sensor $S$ under $H_0$ is dictated by the following Poisson distribution

$$p(S, \lambda_{F_{ov}} | H_0) = \frac{e^{-\lambda_{F_{ov}}} \lambda_{F_{ov}}^S}{S!}$$  \hspace{1cm} (4.4.2)

where $\lambda_{F_{ov}} = \lambda F_{ov}^2 p_{fa}$. Let us define with $Z = \max (S_i, i = 1, \ldots, W)$ the highest number of counts detected over the all windows. It can be easily shown that the probability density function of $Z$ is given by:

\footnote{We use interchangeably the terms window and field of view.}
\[
P(Z = l) = \sum_{k=1}^{W} \binom{W}{k} \left( \sum_{n=0}^{l-1} \frac{e^{-\lambda_{Fov} \lambda_{Fov}^n}}{n!} \right)^{W-k} \times \left( \frac{e^{-\lambda_{Fov} \lambda_{Fov}^l}}{l!} \right)^{k} \quad (4.4.3)
\]

At the MA the null hypothesis \( (H_0) \) is rejected whenever \( Z \geq k \), and

\[
P_{fa} = P(Z \geq k) = 1 - \sum_{l=0}^{k-1} P(Z = l) . \quad (4.4.4)
\]

is the probability of false alarm.

In Figure 4.10 the probability of detection for the MA \( P_d \) versus \( x_t \) (the \( x \) target location) for a fixed value \( \alpha \) of the probability of false alarm for the MA \( P_{fa} \) is plotted. The power of the scan statistic test is more robust than the disjoint-window test with respect to the target position (see Figure 4.10). The performances of the disjoint-window test are strictly related to the relative position of the target with respect to the center of mass of the disjoint-window. The two-dimensional scan statistic test outperforms the non-overlapping test as we can see from the Receiver Operating Characteristic (ROC) curves in Figures 4.11 and 4.12, especially at low \( P_{fa}(MA) \).

### 4.4.3 The Counting Rule Test

In [56] the counting decision fusion rule, which is based on the total number of local detections from local sensors, was proposed: according to [55] at the fusion
center the system-level decision is made by first counting the number of detections made by local sensors and then comparing it with a threshold $T$:

$$
\Lambda = \sum_{s=1}^{M} I_s = \sum_{j=1}^{N} \sum_{i=1}^{N} X_{ij} \begin{cases} 
\geq T & \text{decide } \mathcal{H}_1, \\
\leq T & \text{decide } \mathcal{H}_0, 
\end{cases}
$$

(4.4.5)

where $X_{ij} = \{0, 1\}$ is the local decision made by sensor $s$ which is located in the cell $c(i,j)$.

In Figure 4.13, the ROC curves obtained by simulations (based on 1000 Monte Carlo runs) for the counting rule and the scan statistic test are shown. We note that the ROC corresponding to the optimum field-of-view $F_{ov}^{opt}$ of the MA clearly demonstrates the substantial improvement over the ROC obtained for the counting rule.

Figure 4.14 provides the histograms of the number of sensors $N_s$ which report their decision to the MA (scan statistic) and to the fusion center (counting rule). For the simulations carried out, the average number of sensors which report their decision to the MA (we denote it with $N_{scan}^s$) is about 17 when no target is present in the sensor network and 19 when it is. In the counting rule case we have instead that $N_{count}^s$ is respectively 31 and 42.4 – the scan-statistic test is more efficient. We mention in passing that a multiple scan statistic [67] can statistically characterize the number of those windows.
4.5 Conclusions and Future Directions

We have proposed and studied a new sequential test based on two-dimensional scan statistics, based on the number of counts reported by a sliding window “dragged” across an area of interest by a mobile agent. This test is particularly appropriate when the phenomenon of interest is a localized disturbance, such as the magnetic or acoustic signature of a low-observable target that can only be registered at sensors that are nearby. The form of the test is obvious, and matches with the GLRT under some reasonable assumptions; but the key feature of a scan statistic is its precise approximation for the false-alarm rate (exceedance probability). We have shown how the system parameters can affect the detection performance, and numerical results corroborate the theory. An important result is that the proposed fusion rule outperforms both the counting fusion rule and the “hopping window” test. We note that the scan statistic approach might be thought a (two-dimensional) “quickest” detection scheme, in that it stops as soon as a departure from the null hypothesis is declared.

Ongoing research includes using a Multiple Window Discrete Scan Statistic [79] to improve on our model, and to offer analysis for the probability of detection. Further, in [82], [83] anomalies are sought by comparing two successive scans: if in the second there is a region that contains significantly more detections than in the first, a detection is declared. That work is very interesting, and these references
include performance measures for the asymptotic case; but for an exact means to compute the threshold a scan statistic seems very appropriate.

4.6 Appendices

In order to obtain $q_{m,2m}$ we use the following recursive formulas as in [73] but with a slight modification in order to account for the two-dimensional scan statistics:

$$ q_{m,2m} = \sum_{y=0}^{k} b_{2(m)}^{k}(y) \quad (4.6.1) $$

$$ b_{2(m)}^{k}(y) = \begin{cases} 
\sum_{\eta=0}^{y} \sum_{\nu=0}^{k-y+\eta} b_{2m-1}^{k}(\nu, y - \eta) f(\eta) & \text{for } y \leq k, \\
0 & \text{for } y > k
\end{cases} \quad (4.6.2) $$

For $m < i < 2m$, $b_{i}^{k}(x, y)$ are calculated recursively, using

$$ b_{m+1}^{k}(x, y) = \begin{cases} 
\sum_{\nu=0}^{k-x} f(\nu) f(y) f(\nu) f_{m-1}(x) & \text{for } x + y \leq k, \\
0 & \text{for } x + y > k
\end{cases} \quad (4.6.3) $$

and

$$ b_{i}^{k}(x, y) = \begin{cases} 
\sum_{\eta=0}^{y} \sum_{\nu=0}^{k-(x+y)+\eta} b_{i-1}(x + \nu, y - \eta) \times \\
f(\nu) f(\eta) f_{2m-1}(x) f_{2m-1+1}(x+y) & \text{for } x + y \leq k,
0 & \text{for } x + y > k
\end{cases} \quad (4.6.4) $$
In the above equations we used the functions $f_m(y)$ and $f(y)$ which are defined as:

\[
\begin{align*}
    f_m(y) &= P\left\{ \sum_{i=1}^{m} \sum_{j=1}^{m} X_{ij} = y \right\} \quad (4.6.5) \\
    f(y) &= f_1(y)
\end{align*}
\]
Fig. 4.10: Probability of detection for the MA $P_d$ versus the target location in the $x$ direction ($x_t$) is plotted for the scan statistic test and the disjoint-window test respectively. Here we have $b = 5$, $\sigma^2 = 1$, $\sigma_w^2 = 4$, $\lambda = 20$, $N = 25$, $f_{ov} = 5$, $F_{ov} = 1$, $p_{fa} = 0.05$, and $\alpha = 0.1$ for both tests (in order to get the prescribed value $\alpha = 0.1$ the use of the randomized test as previous described is needed). Simulation are based on standard Monte Carlo counting process, involving 10000 runs.
Fig. 4.11: ROC curves obtained by simulations for the scan statistic test and the disjoint-window test. Here we have $b = 5$, $\sigma^2 = 1$, $\sigma^2_w = 4$, $\lambda = 20$, $N = 25$, $f_{ov} = 5$, $F_{ov} = 1$, $p_{fa} = 0.02$. Simulation are based on standard Monte Carlo counting process, involving 10000 runs.
Fig. 4.12: ROC curves in semilogarithmic scale obtained by simulations for the scan statistic test and the disjoint-window test. Here we have $b = 5$, $\sigma^2 = 1$, $\sigma_w^2 = 4$, $\lambda = 20$, $N = 25$, $f_{ov} = 5$, $F_{ov} = 1$, $p_{fa} = 0.02$. Simulation are based on standard Monte Carlo counting process, involving 10000 runs.
Fig. 4.13: ROC curves in semilogarithmic scale obtained by simulations for the scan statistic test and the counting rule test. Here we have $b = 5$, $M = 625$, $\sigma^2 = 1$, $\sigma^2_w = 4$, $N = 25$, $p_{fa} = 0.05$. Simulation are based on standard Monte Carlo counting process, involving 10000 runs.
Fig. 4.14: Number of sensors $N_s$ which report their decision to the MA (scan statistic) and to the fusion center (counting rule). Here we have $b = 5$, $M = 625$, $\sigma^2 = 1$, $\sigma_w^2 = 4$, $N = 25$, $p_{fa} = 0.05$. Simulation are based on standard Monte Carlo counting process, involving 10000 runs.
Chapter 5

A Repeated Significance Test With Applications To
Sequential Detection In Sensor Networks

5.1 Introduction

In recent years, there has been intensive research interest in signal processing with distributed sensors due to their low cost, increased computational capability and wide range of current and potential applications in detection and estimation, reconnaissance and surveillance, etc., e.g. [84]. In the classical centralized architecture, it is assumed that all the local sensors (such as radar, sonar, etc...) communicate all their data to a Fusion Center (FC) that performs optimal detection/estimation of the object of interest based on conventional statistical techniques. In a decentralized architecture some preliminary processing of data is carried out at each sensor and the new summarized information is transmitted from each sensor to the FC. In this paper we consider a decentralized sequential detection problem where the sensors make conditionally independent observations.
about the hypotheses (say $H_0$ and $H_1$). Due to bandwidth constraints, each sensor sends a quantized version of its observation to the FC. Whenever a transmission between the local sensor and the FC takes place, a cost is paid by the SN. Such cost can be associated to the bandwidth/energy which is necessary for the transmission and the communication between the sensor and the FC. Unlike most of the papers in literature, among which we cite [85], [86], [87], here we do not impose an average or peak constraint, but instead a point constraint. More specifically, the FC will stop sampling the sensors either when the cumulative sum of the local sensors’ observations will exceed a prescribed threshold or when the cumulative expenditure of energy will be larger than the total energy budget that the SN is allowed to spend. We should note that the the cumulative expenditure of energy is a random variable and this makes the problem hard to study. The introduction of a randomly truncated sequential test is needed. In order to study this problem, we use the framework of a Repeated Significance Test (RST), introduced by Armitage [88], a useful sequential test that is familiar within the statistics community but that seems not widely known among engineers. We propose a truncated sequential test with truncation time based on a random stopping time. The latter is the time instant when the total energy budget is spent by the SN. Several articles have dealt with decentralized detection with energy constraints [89], [90], [91], while decentralized sequential detection has been studied in [92], [93], [94] and
more recently in [95], [96], [97], [85]. In [93] the authors study the case where the fusion center implements a sequential test using the likelihood ratio as statistic. In order to implement the likelihood ratio a full knowledge of the underlying distributions is needed. In many practical situations the FC may not have information about the conditional distributions of the observations. A more realistic approach to estimation and detection in a SN is to assume a lack of knowledge in the distributions [98], [99], [100]. We assume that the FC does not know the statistical distribution of the sensors’ observation, thus the use of a distribution free test is needed. When the conditional distributions are not known, they can be learned as in [101], [102]. Our approach is completely different, and, by performing a RST at the FC, we exploit the applicability of the RST for a broad class of distributions.

The main results of this chapter are twofold:

- We introduce a randomly truncated sequential test using the framework of the RST and we characterize its performance.

- We apply the proposed test to a new decentralized sequential problem in SN with a point communication constraint where the distribution of the sensors’ observations are unknown to the FC.

The chapter is organized as follows. In Section 5.2, we introduce formally the classical RST and its power function. We present a FCLT that is used in evaluating the probability of type I error, which is of utmost importance to the design
and analysis of the RST and establishing its curvilinear boundaries. We present an example that illustrates the RST being considered as a repeated generalized likelihood ratio test. In Section 5.3, we compare the performance of the RST to that of the SPRT in testing for a normal mean. In Section 5.4, we introduce a new randomly truncated RST based on a stopping time that controls the total available resources for the sequential testing problem. This stopping time is motivated by an important application of the proposed test to decentralized sequential detection in sensor networks with communication constraints, which is discussed in detail in Section 5.5. In Section 5.4 we also derive theoretical results employed in implementing the proposed randomly truncated RST and also present an effective approach to evaluate the power function and investigate its performance. Finally, in Section 5.6, a sequential detection problem with observations at random times is discussed, and we summarize in Section 5.7.

5.2 Classical Repeated Significance Test (RST)

The optimality of the sequential probability ratio (SPRT) test has been extensively studied [103]. However it applies only to simple hypotheses, and further there is no upper bound on the stopping time of a SPRT and of other sequential tests, and therefore the introduction of a truncation is often necessary. The literature on truncated SPRT’s or others is also quite rich, and we could cite [104] for a
fairly complete overview. The interested reader could also refer to [105] and [106]. In [106] the authors derive asymptotically correct approximations for the operating characteristic function and the average sample number of a truncated test that can be recast as a mixture of a SPRT and a fixed sample size test. In all previous works on truncated sequential tests the truncation time is deterministic and it is fixed a priori in order to reduce the maximum sample size of the test.

Repeated significance tests were introduced in Armitage [88], as an alternative to the SPRT and its variants, and since then have been discussed extensively in the statistical literature. Major developments in the theory and applications of repeated significance tests have been presented in [104], [107], [108], [109], [110], [111], [112], [113] and [114]. In [114], Lerche showed that the repeated significance test is optimal, in the sense of minimizing the Bayes risk, in testing the sign of the drift of Brownian motion with a normal prior on the parameter space. Lerche highlighted the optimality of the RST which can be seen as the natural counterpart to Wald’s sequential probability ratio test (SPRT) for testing composite hypotheses without an indifference zone. He concluded that the RST can be viewed as an adaptive SPRT. In [104], the use of RST is discussed for data from a normal or a t distribution. In [108], the use of RST was extended to data from a multi-parameter exponential family. Nonparametric RST’s have been discussed in [115], [116] and [117]. In [104], Chapter IV, Siegmund gives theoretical justi-
fication for viewing the RST as a repeated generalized likelihood ratio test. This feature of the RST provides additional support for its effectiveness in sequential testing of composite hypotheses in parametric and non-parametric models.

More specifically, let \( \{X_i\}_{i \geq 1} \) be a sequence of independent identically distributed (i.i.d.) random variables with mean \( E(X_i) = \mu \) and variance \( \text{Var}(X_i) = \sigma^2 \). Suppose that one wants to test

\[
H_0 : \mu = 0 \quad \text{vs} \quad H_1 : \mu \neq 0,
\]

where the variance \( \sigma^2 \) is generally unknown and may be different under each hypothesis. The classical FCLT (or invariance principle) [118] for \( \{X_i\}_{i \geq 1} \) can be employed to develop a repeated significance test.

More specifically, let \( S_n = X_1 + \cdots + X_n \), with mean \( E_0(X_i) = 0 \) under \( H_0 \). We define \( S_n(t) \) to be a random element of \( C[0,1] \) obtained by a linear interpolation between points

\[
\left( 0, 0 \right), \left( \frac{1}{n}, \frac{S_1}{\sigma \sqrt{n}} \right), \left( \frac{2}{n}, \frac{S_2}{\sigma \sqrt{n}} \right), \ldots, \left( 1, \frac{S_n}{\sigma \sqrt{n}} \right).
\]

(5.2.1)

where with \( C[0,1] \) we denoted the space of all continuous functions with domain the interval \([0,1]\) [119]. Under the verified hypotheses of Donsker’s theorem [for instance, [118], p. 406, Theorem (6.6)] we have that

\[
S_n(\cdot) \xrightarrow{D} W(\cdot), \quad \text{as } n \to \infty,
\]

(5.2.2)

in the sense of \( C[0,1] \) equipped with the uniform metric [119], where \( W(\cdot) \) is the
standard Brownian motion and the symbol \( \xrightarrow{D} \) denotes convergence in distribution [119].

Fix \( 0 < t_0 < 1 \). Let

\[
h(f) = \max_{t_0 \leq t \leq 1} \frac{|f(t)|}{\sqrt{t}} \leq \frac{1}{\sqrt{t_0}} ||f||, \quad (5.2.3)
\]

with \( h(\cdot) \) being a continuous functional \( h(\cdot) : C[0, 1] \rightarrow \mathbb{R} \). With \( ||f|| \) we denote \( ||f|| = \max_{0 \leq t \leq 1} |f(t)| \). Therefore, if \( n_0, n \rightarrow \infty \) in such way that \( n/n_0 \rightarrow 1/t_0 \) by invoking the FCLT [for instance, [118], p. 406, Theorem (6.7)], we have that

\[
h(S_n(\cdot)) \xrightarrow{D} h(W(\cdot)), \text{ as } n_0, n \rightarrow \infty, \quad (5.2.4)
\]

Therefore,

\[
P_0 \left( \max_{n_0 \leq k \leq n} \left| \frac{S_k}{\sigma \sqrt{k}} \right| \geq b \right) = P_0 \left( \max_{n_0 \leq k \leq n} \left| \frac{S_k}{(\sigma \sqrt{n}) \sqrt{k/n}} \right| \geq b \right) \rightarrow \alpha(t_0, b), \quad (5.2.5)
\]

where

\[
\alpha(t_0, b) = P_0 \left( \max_{t_0 \leq t \leq 1} \left| \frac{W(t)}{\sqrt{t}} \right| \geq b \right). \quad (5.2.6)
\]

As was shown in [117] (p. 244) for a given significance level (probability of Type I error) \( \alpha = \alpha(t_0, b) \) and a value of the power function at a specified alternative, one can construct a repeated significance test with the following associated stopping time

\[
\tau = \inf \left\{ k : k \geq n_0 \left| \frac{S_k}{\sigma \sqrt{k}} \geq b \right| \right\}, \quad (5.2.7)
\]
where $n_0 > 0$ is the initial sample size. The repeated significance test stops sampling at $\min(\tau, N)$ (with $N$ being the truncation time) and rejects the null hypothesis $H_0$ if and only if $\tau \leq N$, (see Figure 5.1). Otherwise, the null hypothesis $H_0$ is accepted.

In the more interesting case that, in addition to composite $\mu$, $\sigma^2$ is also unknown. We redefine $S_n(t)$ to be a random element of $C[0,1]$ obtained by a linear interpolation between points

\[
(0,0), \left(\frac{1}{n}, \frac{S_1}{\sqrt{B_n}}\right), \left(\frac{2}{n}, \frac{S_2}{\sqrt{B_n}}\right), \ldots, \left(1, \frac{S_n}{\sqrt{B_n}}\right).
\]

(5.2.8)

where $\frac{1}{n}B_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$ is the sample variance and $\bar{X}$ is the sample mean, $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$. Using the same arguments as above and if $n_0, n \to \infty$ in such
way that $B_n/B_{n_0} \to 1/t_0$, then

$$P_0 \left( \max_{n_0 \leq k \leq n} \left| \frac{S_k}{\sqrt{B_k}} \right| \geq b \right) \to \alpha(t_0, b), \quad (5.2.9)$$

Unlike the linear boundaries as in the SPRTs, the boundaries defined by the equation (5.2.7) are curved in $S_n$. Problems involving curved stopping boundaries are more difficult to investigate and only seldom can one obtain exact results. In order to motivate the use of the curved boundaries as in (5.2.7) we provide the following example.

**Example 5.2.1.** Let $\{X_i\}_{i \geq 1}$ be a sequence of i.i.d. Gaussian random variables with mean $E(X_i) = \mu$ and variance $Var(X_i) = 1$. Let $S_n = X_1 + \cdots + X_n$. Let suppose that one wants to test $H_0 : \mu = 0$ against the composite alternative $H_1 : \mu \neq 0$. A reasonable sequential test could be defined in terms of the following stopping rule:

$$T = \inf \left\{ n : \max_{\mu} l(n, S_n; \mu) \geq \gamma \right\}$$

$$\quad = \inf \left\{ n : S_n^2/2n \geq \gamma' \right\} = \inf \left\{ n : \left| \frac{S_n}{\sqrt{n}} \right| \geq \gamma'' \right\} \quad (5.2.10)$$

where with $l(n, S_n; \mu)$ we have denoted the likelihood ratio between the two hypotheses $H_0$ and $H_1$.

The justification and usefulness of the test in (5.2.7) thus finds a nice interpretation as a repeated generalized likelihood ratio test (5.2.10).
Fig. 5.2: ASN as function of $\mu_1$ is plotted for the SPRT and the RST, respectively. Here we used $n_0 = 1$, $\sigma = 1$. The values of $b$ and $N$ for the RST and the values of $\gamma_1$ and $\gamma_0$ for the SPRT, were chosen such that $\alpha$ and $\beta$ were kept constant to the prescribed nominal value of 0.05.

5.3 SPRT vs RST in testing for a normal mean

In this section we compare the performance of the SPRT with that of the RST for the case of simple hypotheses. More specifically, let $\{X_i\}_{i \geq 1}$ be a sequence of i.i.d. Gaussian random variables with mean $E(X_i) = \mu$ and known variance $Var(X_i) = \sigma^2$. Let us suppose that one wants to test $H_0 : \mu = 0$ against the simple alternative $H_1 : \mu = \mu_1$ (with $\mu_1 > 0$). As is well known, the optimal test that, for a fixed level of performance (probabilities of Type I and Type II error),
minimizes the Average Sample Number (ASN) both under $H_0$ and under $H_1$, is the sequential probability ratio test (SPRT) [120] for $n = 1, 2, \ldots$

$$\sum_{i=1}^{n} \log \frac{f_1(x_i)}{f_0(x_i)} \begin{cases} \geq \gamma_1 & \Rightarrow \text{decide } H_1, \\ \leq \gamma_0 & \Rightarrow \text{decide } H_0, \\ \text{otherwise} & \Rightarrow \text{go ahead}, \end{cases} \quad (5.3.1)$$

where $f_1(x)$ and $f_0(x)$ denote the probability density function $f$ under $H_1$ and $H_0$ respectively. Here the thresholds $\gamma_0$ and $\gamma_1$ are set to achieve the desired performance level in terms of probability of Type I error $\alpha$ and of Type II error $\beta$. Simple approximations for $\gamma_0$ and $\gamma_1$ are given by [120]:

$$\gamma_0 \approx \log \frac{\beta}{1 - \alpha}, \quad \text{and} \quad \gamma_1 \approx \log \frac{1 - \beta}{\alpha} \quad (5.3.2)$$

The approximations in (5.3.2) are obtained by neglecting the “excess over the boundaries” of the stopping time associated with SPRT [104].

The RST for a normal mean, which is defined in terms of the stopping rule (5.2.7) with $n_0 = 1$, was first studied by McPherson [121], who provided an exact computation for the probability of Type I error $\alpha$, probability of Type II error $\beta$ and the ASN. Approximations for such quantities are also given in [122]. In Figure 5.2 the ASN for the RST and the SPRT versus the mean $\mu_1$ are plotted. Not surprisingly, the SPRT has a slightly smaller ASN purchased at the price of the full knowledge of the distribution of the observations, which is needed in order to compute the log-likelihood ratio test.
5.4 A Randomly Truncated Test

Let \( \{X_i\}_{i \geq 1} \) be a sequence of i.i.d. random variables with \( E(X_i) = \mu \) and \( Var(X_i) = \sigma^2 \) (where \( \sigma \), the nuisance parameter, is unknown). Consider a non-negative function \( g(\cdot) \geq 0 \) and a stopping time

\[
\mathcal{N} = \inf \left\{ n : \sum_{i=1}^{n} g(X_i) \geq M \right\}. \tag{5.4.1}
\]

We might consider \( g(X_i) \) to be the cost of observing \( X_i \), and \( M > 0 \) represents the total budget that one is allowed to spend on testing.

Lemma 5.4.1. If \( M \to \infty \) and \( E(g(X)) \) exists and is finite, then

\[
\frac{\mathcal{N}}{M/E(g(X))} \to 1 \quad a.s. \tag{5.4.2}
\]

where with \( E(g(X)) \) we denote the expectation of \( g(X) \).

Proof. The proof is straightforward and can be found in ([118], p. 58), where renewal theory is used. \( \text{Qed} \)

Following (5.2.7), to test

\[
H_0 : \mu = 0 \quad \text{vs} \quad H_1 : \mu \neq 0, \tag{5.4.3}
\]

we introduce the stopping time

\[
\tau = \inf \left\{ k \geq n_0 : \left| \frac{S_k}{\sqrt{B_k}} \right| \geq b \right\}, \tag{5.4.4}
\]
The repeated significance test with random truncation stops sampling at \( \min(\tau, N) \) and rejects the null hypothesis \( H_0 \) if and only if \( \tau \leq N \). Otherwise, the null hypothesis \( H_0 \) is accepted. The test proceeds as follows: for \( k \geq n_0 \)

\[
\begin{aligned}
\text{if } \sum_{i=1}^{k} g(X_i) \geq M \text{ and } |S_k| \leq b \sqrt{B_k} & \quad \text{choose } H_0, \\
\text{if } \sum_{i=1}^{k} g(X_i) < M \text{ and } |S_k| > b \sqrt{B_k} & \quad \text{choose } H_1, \\
\text{otherwise} & \quad \text{take another observation.}
\end{aligned}
\]

(5.4.5)

For all values of \( \mu \), the power function for the repeated significance test is given by:

\[
\pi(\mu) = P_{\mu}(\tau \leq N) = 1 - \beta(\mu),
\]

(5.4.6)

where

\[
\beta(\mu) = P_{\mu} \left( \max_{n_0 \leq k \leq N} \left| \frac{S_k}{\sqrt{B_k}} \right| < b \right).
\]

(5.4.7)

is the probability of type II error function \(^1\). The significance level of this test (probability of type I error) is given by:

\[
\alpha = \pi(0) = P_0(\tau \leq N) = 1 - \beta(0)
\]

\[
= P_0 \left( \max_{n_0 \leq k \leq N} \left| \frac{S_k}{\sqrt{B_k}} \right| \geq b \right)
\]

(5.4.8)

The following result is of major importance for implementing the repeated significance test with random truncation time based on the stopping time \( N \). It implies

\(^{1}\) With \( P_\mu \) and \( P_0 \) we indicate the probability under the hypothesis \( H_1, \mu \neq 0 \) and under \( H_0, \mu = 0 \), respectively.
that if \( n_0 \) and \( M \) are sufficiently large, one can safely substitute the random quantities \( N \) and \( B_n \) by their deterministic asymptotic quantities. Theorem 1 provides an asymptotic relationship between the probability of type I error and the pair \((t_0, b)\) and can be useful for setting the threshold \( b \).

**Theorem 5.4.2.** If \( N \) is a stopping time that is defined as in (5.4.1) and the expected value \( E(g(X)) \) exists and is finite, then

\[
P_0(\tau \leq N) = P_0\left( \max_{n_0 \leq k \leq N} \left| \frac{S_k}{\sqrt{B_k}} \right| \geq b \right) \to \alpha(t_0, b).
\]

(5.4.9)
as \( n_0, M \to \infty \) and \( \frac{n_0}{M/E(g(X))} \to t_0 \) with \( t_0 \in (0, 1) \).

**Proof.** First, we estimate the probability in (5.4.9) from above. Fix \( \varepsilon > 0 \). Let \( N = M/E(g(X)) \), then observe that

\[
P_0\left( \max_{n_0 \leq k \leq N} \left| \frac{S_k}{\sqrt{B_k}} \right| > b \right) \leq P_0\left( \max_{n_0 \leq k \leq (1 + \varepsilon)N} \left| \frac{S_k}{\sqrt{B_k}} \right| > b, N \leq (1 + \varepsilon)N \right) + P_0(\mathcal{N} > (1 + \varepsilon)N)
\]

\[
\leq P_0\left( \max_{n_0 \leq k \leq (1 + \varepsilon)N} \left| \frac{S_k}{\sqrt{B_k}} \right| > b \right) + P_0(\mathcal{N} > (1 + \varepsilon)N)
\]

\[
\leq P_0\left( \max_{n_0 \leq k \leq (1 + \varepsilon)N} \left| \frac{S_k}{\sqrt{\sigma^2_k B_k}} \right| > b, \sup_{k \geq n_0} \frac{B_k}{\sigma^2_k} \leq 1 + \varepsilon \right) + P_0(\mathcal{N} > (1 + \varepsilon)N)
\]

\[
\leq P_0\left( \max_{n_0 \leq k \leq (1 + \varepsilon)N} \left| \frac{S_k}{\sqrt{\sigma^2_k B_k}} \right| > b \sqrt{1 + \varepsilon} \right) + P_0(\mathcal{N} > (1 + \varepsilon)N)
\]

(5.4.10)
By Kolmogorov’s strong law of large numbers [123] we have $B_n/n \to \sigma^2$ a.s. as $n \to \infty$ which implies (for example, [124], p. 253) that

$$P_0 \left( \sup_{k \geq n_0} \left| \frac{B_k}{k} - \sigma^2 \right| > \epsilon \right) \to 0$$

(5.4.11)

for every $\epsilon > 0$. Equation (5.4.11) is equivalent to

$$P_0 \left( \sup_{k \geq n_0} \frac{B_k}{\sigma^2 k} > 1 + \epsilon \right) \to 0,$$

(5.4.12)

$$P_0 \left( \sup_{k \geq n_0} \frac{\sigma^2 k}{B_k} > 1 + \epsilon \right) \to 0,$$

(5.4.13)

as $n_0 \to \infty$. On the other hand, as consequence of Lemma 1 we have that, as $M \to \infty$

$$P_0 (N > (1 + \epsilon)N) \to 0,$$

(5.4.14)

$$P_0 (N < (1 - \epsilon)N) \to 0.$$  

(5.4.15)

By the FCLT [118] we find the following upper bound on the crossing probability in (5.4.9)

$$\lim_{n_0 \to \infty} P_0 \left( \max_{n_0 \leq k \leq N} \left| \frac{S_k}{\sqrt{B_k}} \right| > b \right) \leq \alpha \left( \frac{t_0}{1 + \epsilon}, \frac{b}{\sqrt{1 + \epsilon}} \right).$$

(5.4.16)

We now proceed to obtain similar estimates of the probability in (5.4.9) from below.

First, we estimate the probability in (5.4.9) from above. Fix $0 < \epsilon < 1$. Let
\[ N = M/E(g(X)), \] then observe that

\[ P_0 \left( \max_{n_0 \leq k \leq (1-\epsilon)N} \left| \frac{S_k}{\sqrt{\sigma^2 k}} \right| > b\sqrt{1+\epsilon} \right) \]

\[ \leq P_0 \left( \max_{n_0 \leq k \leq (1-\epsilon)N} \left| \frac{S_k}{\sqrt{\sigma^2 k}} \right| > b\sqrt{1+\epsilon}, \sup_{k \geq n_0} \frac{\sigma^2 k}{B_k} \leq 1 + \epsilon \right) + P_0 \left( \sup_{k \geq n_0} \frac{\sigma^2 k}{B_k} > 1 + \epsilon \right) \]

\[ \leq P_0 \left( \max_{n_0 \leq k \leq (1-\epsilon)N} \left| \frac{S_k}{\text{median} B_k} \right| > b, \sup_{k \geq n_0} \frac{\sigma^2 k}{B_k} \leq 1 + \epsilon \right) + P_0 \left( \sup_{k \geq n_0} \frac{\sigma^2 k}{B_k} > 1 + \epsilon \right) \]

\[ \leq P_0 \left( \max_{n_0 \leq k \leq N} \left| \frac{S_k}{\text{median} B_k} \right| > b, \mathcal{N} \geq (1-\epsilon)N \right) + P_0 \left( \mathcal{N} < (1-\epsilon)N \right) + P_0 \left( \sup_{k \geq n_0} \frac{\sigma^2 k}{B_k} > 1 + \epsilon \right) \]

\[ \leq P_0 \left( \max_{n_0 \leq k \leq N} \left| \frac{S_k}{\sqrt{B_k}} \right| > b \right) + P_0 \left( \mathcal{N} < (1-\epsilon)N \right) + P_0 \left( \sup_{k \geq n_0} \frac{\sigma^2 k}{B_k} > 1 + \epsilon \right) \]  \hspace{1cm} (5.4.18)

Thus, by the FCLT we have:

\[ \lim_{n_0 \to \infty} P_0 \left( \max_{n_0 \leq k \leq N} \left| \frac{S_k}{\sqrt{B_k}} \right| > b \right) \geq \alpha \left( \frac{t_0}{1-\epsilon}, b\sqrt{1+\epsilon} \right). \]  \hspace{1cm} (5.4.19)

Since \( \epsilon > 0 \) is arbitrary, and \( \alpha(\cdot, \cdot) \) is continuous, we have (5.4.9).

Qed

In [125] a randomly truncated RST, based on a stopping time of ratios of variance estimators, has been employed for testing the median of a heavy tailed distribution.

### 5.4.1 Evaluation of \( \alpha(t_0, b) \)

The theoretical critical values \( b \) and the corresponding targeted significance levels have been obtained in [126] where the tables of \( \alpha(t_0, b) \) for a range of values of
Given $b > 0$, $m > 0$ and the truncation instant $m$, define

$$T = \inf \{ t : t \geq m, |W(t)| \geq bt^{1/2} \} \quad (5.4.20)$$

Let consider the test of $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ which stops sampling at $\min(T, m)$ and rejects $H_0$ if and only if $T \leq m$.

**Theorem 5.4.3.** Suppose $b \to \infty$, $m \to \infty$, $m_0 \to \infty$ in such way that for some $0 < \mu_1 < \mu_0 < \infty$, $bm^{-1/2} = \mu_1$ and $bm_0^{-1/2} = \mu_0$. Then for $T$ defined by (5.4.20) as $b \to \infty$

$$P_0(T < m) = \alpha(t_0, b) = (b - b^{-1})\phi(b) \log(1/t_0) + 4b^{-1}\phi(b) + o(b^{-1}\phi(b)) \quad (5.4.21)$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ and $t_0 = m_0/m$.

**Proof.** The proof is given in [104].

An indication of the accuracy of the approximation in Theorem 5.4.3 is given in table 5.1 where also the simulated values for $\alpha(t_0, b)$, using 1,000,000 Monte Carlo runs, are given.

### 5.4.2 Power Function Calculation

In this section we construct an approximation for the power function of the repeated significance test as in (5.4.3). We show that the power function can be
Table 5.1: Approximation and exact values of $\alpha(t_0, b)$. The performance of the approximation increases as $b$ becomes larger.

<table>
<thead>
<tr>
<th>$b$</th>
<th>$t_0$</th>
<th>Approximation</th>
<th>Exact value according to [126]</th>
<th>MC Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.9</td>
<td>0.5</td>
<td>0.2026</td>
<td>0.1902</td>
<td>0.1869</td>
</tr>
<tr>
<td>2.60</td>
<td>0.1</td>
<td>0.0902</td>
<td>0.0923</td>
<td>0.0894</td>
</tr>
<tr>
<td>3.2</td>
<td>0.08</td>
<td>0.024</td>
<td>0.025</td>
<td>0.020</td>
</tr>
</tbody>
</table>

approximated by the probability of crossing square root boundaries by the standard Brownian motion with a linear drift. While a formal proof for the calculation of the power function is impossible to derive, similar arguments to those used in the proof of Theorem 1 can provide an approximation for the power function. In fact, the FCLT for $\{X_i\}_{i \geq 1}$ and the strong laws of large numbers for $\mathcal{N}$, $g_{n_0} = \sum_{i=1}^{n_0} g(X_i)/n_0$, and $B_n$, suggest the following approximation (the accuracy of which will be evaluated via simulation). Let us again set $N = M/E(g(X))$,
then if $H_1: \mu = \mu_1$ is true, as $n_0, M \to \infty$ and \( \frac{n_0}{M/E(g(X))} \to t_0 \), we have

$$
\pi(\mu_1) = P_{\mu_1}(\tau \leq N)
$$

$$
= P_{\mu_1}\left( \max_{n_0 \leq k \leq N} \left| \frac{S_k}{\sqrt{B_k}} \right| \geq b \right)
$$

$$
\approx P_{\mu_1}\left( \max_{n_0 \leq k \leq N} \left| \frac{S_k - \mu_1 k + \mu_1 k}{\sigma \sqrt{k}} \right| \geq b \right)
$$

$$
= P_{\mu_1}\left( \max_{n_0 \leq k \leq N} \left( \frac{S_k - \mu_1 k}{\sigma \sqrt{N}} + \frac{\mu_1}{\sigma} \sqrt{N} \frac{k}{N} \right) / \sqrt{\frac{k}{N}} \geq b \right)
$$

$$
\approx P_{\mu_1}\left( \max_{n_0/N \leq t \leq 1} \left( W_t + \frac{\mu_1}{\sigma} \sqrt{N} t \right) / \sqrt{t} \geq b \right)
$$

$$
= P_{\mu_1}\left( \max_{n_0Eg(X_i)/M \leq t \leq 1} \left| \frac{W_t + \mu_1}{\sigma} \frac{E\left( g(X) \right)}{E(g(X))} \cdot t}{\sqrt{t}} \right| \geq b \right) \quad (5.4.22)
$$

Thus, in the case $\mu = \mu_1$ we deal with the problem of crossing square root boundaries by the standard Brownian motion with the following linear drift:

$$
\frac{\mu_1}{\sigma} \sqrt{\frac{M}{E(g(X))}} t.
$$

Since $\sigma$ is unknown, we use instead the $n_0$-sample standard deviation estimator $s = \sqrt{B_{n_0}/(n_0 - 1)}$. Again, we will use $g_{n_0} = \sum_{i=1}^{n_0} g(X_i - \mu_1)/n_0$ instead of $E_{\mu_1}(g(X))^2$. The power function can be computed numerically by solving an appropriate associated partial differential equation (see [115], Section 3.2).

\[ ^2 \text{With } E_{\mu_1}, \text{ we denote the expectation under the alternative hypothesis } H_1. \]
5.5 Application: Decentralized Sequential Detection in Sensor Networks With Communication Constraints

5.5.1 Problem Formalization

We consider a SN that is designed to test two statistical hypotheses (say $H_1$ versus $H_0$). In our model, the conditional joint distributions of the sensor observations under each hypothesis are unknown. We also make the assumption that the sensor observations $\{X_i\}_{i \geq 1}$ are i.i.d. from sensor to sensor. For $i = 1, 2, \ldots, \infty^3$, we should also remark that the index spans the sensor class which is, in our model, unbounded. We are implicitly assuming that the number of sensors must be large enough to guarantee that the sequential test stops while the FC is scanning the SN. This working hypothesis can also be removed since the FC implements a randomly truncated sequential test. This implies that the sequential test is guaranteed to terminate.

---

3 We should also remark that the index spans the sensor class which is, in our model, unbounded.
representing the index of the sensor in network, we consider the following hypothesis test

\[ H_0 : X_i = W_i \quad \text{vs} \quad H_1 : X_i = W_i + \mu. \]

where \( W_i \) are i.i.d. realizations of a continuous symmetric random variable whose distribution is unknown. Assuming that all the sensors’ observations have the same mean \( \mu \) can be a reasonable model for cases where all the sensors measure a physical parameter (temperature, pressure, chemical composition, biological agents, etc.) which may be constant over large scales. The location parameter \( \mu \) is assumed to be unknown.

### 5.5.2 Energy Constraint by Random Truncation

Due to communication bandwidth and energy constraints, the \( i \)th sensor processes its own observation by applying a two-level quantization \( Y_i \in \{-1, 0, 1\} \); that is, it transmits data if and only if \( X_i \not\in [-\tau, \tau] \) (see Figure 5.3). Otherwise, for \( X_i \in [-\tau, \tau] \) the sensor stays silent and does not transmit. This scheme links to the concept of censoring, [89], [90], [91], as the the symbol \( Y_i = 0 \) does not expend any resource.
The data $Y_i$ delivered by the $i$th sensor are:

$$Y_i = \begin{cases} 
1 & \text{if } X_i > \tau, \\
0 & \text{if } -\tau \leq X_i \leq \tau, \\
-1 & \text{if } X_i < -\tau,
\end{cases} \quad (5.5.1)$$

Let $p_0(\tau) = P\{X_i \notin [-\tau, \tau]|H_0\}$ be the probability that the $i$th sensor’s observation $X_i$ falls outside the censoring zone, under the null hypothesis $H_0$. Let also $p_1^+(\tau) = P\{X_i > \tau|H_1\}$ and $p_1^-(\tau) = P\{X_i < -\tau|H_1\}$ be probabilities that the $i$th sensor’s observation $X_i$ is greater than $\tau$ and it is smaller than $-\tau$ under the alternative hypothesis $H_1$, respectively. Let $Y_i^j \in \{-1, 0, 1\}$ be the discrete random variable under the hypothesis $j$ (with $j = 0, 1$) which is associated to the transmission by the $i$th sensor. The corresponding unknown probability mass functions (pmf) for $Y_i^j$ are $[\frac{p_0(\tau)}{2}, 1 - p_0(\tau), \frac{p_0(\tau)}{2}]$ and $[p_1^-(\tau), 1 - (p_1^+ + p_0^-)(\tau), p_1^+(\tau)]$ for the hypothesis $H_0$ and the hypothesis $H_1$, respectively. It is straightforward to note that $Y_i$ have zero mean under the null hypothesis $H_0$. By using the data $Y_i$, the FC wants to test:

$$H_0 : \mu = 0 \quad \text{vs} \quad H_1 : \mu \neq 0$$

Whenever the $i$th sensor transmits its data we define $A_i = A$ (constant for all the sensors) the cost or resource which is paid by the $i$th sensor (see Figure 5.3). Such cost can relate to the bandwidth/energy which is necessary for the transmission and the communication between the sensor and the fusion center.
Fig. 5.4: Sensor network for decentralized sequential detection. Due to bandwidth constraints the observations have to be quantized and then transmitted to a FC which implements a truncated RST. In the unlikely case that the RST crosses the boundaries at the same sensor’s index $i$ for which $\sum_{i=1}^{n} g(x_i) \geq M$ the FC decides in favor of $H_1$.

Unlike most of the papers in literature, among which we cite [85], [86], [87], here we do not impose an *average* or *peak* constraint but instead a *point* constraint. More formally we impose the following: the cumulative expenditure of the resources $\sum_{i=1}^{n} g(X_i)$, when $n$ sensors are polled (sampled), is computed and whenever $\sum_{i=1}^{n} g(X_i) > M$ (with $M$ being the total amount of available resources that the system is allowed to spend) the FC stops sequentially sampling the sensors. The FC implements the randomly truncated sequential test which was previously
introduced in Section 5.4; see the conceptual scheme depicted in Figure 5.4.

5.5.3 Minimum Amount of System Resources $M$

In this subsection we are interested in finding the minimum amount of system resources that the SN is willing to spend in order to achieve the desired performance in terms of probability of Type I error $\alpha$ and of Type II error $\beta$.

The RST implemented by the FC is subject to the following requirements:

- the initial sample size $n_0$ is given (the FC polled the first $n_0$ sensors and the system did not reach the energy constraint yet, that is $\sum_{i=1}^{n_0} g(X_i) < M$;

- the probability of Type I error, the significance level $\alpha$, is given;

- the probability of Type II error $\beta$ at $H_1: \mu = \mu_1$ must be less than a given level $\beta^*$;

- the cost function $g(\cdot)$ is given;

- the total resources $M$ budgeted ought to be made as small as possible.

For a given significance level $\alpha$ we have a curve of feasible pairs $(t_0, b)$ that is defined by equation

$$\alpha(t_0, b) = \alpha. \tag{5.5.2}$$

Note that $t_0 = \frac{n_0}{M/Eg(X_i)}$. Since $n_0$ is fixed and $Eg(X_i)$ can be substituted by $g_{n_0} = \sum_{i=1}^{n_0} g(X_i)/n_0$, equation (5.5.2) gives, in fact, a curve of feasible pairs
Fig. 5.5: In \((M,b)\) plane the curve above is a plot of an increasing function. Moreover, it easy to check that in the \((M,b)\)-plane this curve is an increasing function (see Fig. 5.5). In order to solve for \(b\) and \(M\) we need equation (5.4.22) which computes the power of the test.

**Example 5.5.1.** Let \(\{X_i\}_{i \geq 1}\) be a sequence i.i.d. Gaussian random variables with mean \(E(X_i) = \mu\) and variance \(Var(X_i) = \sigma^2\). The system parameters are given in the Table 5.2.

After performing a numerical optimization, we found that the minimum amount of resources \(M_{\text{min}}\) that the must be budgeted by the SN in order to achieve the preassigned detection performances is \(M_{\text{min}} = 2508\) (see Table 5.3). In Fig. 5.6
the simulated probability of Type II error $\beta^{sim}$ as function of $M$ is plotted.

<table>
<thead>
<tr>
<th>Table 5.2: Optimization Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5.3: Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b^*$</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>2.85</td>
</tr>
</tbody>
</table>

As we can see, the value $M_{min} = 2508$, which has numerically been found using the theory of the randomly truncated RST, is a good approximation of the value $M_{min}^{sim} = 2490$ which has been computed by simulation.

### 5.5.4 Optimal Censoring Threshold $\tau$

Here we want to find the optimal censoring threshold $\tau$ which maximizes the power of the test subject to the constraints that the significance level $\alpha(t_0, b) = \alpha$ and that the resources of the SN are confined to $M$. The requirements for the RST implemented by the FC are:

- the initial sample size $n_0$ is given (the FC polled the first $n_0$ sensors and the system did not reach the energy constraint yet, that is $\sum_{i=1}^{n_0} g(X_i) < M$;
Fig. 5.6: Simulated probability of Type II error $\beta^{\text{sim}}$ as function of $M$ is plotted.

Simulations are based on 10,000 runs.

- the probability of Type I error, the significance level $\alpha$, is given;

- the total resources $M$ is budgeted.

- the cost function $g(\cdot)$, parametrized only by the threshold $\tau$ (the parameter $A$ is fixed in advance), should be optimized in order to maximize the power of the test $\pi(\mu)$ for a specific alternative $\mu = \mu_1$.

Example 5.5.2. Let $\{X_i\}_{i\geq 1}$ be a sequence i.i.d. Gaussian random variables with mean $E(X_i) = \mu$ and variance $Var(X_i) = \sigma^2$. The system parameters are given in the Table 5.4.
Our goal is to maximize the power of the test $\pi$ (minimize $\beta$) with respect to the threshold $b$ of the RST and the censoring threshold $\tau$ subject to the constraint that the significance level $\alpha(t_0(\tau), b) = \alpha$.

**Table 5.4:** Optimization Parameters 10,000 MonteCarlo simulations

<table>
<thead>
<tr>
<th>$A$</th>
<th>$M$</th>
<th>$\mu_0$</th>
<th>$\mu_1$</th>
<th>$\sigma$</th>
<th>$n_0$</th>
<th>$\alpha^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2000</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1000</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>2000</td>
<td>0.05</td>
<td>1</td>
<td>1000</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

**Table 5.5:** Results

<table>
<thead>
<tr>
<th>$b_{opt}$</th>
<th>$\tau_{opt}$</th>
<th>$\alpha(b_{opt}, \tau_{opt})$</th>
<th>$\pi(b_{opt}, \tau_{opt})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.633</td>
<td>0.1867</td>
<td>0.05</td>
<td>1</td>
</tr>
<tr>
<td>2.9133</td>
<td>1.8773</td>
<td>0.05</td>
<td>0.9880</td>
</tr>
</tbody>
</table>

We study two cases:

1. $\mu_1 >> \mu_0$ (the two hypotheses are highly distinguishable)

2. $\mu_1 \rightarrow \mu_0$ (The two hypotheses are close to each other)

In the Tables 5.5 the optimal values for $b$ and $t$ are given both for the first ($\mu_1 >> \mu_0$) and second cases ($\mu_1 \rightarrow \mu_0$). As we can see from Figure (5.7), the optimal censoring threshold $\tau_{opt}$ is small, while for the case $\mu_1 \rightarrow \mu_0$ it is large (see Figure
Fig. 5.7: Censoring function when (a) $\mu_1 >> \mu_0$: highly distinguishable hypotheses.

(5.8). A nice interpretation of these results is the following: in a constrained framework (the system can spend at most $M$), when the sensors are not sure about the hypothesis, they do not transmit. In this way they do not waste resources and the current sum of all the resources spent by the previous sensors does not increase.

5.6 Application: Sequential detection with observations at random times

Here we apply the test proposed in Section 5.4 to a sequential detection problem with observations at random times and a maximum-time constraint. A standard approach in sequential detection is that the observations are available at the Central Unit (CU) either immediately or with a constant and known delay.
However, in certain new applications [127], [128], [129], observations arrive at the CU through communication channels with considerable unknown delays. In this section we model the process of these observations as a one-dimensional Poisson process with arrival intensity $\lambda$ [81], see Figure 5.9. Let $\{X_i\}_{i \geq 1}$ be a sequence of independent i.i.d. Gaussian random variables with mean $E(X_i) = \mu$ and variance $Var(X_i) = \sigma^2$, which is generally unknown. $\{X_i\}_{i \geq 1}$ represents a set of observations which are collected at random times $t_i$. The FC center wants to test $H_0 : \mu_0 = 0$ against the composite alternative $H_1 : \mu_1 \neq 0$ using a truncated RST with the following stopping time:

$$
\mathcal{N} = \inf \{n : \sum_{i=1}^{n} Y_i \geq T\}. \quad (5.6.1)
$$

where the $Y_i$’s denote the interarrival times ($Y_i = t_i - t_{i-1}$ with $Y_1 = t_1$) which are governed by an exponential distribution with parameter $\lambda$ [81]. The constraint in (5.6.1) implies that the decision has to be made within a maximum-time $T$. Using
Fig. 5.9: Arrival process of the observations is described by a one-dimensional Poisson process. The interarrival times are independent and obey the exponential distribution with parameter $\lambda$.

Theorem 5.4.2 and the power approximation as in (5.4.22) we can characterize the probability of type I error $\alpha$ and the power of the test as follows:

$$P_0(\tau \leq N) = P_0\left(\max_{n_0 \leq k \leq N} \left| \frac{S_k}{\sqrt{B_k}} \right| \geq b\right) = \alpha(t_0, b) \quad (5.6.2)$$

with $t_0 = \frac{n_0}{\lambda T}$ and

$$\pi(\mu_1) = P_{\mu_1}(\tau \leq N) \approx P_{\mu_1}\left(\max_{n_0/\lambda T \leq t \leq 1} \left| \frac{W_t + \frac{\mu_1}{\sigma} \sqrt{\lambda T} \cdot t}{\sqrt{t}} \right| \geq b\right). \quad (5.6.3)$$

In Figure 5.10 the probability of type I error and the power of the test obtained by simulation and by theory, respectively, are plotted as function of the parameter $T$. As we can see the parameter $T$ could be designed to achieve some desired performance in terms of probability of type I error $\alpha$ and the power of the test $\pi(\mu_1)$ at a specific alternative $\mu_1$. For example, one has $\alpha = 0.07$ and $\pi(\mu_1)|_{\mu_1=0.3} = 0.9$. 
Fig. 5.10: The probability of type I error and the power of the test obtained by simulation and by theory, respectively, are plotted as function of $T$. Here, we used, $\lambda = 7$, $n_0 = 40$, $\sigma = 1$, $b = 2.5$ and $\mu_1 = 0.3$. Simulations are based on 10,000 runs.

for $T = 950$. One could jointly choose the RST threshold $b$ together with the parameter $T$.

5.7 Conclusions

In this chapter we have introduced a new randomly truncated RST based on a random stopping time associated with an application in decentralized sequential detection problem with communication constraints in a sensor network. In certain applications, two discussed in this article, the truncation time is based on a ran-
dom stopping time associated with total available resources. We have developed an effective method for the design and analysis of the proposed randomly truncated sequential test, including accurate calculations of the probability of type I error and the power function. We also studied two applications of the proposed test: a decentralized detection problem in sensor networks with communication constraints with the FC performing a randomly truncated test and a sequential detection problem with measurements at random times.
Bibliography


