Multi-Sensor Detection Schemes Using Spatio-Temporal Models

by

Sadiq Ali

Advisors: José A. López-Salcedo and Gonzalo Seco-Granados

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Department of Telecommunication and Systems Engineering
School of Engineering

February 2014

Multi-Sensor Detection Schemes Using Spatio-Temporal Models

Author: Sadiq Ali

Thesis Advisor: José A. López-Salcedo

Thesis Advisor: Gonzalo Seco-Granados

Abstract

This thesis is focused on multi-sensor event detection schemes in Wireless Sensor Network (WSN). In the WSN, sensors individually take their observations, do some initial local processing and then send the result to the fusion center. Apart from this information, the fusion center may have prior information such as the known positions of sensors, the topology of the network or structures, features and patterns present in the received observations. In order to achieve better detection performance, the fusion center must exploit all this prior information in the best possible way.

Keeping this in mind, in this thesis we focus on the exploitation of available a-priori information with the aim of developing enhanced and robust detection mechanisms. For instance, we exploit the fact that in most applications, the signal emitted from the event becomes a local physical phenomenon that only affects a small subset of sensors. Moreover, the affected sensors will be located close to the event as well as close to each other in the form of a spatial cluster. Hence, novel detection schemes are presented with a two-fold motivation: first, the exploitation of the relevant set of sensors, which helps in rejecting the noise; second, to take advantage of the signal correlation by using a-priori information about the positions of sensors.

Based on these results, we move one step further and concentrate on the exploitation of both spatial and temporal correlation in the received observations. In this case, we propose detection schemes that explore different matrix structures embedded in the observed covariance matrix, which naturally arise due to the underlying topologies of the multiple sensors.

Numerical results are presented for all of the proposed schemes that show important advantages compared to traditional schemes.

Contents

List of Figures

List of Tables

List of Abbreviations

WSN	Wireless Sensor Network
LRT	Likelihood Ratio Test
ROC	Receiver Operating Characteristic
AUC	Area Under ROC Curve
SNR	Signal to Noise Ratio
GLRT	Generalize Likelihood Ratio Test
MLE	Maximum Likelihood Estimate
MFLRT	Multi-family Likelihood Ratio Test
AIC	Akaike information criterion
MDL	Minimum description length
CAV	Non-destructive Analysis
SPKP	Single Pair Kronecker Product
MPKP	Multi-Pair Kronecker Product
dB	Decibel
CUSUM	Cumulative Sum
MCUSUM	Multivariate Cumulative Sum
KLD	Kullback-Leibler Divergence
CRN	cognitive radio network
CR	Cognitive Radio
PU	Primary user or licensed user
SU	Secondary User or unlicensed user
CLT	Central Limit Theorem
ST-GLRT	Spatio-Temporal GLRT
P-GLRT	Persymeteric based Spatio-Temporal GLRT
KR-GLRT	Kronecker product based Spatio-Temporal GLRT
PK-GLRT	Persymeteric and Kronecker product based Spatio-Temporal GLRT
SIMO	Single Input Multiple Output.
i.i.d.	independent and identically distributed.
PDF	Probability density function.
i.e.	in effect.
e.g.	example given.
Fig.	Figure

List of Symbols

K	Total number of sensors.
$\mathcal{K}$	Subset of total sensors.
L	Number of antennas.
$\mathcal{H}_0$	Null hypothesis.
$\mathcal{H}_1$	Alternate hypothesis.
$\mathcal{R}$	Bayes risk
$\Lambda(\mathbf{X})$	Likelihood ratio.
$\mathcal{L}(\mathbf{x})$	Log-likelihood ratio.
log	Natural logarithm.
γ	Threshold.
P_{FA}	Probability of false alarm.
P_D	Probability of miss-detection.
P_M	Kronecker product.
λ	Lagrange multiplier.
$\|.\|$	Represents both determinant and absolute value.
$\mathbf{H}$	Signature matrix .
E[.]	Expected value/expectation.
$\boldsymbol{\mu}_{\mathcal{H}_i}$	Statistical mean under $\mathcal{H}_i$.
$\mathcal{N}(\mu, \sigma^2)$	Gaussian Distribution with mean μ and variance σ^2.
σ^2	Noise variance/power.
T_Δ	Average Detection delay time .
$\rho_{i,j}$	Correlation between two points i, j.
sup	Denotes Supremum.
$esssup$	Denotes essential supremum.
χ^2	Chi-square distribution.
$\mathbf{J}$	Denotes the reversal matrix.
$\otimes$	Kronecker product .
$\mathcal{D}\left(f_{\mathcal{H}_1} \parallel f_{\mathcal{H}_0}\right)$	Kullback-Leibler Divergence(KLD).
$\triangleq$	$A \triangleq B$ means A is equal by definition to B.

$\sum$	Summation.
$\prod$	Product.
∂	Partial derivative.
$\in$	is an element of.
$\approx$	is approximately equal to.
$\mathbb{C}$	Complex numbers.
$\mathbb{R}$	Real numbers.
$(.)^T$	Transpose.
$(.)^H$	Conjugate transpose.
π	Pi.
$\bar{x}$	(often read as "x bar") is the mean (average value of x_i).
$\sqrt{.}$	Square root.
$\int$	Integration.
$\|.\|_F^2$	Frobenious norm.
Tr $(\mathbf{A})$	Trace of $\mathbf{A}$.
$[\mathbf{A}]_{i,j}$	The $(i-\text{th}, j-\text{th})$ element of matrix $\mathbf{A}$.
$\mathbf{A}^{-1}$	Inverse of a full-rank matrix $\mathbf{A}$.
E $[.]$	Statistical expectation.
exp(x)	Exponential function. That is, e^x.
$=$	Equal to.
$\neq$	Not equal to.
$\hat{\mathbf{a}}$	Estimate of vector $\mathbf{a}$.
vec$(\mathbf{A})$	Vec operator or column-wise stacking of matrix $\mathbf{A}$.
vech$(\mathbf{A})$	Vech operator or column-wise stacking of matrix $\mathbf{A}$ by removing the terms above the main diagonal of $\mathbf{A}$.
$a < b$	a smaller than b.
$a > b$	a greater than b.
$a \ll b$	a much smaller than b.
$a \gg b$	a much greater than b.
$a \leq b$	a smaller or equal to b.
$a \geq b$	a greater or equal to b.

Dedication

I dedicate this thesis to my parents Mr. Maqsood Ali and Mrs. Noor, and my late uncle Mr. Mahboob Ali (1945-2004).

Acknowledgments

For me completing my Ph.D. degree is probably the most challenging activity of my life. The best and worst moments of my doctoral journey have been shared with many people. It has been a great privilege to spend several years at Signal Processing for COMmunication and NAVigation (SPCOMNAV) in the Department of Telecommunication and Systems Engineering, Engineering School of Universitat Autònoma de Barcelona (UAB) and its members will always remain dear to me.

First and foremost I want to thank my advisor José Antonio López-Salcedo for his invaluable and insightful guidance throughout the course of my Ph.D. study at the UAB. It has been a pleasure to be his Ph.D. student. I appreciate all his contributions of time and ideas to make my Ph.D. experience productive and stimulating. The joy and enthusiasm he has for his research was contagious and motivational for me, even during tough times in the Ph.D. pursuit. I am always thankful for his support and all I have learned from his great professional vision and scientific insight. This would not be possible without him. Secondly, I am thankful to Prof. Gonzalo Seco-Granados for his timely and valuable support being my co-advisor during my Ph.D. studies.

Besides my advisors, my sincere thanks also goes to Prof. Magnus Jansson for offering me the opportunity in the Signal Processing Lab, School of Electrical Engineering, KTH-Royal Institute of Technology, Sweden and leading me working on diverse exciting aspects of signal processing..

Sadiq Ali

UAB, Bellatera
February 2014

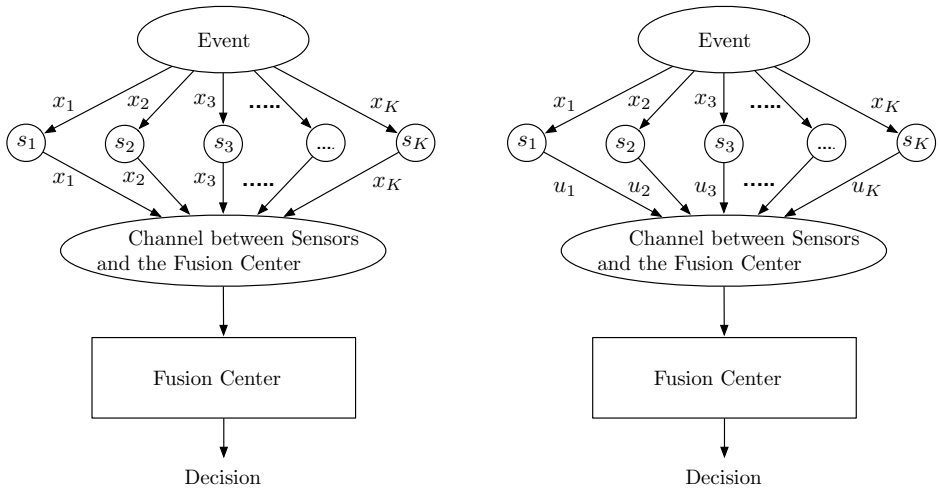
CHAPTER 1

Introduction

Event detection is one of the main purposes of most Wireless Sensor Network (WSN) applications [1]. For example WSNs for military applications are deployed to detect the invasion of enemy forces, health monitoring sensor networks are deployed to detect abnormal patient behavior, fire detection sensor networks are deployed to set an alarm if a fire starts somewhere in the monitored area. Similarly, in the field of cognitive radio, event detection involves the collaborative detection of signals from licensed users, in order to detect empty gaps at which unlicensed users can transmit[2, 3]. In this application the unlicensed users can be considered as the sensors. Regardless of the specific application, the goal is to alert a centralized unit about the occurrence of significant events so that it can respond appropriately to them. During the detection process in WSN, after taking their local observations and doing some initial local processing, the sensors send the information to the fusion center. The fusion center manages and fuses this information to make the final decision.

Managing multiple sensors for signal detection purposes is a challenging task for which a wide range of different strategies can be devised. For instance, *centralized detection* (as shown in Fig. 1.1a) is the most straightforward and simple approach. It involves placing all the intelligence of the network on a single point, where all sensor measurements will be processed. Based on these received data, the fusion center will issue a global decision on whether the event of interest was present or not. Certainly, this simple approach raises many practical problems in terms of network reliability and bandwidth constraints. A step forward for solving these drawbacks leads to the concept of *decentralized* detection as depicted in Fig. 1.1b. In this approach, sensors do perform some kind of pre-processing for data compression or even single-decision making, and then they report their results to the fusion center [4], [5],[6]. Resource constraints can be alleviated by using this approach, but there is still a critical dependence on the fusion center. Since the fusion center based detection scheme can largely resort to classical detection theory, many works focus on distributed detection [7]. However, it is matter of fact that some issues are still open in centralized detection, and thus they demand attention from researchers.

Keeping the above facts into consideration, in this thesis we focus on the development of fusion rules based on the soft information provided by randomly distributed sensors. Specifically, we consider the fact that the fusion rule should exploit all of the available a-priori known side information about the sensors, which often depends on the topology of a specific



(a) Centralized detection scheme

(b) Decentralized detection scheme

Figure 1.1: Schematic representation of detection schemes with fusion center in WSN

network. Consequently, the fusion rule should take into account all features, structures and patterns etc, that are inherently present in the received observations due to the underlying topology of the WSN.

1.1 Motivation and objectives

In some application where the local sensors receive strong and reliable signals, the sensors can readily make their local decisions and send them to the fusion center where the final decision is made. The fusion of local decisions at the fusion center is termed as *hard fusion* [8]. In other words, the hard decision fusion is the one in which, the local sensors make a one-bit decision regarding the existence of the event, then this one-bit decision (0 or 1) is forwarded to the fusion center for the *global decision*. The most common hard decision rules are "OR", "AND " and "Counting" rules as given in Fig. 1.2. Nevertheless, in many applications, the sensors may do some local processing of the observations instead of hard decisions, and then they send their processed information to the fusion center for the final decision. The combination of such soft information at the fusion center is known as the *soft fusion*. The scheme based on the soft information fusion has applications in situations where the local sensors receive week signals in the presence of harsh conditions (shadowing, low signal-to-noise ratio etc) and local decisions are unreliable. It is demonstrated in [9] that soft combination schemes have significant performance improvement over the hard combination in harsh working conditions. The drawback of soft data combination is due to the fact that when each sensor transmits the real value of its sensing data to the fusion center, theoretically infinite bits are required. Hence, this results in a wide communication bandwidth. However, this problem can be solved by quantization of local observations [9]. Nevertheless, when high performance detection is required, it is better that instead of making one bit local decisions,

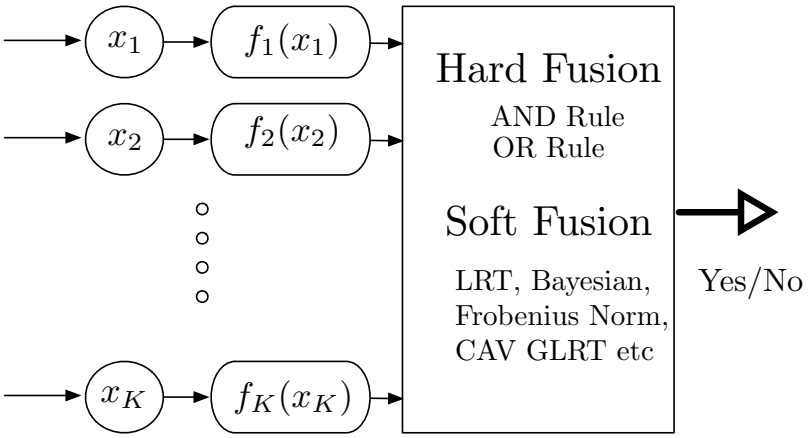


Figure 1.2: Summary of multi-sensor fusion rules.

the sensors send quantized soft measurements to the fusion center. In the presence of this soft information, the fusion center can easily utilize all of the available information and exploit the side information for improved detection performance.

Keeping this in mind, we consider the fact that in order to bring decision about the presence or absence of any event, the decision mechanism should process and exploit all of the available information in the best possible way. Therefore, the fusion rule should not only use the energy and power of the signal but also the available a-priori information. A-priori information can be either the positions of the sensor nodes or the statistical topology of the network in terms of the sensors, spatial and temporal correlations. Hence, **the main focus of this work is to develop detection mechanisms at the fusion center that take into account the side information to enhance the robustness**.

Before starting with the technical contents of this thesis, we will briefly discuss some of the existing traditional detection schemes that will be used as a reference herein. The most popular detection schemes that have been previously considered include the energy detection, the matched filtering and cyclostationary detection [10, 11, 12, 13]. These algorithms have different requirements and advantages/disadvantages. The detection performance of these detection schemes depends on the available prior knowledge. For example if the signal is fixed and a-priori known to the receiver, matched filtering gives the optimum detection performance. If the signal is unknown then the most popular and widely used scheme is the energy based detection scheme. Unlike other methods, energy detection does not need any information of the signal to be detected and is robust to unknown multipath fading [10, 11]. Moreover, if the signal is unknown, signal samples are often modeled as independent and identically distributed (i.i.d.) random variables, and it that case the energy detector gives the optimum performance. However, even though the energy detector doesn't require the desired signal to be known, it assumes the noise to be known. Hence, energy detection is vulnerable to the noise uncertainty, because the method requires the exact knowledge of the noise power [11, 14, 15, 16]. In practice, it is very difficult to obtain accurate noise power estimates. Therefore, one should look for mechanisms that take into account the fact that

noise is unknown in practice. However, in many cases especially in practical communication systems, received signals to sensors always contain distinctive features that can be exploited for detection and that enable us to achieve a detection performance that substantially surpasses that of the energy detector. Perhaps even more importantly, known signal features can be exploited to estimate unknown parameters such as the noise power. Therefore**, making use of known signal features or a-priori information can effectively circumvent the problem faced by traditional energy based detection mechanisms.**

Herein, the term feature detection refers to the exploitation of the known statistical properties of the signal [13]. Depending on the application, the signal features referred to may be manifested both in time and space. For instance, one of the features that can be exploited as a side information is the correlation of the received observations. For example, dense deployment of sensor nodes makes the observations highly correlated in the space domain [17, 18]. The existence of spatial correlation implies that the readings from sensor nodes which are geographically close to each other are expected to be largely correlated. In addition to being spatially correlated, the received signal samples are usually correlated in time due to the presence of temporal correlated channels, due to oversampling at the receiver or because the transmitted signal was correlated in time[19, 20]. However, for the sake of simplicity, many of the previous contributions ignore the correlation in the formulation of detection schemes. On the other hand, many of the schemes consider the correlation as a deleterious effect [21, 22]. Hence, these schemes focus on techniques to counteract (i.e. whitening or decorrelation) correlation effects. However, the correlation can be considered as a prior information or side information that can be used to avoid the pitfalls of the noise uncertainty or to bring robustness to the detection process.

Beside the exploitation of spatio-temporal correlation, one should also take into account the fact that for the case of dense sensor networks, the signal of emitted by the event of interest will only effect a subset of sensors (i.e. those sensors located close to the event), which will typically be closely spaced, forming a cluster, thus having highly correlated observations [17]. This scenario is depicted in the upper picture of Fig. 1.3. Based on intuition, we can see that since active sensors are in the form of clusters, their physically closeness can provide us extra-information and this should be considered in the formulation of the fusion rule.

Similarly, a particular case of interest is the one where multiple sensors are uniformly and closely placed to each other in a single device (i.e. multi-antenna transceiver) and it will be also interested to exploit a-priori information for such scenario. In addition, one should also consider a WSN that consists of multiple distributed nodes where each node has multiple antennas. Such setup can easily be considered in the field of cognitive radios since multi-antenna receivers have already become an integral part of many cognitive radios. The existence of such hardware setup/platform gives us the chance to consider multi-antenna techniques for cooperative spectrum sensing. Hence, for this type of setup it will be interesting to devise novel detection schemes that exploit the correlation structure of the received observation for improved detection performance.

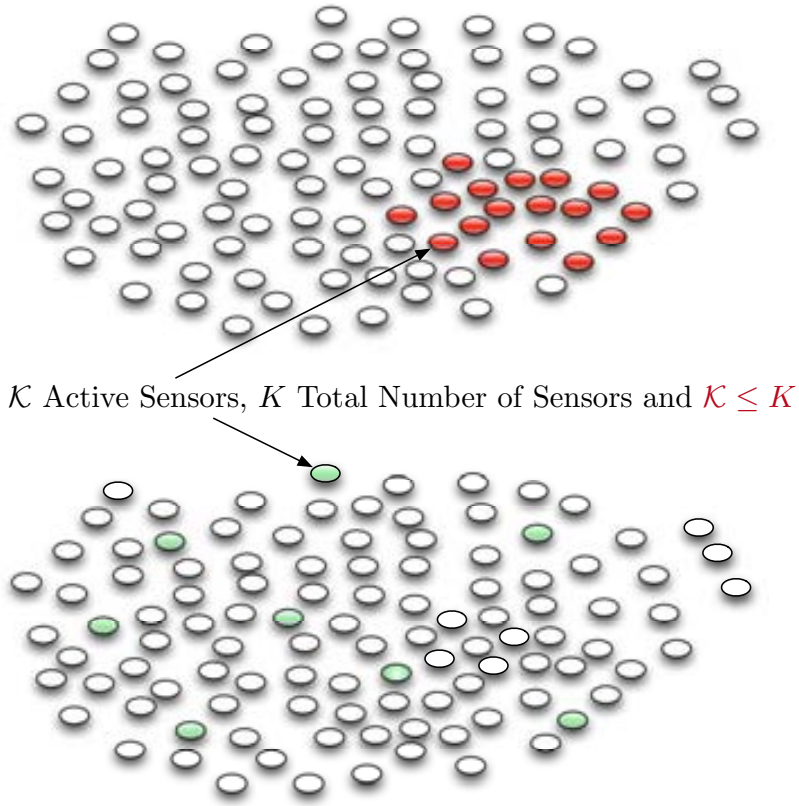


Figure 1.3: Pictorial view of the event region.

1.2 Thesis outline and research contributions

Based on the above motivating discussions, the main contributions are organized as follow:

Chapter 2

In this chapter, we introduce the reader to the basics of hypothesis testing and we provide a brief review of the fundamentals of the signal detection theory. These fundamentals are presented with the aim of of facilitating the discussion for the reader in the remaining chapters.

Chapter 3

Based on the fundamental principles of statistical signal processing, in this chapter we focus on the centralized detection scheme, where the local sensors send soft information to the fusion center. Since the centralized detection can largely resort to classical detection theory, many existing results are already available. However, there are still some problems left for centralized detection, especially in the case of composite hypothesis. We present detailed discussion on multi-sensor composite hypothesis schemes. Moreover, we consider the fact that the region where the event happens within a dense WSN usually spans across an area

which includes just a subset of all the sensors. In order to achieve detection schemes with improved SNR, we propose mechanism to select the most relevant sensors and ignore rest of the sensors in decisions making. The performance of the detection schemes is evaluated by means of numerical simulations, showing important advantages with respect to the traditional schemes.

Partial results of this chapter have been presented in the following paper:

- S. Ali, J. A. Lopez-Salcedo, G. Seco-Granados, "Sensor-to-Sensor Assistance for Distributed Signal Detection", *Proc. IEEE ICASSP, Dallas (USA), Mar 18 2010.*

Chapter 4

In this chapter we exploit the spatial correlation information based on the spatial proximity of sensors with each other. In order to use this information we focus on two different detection mechanisms. The first scheme is based on the use of the Generalized Likelihood Ratio Test (GLRT) and model order selection techniques, leading to improved event detectors by incorporating the spatial correlation already present in measurements coming from neighboring sensors through a novel concept of *signature vectors*. For the resulting detector, analytical and simulation results are provided, showing a significant gain in performance compared to traditional approaches.

In the second part of this chapter, we present a novel quickest detection scheme to sequentially detect the emergence of an event with the help of multiple sensors. In the proposed scheme, we exploit the fact that the observations of the spatially proximal sensor nodes are highly correlated due to correlated shadowing effects. However, we will see that when it comes to infer the spatial correlation, the estimate of the spatially structured covariance matrix is not feasible, since the proposed quickest detection scheme is a recursive method and operates with single snapshot. Hence, we propose to model the spatial covariance structure by using the a-priori information about the locations of the sensors. Moreover, the proposed scheme also takes into account a scenario where only a subset of sensors are affected by the event's signal instead of the whole sensor field. Therefore, it takes into account both the exploitation of the spatial structure and the selection of the subset of sensors in the process of detection. Both analytical and numerical results are developed for the mean detection delay, showing important advantages.

The results in this chapter have been published in the following two papers:

- S. Ali, G. Seco-Granados, J. A. Lopez-Salcedo, "Multi-sensor Quickest Detection by Exploiting Radio Correlation", *Proc. European Signal Processing Conference (EUSIPCO), Sep 09 2013.*
- S. Ali, J. A. Lopez-Salcedo, G. Seco-Granados, "Improved GLRT Based on the Exploitation of Spatial Correlation Between Neighboring Sensors", *Proc. 19th European Signal Processing Conference (EUSIPCO), Aug 31 2011.*

Chapter 5

In this chapter, we adopt the concept of exploiting spatial correlation proposed in Chapter 4, to apply it to the system of cognitive radios that collaborate with each other with the aim of detecting the random waveforms being emitted from licensed users. In order to do so, we adopt and extend the proposed scheme in the previous chapter by considering the fact that the signal is stochastic rather deterministic signal as it was in the previous chapter. We derive a cognitive detector (for collaborative spectrum sensing) based on the GLRT and the use of spatial signatures. Adopting model order detection allows the detector at the base station to operate with a rank-reduced version of the observed covariance matrix. Since the estimation of this matrix may be a challenge in large-scale networks, we study the application of shrinkage techniques to cope with the problem of having more sensors than available observations. For the proposed detector, numerical results are drawn, showing a significant gain in performance compared to traditional approaches.

The results in this chapter are summarized in one journal paper:

- S. Ali, G. Seco-Granados, J. A. Lopez-Salcedo, "Spectrum Sensing with Spatial Signatures in the Presence of Noise Uncertainty and Shadowing", *EURASIP Journal on Wireless Communications and Networking, vol 2013, Jun 03 2013.*

Chapter 6

In Chapters 4 & 5, we exploited spatial proximity of sensors as a side information to design robust multi-sensor detection schemes. Herein, we focus on exploiting the underlying spatial and temporal features/patterns present in the received covariance matrix. Having this in mind, first we propose a novel mechanism for spectrum sensing that leads us to exploit the spatio-temporal correlation present in the received signal at a multi-antenna receiver. We also extend and apply the proposed mechanisms to event detection in WSN. Towards the end of this chapter, we present novel spectrum sensing mechanisms for a cognitive radio network with multiple distributed radios where each radio has multiple antennas. The proposed spectrum sensing scheme exploits the fact that when any primary signal is present, measurements are spatially correlated due to inter-antennas and inter-radios spatial correlation. In order to exploit this correlation, we propose novel detection schemes that exploit the embedded inter-antennas and inter-radios structure in the received observations.

For all proposed mechanisms, we formulate the spectrum sensing scheme by adopting the GLRT paradigm. Unfortunately, the GLRT degenerates in the of case in the case of high dimensionality and small sample size. To circumvent this problem, then several extensions are proposed that bring robustness to the GLRT in the limited sample support regime by exploiting covariance structure and factoring the large spatio-temporal covariance matrix. The performance of the proposed detectors is evaluated by means of numerical simulations, showing important advantages.

The results in this chapter have been published in two conference papers and are currently under review in one journal paper:

- S. Ali, M. Jansson, G. Seco-Granados, J. A. Lopez-Salcedo, "Kronecker-based Cooperative Spectrum Sensing with Multi-Antenna Receivers", *submitted to Elsevier Signal Processing Feb. 2014.*
- S. Ali, M. Jansson, G. Seco-Granados, J. A. Lopez-Salcedo, "Novel Collaborative Spectrum Sensing Based on Spatial Covariance Structure", *Proc. European Signal Processing Conference (EUSIPCO), Sep 09 2013.*
- S. Ali, J. A. Lopez-Salcedo, G. Seco-Granados, "Exploiting structure of spatio-temporal correlation for detection in wireless sensor networks", *Proc. 20th European Signal Processing Conference (EUSIPCO), Aug 2012.*

Chapter 7

We conclude and present the possible extension of the present work.

CHAPTER 2

Elements of Detection Theory

In order to facilitate the discussions in the remaining chapters, in this chapter we discuss some important concepts based on the classical signal detection theory. Applications of signal detection theory are found in many areas, such as communications, event detection and automatic control. For example, in communications applications detection and estimation provide the theoretical and analytical basis for the design of effective communication receivers. In general terms, the detection applications involve making inferences from observations that are distorted or corrupted by the presence of some disturbing effect (e.g. noise).

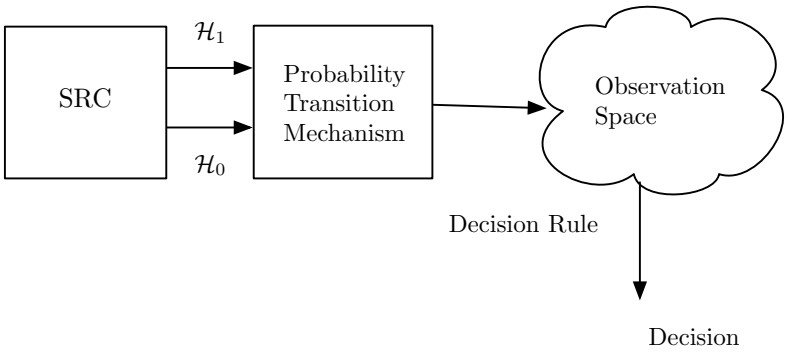


Figure 2.1: Components of a detection theory problem

The basic components of detection theory are shown in Fig. 2.1. The first component is the *source* that generates an event. In the simplest case this output is one of two choices. We refer to them as hypotheses and label them $\mathcal{H}_0$ and $\mathcal{H}_1$. We can explain this with the help of a simple example, i.e., signal detection in the presence of noise. In a typical signal detection process the noise is the unwanted energy that interferes with the ability of the detector (sensor) to detect the wanted signal (output of the source). Assume that for some physical measurement a sensor produces an output observation x. The observation may have been produced by the noise alone or by an impinging wanted or desired signal corrupted by noise. These two possibilities are called the null hypothesis $\mathcal{H}_0$ and the alternative hypothesis $\mathcal{H}_1$, respectively, and are commonly written in the compact notation :

$$\begin{aligned} \mathcal{H}_0 &: x = \text{noise alone}, \\ \mathcal{H}_1 &: x = \text{signal} + \text{noise}. \end{aligned} \tag{2.1}$$

The second component of the problem is a *probabilistic transition mechanism* and the third is an *observation space* [23]. The transition mechanism can be viewed as a device that knows which hypothesis is true. Based on this knowledge, it generates a point in the observation space according to some probability law. After observing the outcome in the observation space we shall guess which hypothesis was true between the null $(\mathcal{H}_0)$ and alternative $(\mathcal{H}_1)$ hypotheses . To decide between the null and alternative hypotheses one might apply some threshold in such a way that if the signal crosses this threshold, it means that we have $\mathcal{H}_1$ true. Now we are faced with the practical question of where to set the threshold so as to ensure that the number of decision errors is small. Note that there are two types of error possible, P_M, the error of missing the signal (decide $\mathcal{H}_0$ while $\mathcal{H}_1$ was actually true) and P_{FA}, the error of false alarm (decide $\mathcal{H}_1$ while $\mathcal{H}_0$ was true). Considering these errors, there is always a compromise between choosing a high threshold to make the average number of false alarms small versus choosing a low threshold to make the average number of misses small. To quantify this compromise it becomes necessary to specify the statistical distribution of x under each of the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ and a robust *decision rule.* The latter is, therefore, the fourth and most important component of any detection problem. In the following we demonstrate how these four components of the detection theory fit together to form the total decision or hypothesis-testing problem.

2.1 Simple binary hypothesis tests

Hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are called simple hypotheses when the statistical distributions of the observations under $\mathcal{H}_0$ and $\mathcal{H}_1$ are completely known. That is to say, when these distributions involve no unknown parameters such as signal amplitude, signal phase, or noise power. In binary hypothesis testing, we consider the decision problem in which each of two source outputs corresponds to a hypothesis. We assume that the observation space corresponds to a set of K observations: $x_1, x_2, x_3, ..., x_K$. Thus each set can be thought of as a point in an Kdimensional space and can be denoted by a vector $\mathbf{x}$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \end{bmatrix}. \tag{2.2}$$

Based on the received $\mathbf{x}$, the probabilistic transition mechanism generates points in accordance with the two known probability densities $f_{\mathbf{x}|\mathcal{H}_1}(\mathbf{X} \mid \mathcal{H}_1)$ and $f_{\mathbf{x}|\mathcal{H}_0}(\mathbf{X} \mid \mathcal{H}_0)$. The objective is to use these known probability densities to develop a suitable decision rule. Thus each time the experiment is conducted one of four things can happen [23] :

1. $\mathcal{H}_0$ true; choose $\mathcal{H}_0$.
2. $\mathcal{H}_0$ true; choose $\mathcal{H}_1$.

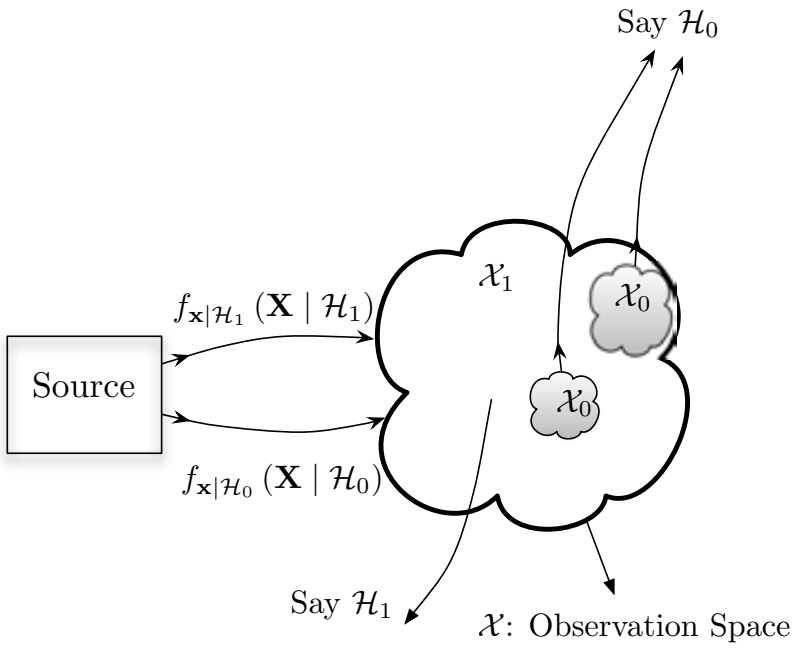


Figure 2.2: Observation space and decision regions

3. $\mathcal{H}_1$ true; choose $\mathcal{H}_1$
4. $\mathcal{H}_1$ true; choose $\mathcal{H}_0$

The first and third alternatives correspond to correct choices. The second and fourth alternatives correspond to errors. The purpose of a decision criterion is to attach some relative importance to the four possible courses of action. It might be expected that the method of processing the received data ($\mathbf{x}$) would depend on the decision criterion we select. In the following two subsections we discuss the two most important criterion, the Bayes and the Neyman-Pearson. We show that for these criteria of special interest, the operations on observation $\mathbf{x}$ are identical.

2.1.1 Bayes criterion

The Bayes test is based on two assumptions. The first is that the source outputs are governed by probability assignments, which are denoted by p_1 and p_0, respectively, and called the a-priori probabilities for hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$. These probabilities represent the observer's a-priori known information about the source before the experiment is conducted. The second assumption is that a cost is assigned to each possible course of action. We denote the cost for the four courses of action as $C_{00}, C_{10}, C_{11}, C_{01}$, respectively. The first subscript indicates the hypothesis chosen and the second, the hypothesis that was true. Each time the experiment is conducted a certain cost will be incurred. The aim will be to design our decision rule in such a way that on the average the cost will be as small as possible. To do this we first write an expression for the expected value of the cost. We see that there are two probabilities that

we must average over; the a priori probability and the probability that a particular course of action will be taken. Denoting the expected value of the cost as the risk $\mathcal{R}$, we have:

$$\begin{aligned}\mathcal{R} = & C_{00}p_0\text{Pr}\left(\text{say } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ is true }\right) \\ & + C_{10}p_0\text{Pr}\left(\text{say } \mathcal{H}_1 \mid \mathcal{H}_0 \text{ is true }\right) \\ & + C_{11}p_0\text{Pr}\left(\text{say } \mathcal{H}_1 \mid \mathcal{H}_1 \text{ is true }\right) \\ & + C_{01}p_0\text{Pr}\left(\text{say } \mathcal{H}_0 \mid \mathcal{H}_1 \text{ is true }\right),\end{aligned} \tag{2.3}$$

where Pr $(.)$ denotes probability of an event. Because we have assumed that the decision rule must say either $\mathcal{H}_0$ or $\mathcal{H}_1$, we can view it as a rule for dividing the total observation space $\mathcal{X}$ into two parts, $\mathcal{X}_0$ and $\mathcal{X}_1$, as shown in Fig. 2.2 . Whenever an observation falls in $\mathcal{X}_0$ we say $\mathcal{H}_0$, and whenever an observation falls in $\mathcal{X}_1$ we say $\mathcal{H}_1$. We can now write the the expression for the risk in terms of the transition probabilities and decision regions :

$$\begin{aligned}\mathcal{R} = & C_{00}p_0\int_{\mathcal{X}_0} f_{\mathbf{x}|\mathcal{H}_0}\left(\mathbf{X} \mid \mathcal{H}_0\right)d\mathbf{X} \\ & + C_{10}p_0\int_{\mathcal{X}_1} f_{\mathbf{x}|\mathcal{H}_0}\left(\mathbf{X} \mid \mathcal{H}_0\right)d\mathbf{X} \\ & + C_{11}p_1\int_{\mathcal{X}_1} f_{\mathbf{x}|\mathcal{H}_1}\left(\mathbf{X} \mid \mathcal{H}_1\right)d\mathbf{X} \\ & + C_{01}p_1\int_{\mathcal{X}_0} f_{\mathbf{x}|\mathcal{H}_1}\left(\mathbf{X} \mid \mathcal{H}_1\right)d\mathbf{X}.\end{aligned} \tag{2.4}$$

It is to be noted that for a K-dimensional observation space the integrals in (2.4) are K-fold integrals. No doubt that in practice the biggest issue is ending up with the wrong decision (i.e. miss detection and false alarm), hence, we shall easily assume throughout this chapter that the cost of a wrong decision is higher than the cost of a correct decision. In other words,

$$\begin{aligned}C_{10} > C_{00}, \\ C_{01} > C_{11.}\end{aligned} \tag{2.5}$$

Now, to find the Bayes test we must choose the decision regions $\mathcal{X}_0$ and $\mathcal{X}_1$ in such a manner that the risk will be minimized. Because we require that a decision be made, this means that we must assign each point $\mathbf{X}$ in the observation space $\mathcal{X}$ to either $\mathcal{X}_0$ or $\mathcal{X}_1$. Thus

$$\mathcal{X} = \mathcal{X}_0 + \mathcal{X}_1 \triangleq \mathcal{X}_0 \cup \mathcal{X}_1. \tag{2.6}$$

Based on the above facts, we can rewrite (2.4) as:

$$\begin{aligned}\mathcal{R} = & p_0C_{00}\int_{\mathcal{X}_0} f_{\mathbf{x}|\mathcal{H}_0}\left(\mathbf{X} \mid \mathcal{H}_0\right)d\mathbf{X} + p_0C_{10}\int_{\mathcal{X}-\mathcal{X}_0} f_{\mathbf{x}|\mathcal{H}_0}\left(\mathbf{X} \mid \mathcal{H}_0\right)d\mathbf{X} \\ & + p_1C_{11}\int_{\mathcal{X}_1} f_{\mathbf{x}|\mathcal{H}_1}\left(\mathbf{X} \mid \mathcal{H}_1\right)d\mathbf{X} + p_1C_{01}\int_{\mathcal{X}-\mathcal{X}_0} f_{\mathbf{x}|\mathcal{H}_1}\left(\mathbf{X} \mid \mathcal{H}_1\right)d\mathbf{X}.\end{aligned} \tag{2.7}$$

From (2.7), we can observe that

$$\int_{\mathcal{X}} f_{\mathbf{x}|\mathcal{H}_0}\left(\mathbf{X} \mid \mathcal{H}_0\right) d\mathbf{X} = \int_{\mathcal{X}} f_{\mathbf{x}|\mathcal{H}_1}\left(\mathbf{X} \mid \mathcal{H}_1\right) d\mathbf{X} = 1, \tag{2.8}$$

then (2.7) will reduce to

$$\begin{aligned}\mathcal{R} &= p_0 C_{00} + p_1 C_{11} \\ &+ \int_{\mathcal{X}_0} \left[\left\{p_1 \left(C_{01} - C_{11}\right) f_{\mathbf{x}|\mathcal{H}_1}\left(\mathbf{X} \mid \mathcal{H}_1\right)\right\} - \left\{p_0 \left(C_{10} - C_{00}\right) f_{\mathbf{x}|\mathcal{H}_0}\left(\mathbf{X} \mid \mathcal{H}_0\right)\right\}\right] d\mathbf{X}.\end{aligned} \tag{2.9}$$

In (2.9), the first two terms represent the fixed cost. The term with integral represents the cost controlled by those points $\mathbf{X}$ that we assign to $\mathcal{X}_0$. The assumption in (2.6) implies that the two terms inside the brackets are positive [23]. Therefore, all values of $\mathbf{X}$ where the second term is larger than the first should be included in $\mathcal{X}_0$ because they contribute a negative amount to the integral. Similarly, all values of $\mathbf{X}$ where the first term is larger than the second should be excluded from $\mathcal{X}_0$ (assigned to $\mathcal{X}_1$) because they would contribute a positive amount to the integral. Values of $\mathbf{X}$ where the two terms are equal have no effect on the cost and may be assigned arbitrarily. We shall assume that these points are assigned to $\mathcal{H}_1$ and ignore them in our subsequent discussion. Thus the decision regions are defined by the statement:

If

$$p_1 \left(C_{01} - C_{11}\right) f_{\mathbf{x}|\mathcal{H}_1}\left(\mathbf{X} \mid \mathcal{H}_1\right) \geqq p_0 \left(C_{10} - C_{00}\right) f_{\mathbf{x}|\mathcal{H}_0}\left(\mathbf{X} \mid \mathcal{H}_0\right), \tag{2.10}$$

assign $\mathbf{X}$ to $\mathcal{X}_1$ and consequently say that $\mathcal{H}_1$ is true. Otherwise assign $\mathbf{X}$ to $\mathcal{X}_0$ and consequently say that $\mathcal{H}_0$ is true. Alternatively, we may write

$$\frac{f_{\mathbf{x}|\mathcal{H}_1}\left(\mathbf{X} \mid \mathcal{H}_1\right)}{f_{\mathbf{x}|\mathcal{H}_0}\left(\mathbf{X} \mid \mathcal{H}_0\right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \frac{p_0 \left(C_{10} - C_{00}\right)}{p_1 \left(C_{01} - C_{11}\right)}. \tag{2.11}$$

The quantity on the left is called the *likelihood ratio* and denoted by $\Lambda\left(\mathbf{X}\right)$

$$\Lambda\left(\mathbf{X}\right) \triangleq \frac{f_{\mathbf{x}|\mathcal{H}_1}\left(\mathbf{X} \mid \mathcal{H}_1\right)}{f_{\mathbf{x}|\mathcal{H}_0}\left(\mathbf{X} \mid \mathcal{H}_0\right)}. \tag{2.12}$$

Because it is the ratio of two functions of a random variable, it is also a random variable. We see that regardless of the dimensionality of $\mathbf{X}$, $\Lambda\left(\mathbf{X}\right)$ is a one-dimensional variable. Similarly, the quantity on the right of (2.11) is the threshold of the test and is denoted by γ_B:

$$\gamma_B \triangleq \frac{p_0 \left(C_{10} - C_{00}\right)}{p_1 \left(C_{01} - C_{11}\right)}. \tag{2.13}$$

Thus Bayes criterion leads us to a *likelihood ratio test* (LRT)

$$\Lambda\left(\mathbf{X}\right) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma_B \tag{2.14}$$

We see that all the data processing is involved in computing $\Lambda(\mathbf{X})$ in (2.14) and is not affected by a priori probabilities or cost assignments. This invariance of the data processing is of considerable practical importance. Frequently the costs and a priori probabilities are merely educated guesses. The result in (2.14) enables us to build the entire processor and leave γ_B as a variable threshold to accommodate changes in our estimates of a priori probabilities and costs. As the decision to be drawn depends only on whether the $\Lambda(\mathbf{X})$ exceeds the threshold or not, any monotone increasing [1] operation can be performed on both sides of (2.14) without affecting the values of observed data $\mathbf{x}$ that cause the threshold to be exceeded, and therefore without affecting P_D (choose $\mathcal{H}_1$ and $\mathcal{H}_1$ was true or choose $\mathcal{H}_0$ and $\mathcal{H}_0$ was true) and P_{FA} (choose $\mathcal{H}_1$ and $\mathcal{H}_0$ was true). Knowing this one should choose some sort of transformation to simplify the final expression $\Lambda(\mathbf{X})$. Most common is to take the natural logarithm of both sides of (2.14) to obtain the log-likelihood ratio test:

$$\begin{aligned} \log\Lambda(\mathbf{X}) &\gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \log\gamma_B, \\ &\gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma, \end{aligned} \tag{2.15}$$

where $\gamma = \log\gamma_B$.

Before continuing our discussion of likelihood ratio tests in general and then composite hypothesis testing we shall discuss a second criterion and prove that it also leads to a likelihood ratio test.

2.1.2 Neyman-Pearson criterion

In many physical situations it is difficult to assign realistic costs or a priori probabilities. A simple procedure to by-pass this difficulty is to concentrate on the problem of maximizing P_D for fixed P_{FA}, where we have

$$P_D = \int_{\mathcal{X}_1} f_{\mathbf{x}|\mathcal{H}_1}(\mathbf{X} \mid \mathcal{H}_1)\, d\mathbf{x}, \tag{2.16}$$

$$P_{FA} = \int_{\mathcal{X}_1} f_{\mathbf{x}|\mathcal{H}_0}(\mathbf{X} \mid \mathcal{H}_0)\, d\mathbf{x}. \tag{2.17}$$

In general, we would like to make P_{FA}, as small as possible and P_D as large as possible. However, for most problems of practical importance these are conflicting objectives. An obvious criterion is to constrain one of the probabilities and maximize (or minimize) the other. A specific statement of this criterion is called Nyeman-Pearson Criterion and can be define as [23]:

Neyman-Pearson criterion: Constrain $P_{FA} = \alpha_{NP} \leq \alpha$ and design a test to maximize P_D under this constraint.

This optimization problem can be solved by the method of Lagrange multipliers. In order

[1] A monotone decreasing operation would simply invert the sense of the threshold test.

to do so we can construct the function [23]

$$\mathcal{F}_{\mathcal{L}} = P_D + \lambda \left(P_{FA} - \alpha_{NP} \right). \tag{2.18}$$

Now to find the optimum solution, we need to maximize $\mathcal{F}_{\mathcal{L}}$ and then choose λ to satisfy the constraint criterion $P_{FA} = \alpha_{NP}$. Substituting (2.16) and (2.17) into (2.18), we get

$$\begin{aligned} \mathcal{F}_{\mathcal{L}} &= \int_{\mathcal{X}_1} f_{\mathbf{x}|\mathcal{H}_1}\left(\mathbf{X} \mid \mathcal{H}_1\right) d\mathbf{x} + \lambda \left(\int_{\mathcal{X}_1} f_{\mathbf{x}|\mathcal{H}_0}\left(\mathbf{X} \mid \mathcal{H}_0\right) d\mathbf{x} - \alpha_{NP} \right) \\ &= -\lambda \alpha_{NP} + \int_{\mathcal{X}_1} \left\{ f_{\mathbf{x}|\mathcal{H}_1}\left(\mathbf{X} \mid \mathcal{H}_1\right) + \lambda f_{\mathbf{x}|\mathcal{H}_0}\left(\mathbf{X} \mid \mathcal{H}_0\right) \right\} d\mathbf{x} \end{aligned}. \tag{2.19}$$

Remember that the design variable here is the choice of the region $\mathcal{X}_1$. The first term in the second line of (2.19) does not depend on $\mathcal{X}_1$, so $\mathcal{F}_{\mathcal{L}}$ is maximized by maximizing the value of the integral over $\mathcal{X}_1$. Since λ could be negative, the integrand can be either positive or negative, depending on the values of λ and the relative values of $f_{\mathbf{x}|\mathcal{H}_0}\left(\mathbf{X} \mid \mathcal{H}_0\right)$ and $f_{\mathbf{x}|\mathcal{H}_1}\left(\mathbf{X} \mid \mathcal{H}_1\right)$. The integral is therefore maximized by including in $\mathcal{X}_1$ all the points, and only the points, in the K-dimensional space for which $f_{\mathbf{x}|\mathcal{H}_1}\left(\mathbf{X} \mid \mathcal{H}_1\right) + \lambda f_{\mathbf{x}|\mathcal{H}_0}\left(\mathbf{X} \mid \mathcal{H}_0\right) > 0$, that is, $\mathcal{X}_1$ is all points $\mathbf{x}$ for which $f_{\mathbf{x}|\mathcal{H}_1}\left(\mathbf{X} \mid \mathcal{H}_1\right) > -\lambda f_{\mathbf{x}|\mathcal{H}_0}\left(\mathbf{X} \mid \mathcal{H}_0\right)$ [24]. This leads directly to the decision rule

$$\Lambda\left(\mathbf{X}\right) = \frac{f_{\mathbf{x}|\mathcal{H}_1}\left(\mathbf{X} \mid \mathcal{H}_1\right)}{f_{\mathbf{x}|\mathcal{H}_0}\left(\mathbf{X} \mid \mathcal{H}_0\right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma_\lambda, \tag{2.20}$$

where we used $\gamma_\lambda \triangleq -\lambda$. We term (2.20) as likelihood ratio test and we can see that the left hand side of (2.20) is similar to (2.14). The detection rule in (2.20) states that the ratio of the two PDFs, each evaluated for the particular observed data $\mathbf{x}$, should be compared to a threshold. If that "likelihood ratio" exceeds the threshold, choose hypothesis $\mathcal{H}_1$, i.e., declare a event to be present. If it does not exceed the threshold, choose $\mathcal{H}_0$ and declare that a event is not present. It is to be noted that under the Neyman-Pearson optimization criterion, the P_{FA} cannot exceed the original design value and that $f_{\mathbf{x}|\mathcal{H}_1}\left(\mathbf{X} \mid \mathcal{H}_1\right)$ and $f_{\mathbf{x}|\mathcal{H}_0}\left(\mathbf{X} \mid \mathcal{H}_0\right)$ should be known beforehand.

We see that similar to the Bayes criterion, all the data processing is involved in computing $\Lambda\left(\mathbf{X}\right)$ in (2.20). Similarly, as the natural logarithm is a monotonic function, and both sides of (2.20) are positive, an equivalent test is

$$\begin{aligned} \log\Lambda\left(\mathbf{X}\right) &\underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \log\gamma_\lambda \\ &\underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma \end{aligned} \tag{2.21}$$

where we used $\gamma = \log\gamma_\lambda$.

Hence, we conclude that the LRT test arises as the solution to the hypothesis testing problem under both the Bayes Criterion and Neyman-Pearson criterion, whereas the difference is just the threshold where to compare the LRT [24].

2.1.3 General multivariate Gaussian

In the following chapters we will see that the focus of this thesis is to develop detection mechanisms based on observations provided by multiple sensors. Hence, in the above we have provided a general multivariate likelihood ratio rule that will facilitate most of our discussion in the following chapters. We mentioned that to use the LRT, one should exactly know the models of $f_{\mathbf{x}|\mathcal{H}_1}(\mathbf{X} \mid \mathcal{H}_1)$ and $f_{\mathbf{x}|\mathcal{H}_0}(\mathbf{X} \mid \mathcal{H}_0)$. In most detection problems, the statistical models of the observations are Gaussian, hence, as an example, we particularize the LRT introduced before to the specific case of Gaussian measurements. We consider that under the hypothesis $\mathcal{H}_i$ for $i = \{0, 1\}$ we assume that $\mathbf{x}$ is a Gaussian random vector, which is completely specified by its mean vector and covariance matrix. We denote the mean vector as:

$$\mathrm{E}[\mathbf{x} \mid \mathcal{H}_i] = \begin{bmatrix} \mathrm{E}[x_1 \mid \mathcal{H}_i] \\ \mathrm{E}[x_2 \mid \mathcal{H}_i] \\ \vdots \\ \mathrm{E}[x_K \mid \mathcal{H}_i] \end{bmatrix} \triangleq \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_K \end{bmatrix} \triangleq \boldsymbol{\mu}_{\mathcal{H}_i}. \tag{2.22}$$

Similarly, the covariance matrix under the hypothesis $\mathcal{H}_i$ is denoted as:

$$\boldsymbol{\Sigma}_{\mathcal{H}_i} \triangleq \mathrm{E}\left[(\mathbf{x} - \boldsymbol{\mu}_{\mathcal{H}_i})(\mathbf{x} - \boldsymbol{\mu}_{\mathcal{H}_i})^H \mid \mathcal{H}_i\right]. \tag{2.23}$$

We define the inverse of $\boldsymbol{\Sigma}_{\mathcal{H}_i}$as:

$$\mathbf{Q}_i = \boldsymbol{\Sigma}_{\mathcal{H}_i}^{-1}. \tag{2.24}$$

Using the above definitions we may write the probability density functions of $\mathbf{x}$ on hypothesis $\mathcal{H}_1$ as:

$$f_{\mathbf{x}|\mathcal{H}_1}(\mathbf{X} \mid \mathcal{H}_1) = \left[(2\pi)^{K/2} |\boldsymbol{\Sigma}_{\mathcal{H}_1}|^{1/2}\right]^{-1} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{\mathcal{H}_1})^H \mathbf{Q}_1 (\mathbf{x} - \boldsymbol{\mu}_{\mathcal{H}_{1i}})\right]. \tag{2.25}$$

Going through a similar set of definitions for $\mathcal{H}_0$, we obtain the probability density

$$f_{\mathbf{x}|\mathcal{H}_0}(\mathbf{X} \mid \mathcal{H}_0) = \left[(2\pi)^{K/2} |\boldsymbol{\Sigma}_{\mathcal{H}_0}|^{1/2}\right]^{-1} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{\mathcal{H}_0})^H \mathbf{Q}_0 (\mathbf{x} - \boldsymbol{\mu}_{\mathcal{H}_0})\right]. \tag{2.26}$$

Using the definition in (2.20), the likelihood ratio test follows easily:

$$\begin{aligned} \Lambda(\mathbf{X}) &= \frac{f_{\mathbf{x}|\mathcal{H}_1}(\mathbf{X} \mid \mathcal{H}_1)}{f_{\mathbf{x}|\mathcal{H}_0}(\mathbf{X} \mid \mathcal{H}_0)} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma_{initial} \\ &= \frac{\left[(2\pi)^{K/2} |\boldsymbol{\Sigma}_{\mathcal{H}_1}|^{1/2}\right]^{-1} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{\mathcal{H}_1})^H \mathbf{Q}_1 (\mathbf{x} - \boldsymbol{\mu}_{\mathcal{H}_{1i}})\right]}{\left[(2\pi)^{K/2} |\boldsymbol{\Sigma}_{\mathcal{H}_0}|^{1/2}\right]^{-1} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{\mathcal{H}_0})^H \mathbf{Q}_0 (\mathbf{x} - \boldsymbol{\mu}_{\mathcal{H}_0})\right]} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma_{initial}. \end{aligned} \tag{2.27}$$

Taking logarithms, we obtain

$$\begin{aligned}\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}_{\mathcal{H}_0}\right)^H \mathbf{Q}_0 \left(\mathbf{x}-\boldsymbol{\mu}_{\mathcal{H}_0}\right) - \frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}_{\mathcal{H}_1}\right)^H \mathbf{Q}_1 \left(\mathbf{x}-\boldsymbol{\mu}_{\mathcal{H}_{1i}}\right) \\ =\gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \log\gamma_{initial} + \frac{1}{2}\log\left|\boldsymbol{\Sigma}_{\mathcal{H}_1}\right| - \frac{1}{2}\log\left|\boldsymbol{\Sigma}_{\mathcal{H}_0}\right| \triangleq \gamma\end{aligned} \tag{2.28}$$

We see that the test consists of finding the difference between two quadratic forms. In many cases of interest, the specific form of the log-likelihood ratio can be further rearranged to isolate only those terms on the left hand side of the equation that explicitly depend on the data samples $\mathbf{x}$ and moving all other constants to the right hand side. Equation (2.28) is such a rearrangement of (2.27).

2.2 Composite binary hypothesis testing

In many real world problems, it is difficult to precisely know the probability distributions of the received observations under $\mathcal{H}_0$ and $\mathcal{H}_1$. Quite often, our models may involve unknown parameters and then the hypothesis testing is called *composite* hypothesis testing. Hypothesis testing problems discussed in the preceding sections are simple hypothesis-testing problems because each of the two hypotheses were completely known. In that case an optimal detector exists with the highest possible power P_D for any fixed level P_{FA}. This optimal detector is called the *most powerful* (MP) test and is specified by the ubiquitous likelihood ratio test [25]. In the more common case where the signal and/or noise are described by unknown parameters, at least one hypothesis is composite, and the resulting detectors have different performances for different values of the parameters. Unfortunately, there seldom exists a uniformly most powerful detector whose performances remain upper bounds for the entire range of unknown parameters. Therefore, for composite hypotheses, other design strategies beyond the Neyman-Pearson criterion or Bayes rule must generally be adopted to ensure reliable detection performance.

In the case of composite hypothesis testing there are two major approaches. The first one is to consider the unknown parameters as realizations of random variables and to assign a prior PDF. The second is to estimate the unknown parameters for use in a likelihood ratio test. The first method is termed as Bayesian approach and the second, is the GLRT. The Bayesian approach employs the Bayesian philosophy. It requires prior knowledge of the unknown parameters whereas the GLRT does not.

Hence, in the composite hypotheses testing, the general problem is to decide between $\mathcal{H}_0$ and $\mathcal{H}_1$when the PDFs $f_{\mathbf{x}|\mathcal{H}_1}\left(\mathbf{X} \mid \mathcal{H}_1\right)$ and $f_{\mathbf{x}|\mathcal{H}_0}\left(\mathbf{X} \mid \mathcal{H}_0\right)$ depend on a set of unknown parameters, which we will stack in a vector and refer to as $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$, respectively. In the following, we first discus the Bayesian approach followed by the GLRT. under $\mathcal{H}_0$, we assume that the vector parameter

2.2.1 Bayesian Approach

The Bayesian approach assigns prior PDFs to $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$. In doing so it models the unknown parameters as realizations of vector random variable. If the prior PDFs are denoted by $f(\boldsymbol{\theta}_0)$ and $f(\boldsymbol{\theta}_1)$, respectively, the PDF of the data are

$$f_{\mathbf{x}|\mathcal{H}_0}(\mathbf{X} \mid \mathcal{H}_0) = \int f_{\mathbf{x}|\mathcal{H}_0}(\mathbf{X} \mid \boldsymbol{\theta}_0; \mathcal{H}_0) f(\boldsymbol{\theta}_0) d\boldsymbol{\theta}_0, \tag{2.29}$$

$$f_{\mathbf{x}|\mathcal{H}_1}(\mathbf{X} \mid \mathcal{H}_1) = \int f_{\mathbf{x}|\mathcal{H}_1}(\mathbf{X} \mid \boldsymbol{\theta}_1; \mathcal{H}_1) f(\boldsymbol{\theta}_1) d\boldsymbol{\theta}_1, \tag{2.30}$$

where $f_{\mathbf{x}|\mathcal{H}_0}(\mathbf{X} \mid \boldsymbol{\theta}_0; \mathcal{H}_0)$ and $f_{\mathbf{x}|\mathcal{H}_1}(\mathbf{X} \mid \boldsymbol{\theta}_1; \mathcal{H}_1)$are the conditional PDFs of $\mathbf{x}$, conditioned on $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$, assuming $\mathcal{H}_0$ and $\mathcal{H}_1$ are true, respectively. The unconditional PDFs $f_{\mathbf{x}|\mathcal{H}_1}(\mathbf{X} \mid \mathcal{H}_1)$ and $f_{\mathbf{x}|\mathcal{H}_0}(\mathbf{X} \mid \mathcal{H}_0)$are now completely specified, no longer dependent on the unknown parameters. With the Bayesian approach the optimal likelihood ratio detector decides in favor of $\mathcal{H}_1$if

$$\Lambda(\mathbf{X}) = \frac{f_{\mathbf{x}|\mathcal{H}_1}(\mathbf{X} \mid \mathcal{H}_1)}{f_{\mathbf{x}|\mathcal{H}_0}(\mathbf{X} \mid \mathcal{H}_0)} = \frac{\int f_{\mathbf{x}|\mathcal{H}_1}(\mathbf{X} \mid \boldsymbol{\theta}_1; \mathcal{H}_1) f(\boldsymbol{\theta}_1) d\boldsymbol{\theta}_1}{\int f_{\mathbf{x}|\mathcal{H}_0}(\mathbf{X} \mid \boldsymbol{\theta}_0; \mathcal{H}_0) f(\boldsymbol{\theta}_0) d\boldsymbol{\theta}_0} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma_{\text{Bay}}. \tag{2.31}$$

In (2.31) the required integrations are multidimensional with dimension equal to the unknown parameter dimension, and this makes difficult to end up with a closed-form solution.

2.2.2 GLRT

The GLRT replaces the unknown parameters by their maximum likelihood estimates. Although there is no optimality associated with the GLRT, in practice it appears to work quite well and some asymptotic optimality has been reported [25]. In general, the GLRT decides in favor of $\mathcal{H}_1$if

$$\begin{aligned} \Lambda_G(\mathbf{X}) &= \frac{\max_{\boldsymbol{\theta}_1} f_{\mathbf{x}|\mathcal{H}_1}(\mathbf{X} \mid \boldsymbol{\theta}_1; \mathcal{H}_1)}{\max_{\boldsymbol{\theta}_0} f_{\mathbf{x}|\mathcal{H}_0}(\mathbf{X} \mid \boldsymbol{\theta}_0; \mathcal{H}_0)} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma_{\text{G}}, \\ &= \frac{f_{\mathbf{x}|\mathcal{H}_1}\left(\mathbf{X}; \hat{\boldsymbol{\theta}}_1, \mathcal{H}_1\right)}{f_{\mathbf{x}|\mathcal{H}_0}\left(\mathbf{X}; \hat{\boldsymbol{\theta}}_0, \mathcal{H}_0\right)} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma_{\text{G}}, \end{aligned} \tag{2.32}$$

where $\hat{\boldsymbol{\theta}}_1$ is the maximum likelihood estimate of $\boldsymbol{\theta}_1$assuming $\mathcal{H}_1$as:

$$\hat{\boldsymbol{\theta}}_1 = \arg\max f_{\mathbf{x}|\mathcal{H}_1}(\mathbf{X} \mid \boldsymbol{\theta}_1; \mathcal{H}_1), \tag{2.33}$$

and $\hat{\boldsymbol{\theta}}_0$ is the maximum likelihood estimate of $\boldsymbol{\theta}_0$assuming $\mathcal{H}_0$as:

$$\hat{\boldsymbol{\theta}}_0 = \arg\max f_{\mathbf{x}|\mathcal{H}_0}(\mathbf{X} \mid \boldsymbol{\theta}_0; \mathcal{H}_0). \tag{2.34}$$

Thus, the GLRT is like the *clairvoyant* detector except that instead of knowing the actual parameter values, it estimates them from the measurement $\mathbf{x}$, using maximum likelihood

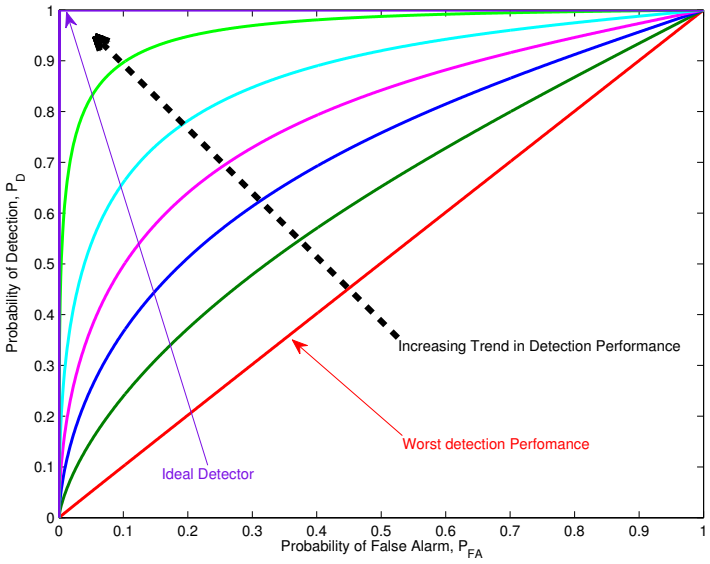


Figure 2.3: Typical Receiver Operating Characteristic (ROC) curves

[26]. Note that a true clairvoyant detector assumes perfect knowledge of an unknown parameter [25, Chap 5-6].

2.3 Performance analysis techniques

The performance of detection schemes are typically analyzed in terms of probability of detection P_D and probability of false alarm P_{FA}. In simple detection schemes, these two probabilities often have closed form expression, and thus it is easy to develop explicit expressions for these two detection metrics. However, for multi-sensor and composite detection schemes no closed form expressions for the P_D and P_{FA} are usually available. In that case numerical techniques are often required to analyze the performance of the detection schemes. The most popular approach is based on the receiver operating characteristics (ROC) and some metrics derived from this tool. We will briefly introduce these metric in the following subsections.

2.3.1 Receiver Operating Characteristics curve

A receiver operating characteristics (ROC) is a two-dimensional plot of the accuracy of a specific signal detector. In particular, it is a technique for visualizing, organizing and selecting detection schemes based on their performance. It has long been used in signal detection theory to depict the tradeoff between correct detections P_D and false alarm rates P_{FA} of a detector. In other words, the plot of the pair P_{FA} and P_D over the range of thresholds $-\infty < \gamma < \infty$ produces the ROC curve which completely describes the error rate of the detector as a function of γ [27][28].

This accuracy information apparent in the shape of the curve is two-dimensional because there are two kinds of events, and hence two kinds of accuracies possible. The first dimension is P_D of signal events, which is shown along the vertical axis. The second dimension is the error rate of falsely identifying noise (i.e, P_{FA}) , which is shown along the horizontal axis as shown in Fig. 2.3. Since success is good and error is bad, an ideal ROC curve will have vertical values which grow at a faster rate than its horizontal values, resulting in a curve shape which rises swiftly upward. As the decision threshold changes to become more and more lenient, the error values for noise (horizontal values) must also grow large, catching up with the success values for signals (horizontal values). This makes the curve bend over to the right, until it touches the point (0,1). To be more specific, a good detectors have ROC curves which have desirable properties such as concavity (negative curvature), monotone increase in P_D as P_{FA} increases, high slope of P_D at the point $(P_{FA}, P_D) = (0, 0)$, etc [27][28].

2.3.2 Area under the ROC curve

Although the ROC curves fully characterize the performance, it is also desirable to have a single and quantitative figure of merit in order to compare different detectors. The most general measure to characterize the performance is the area under the curve (AUC). The AUC summarizes the total accuracy of the detector in a way that accounts for both the gains in P_D and the losses in P_{FA}. The number for AUC always ranges from 0.5 to 1.0, because the very worst ROC curve (due to chance alone) lies along the positive diagonal and has the corresponding area of 0.5. The very best ROC curve, passing through the "northwest" corner point $(0, 1)$, has an area of 1, or the unit square. So it varies between 0.5 (poor performance) and 1 (good performance).

Next we need to know how to calculate AUC. To find the AUC, simply calculate the trapezoidal area under each vertical slice of an empirical ROC curve having a straight line segment as its top; then sum all the individual areas. For instance, if a binary detector produces only a single data point, then there will be two trapezoidal regions total for which to calculate area (one to the left of the point, and one to the right of the point). The process for AUC is mathematically expressed as:

$$\text{AUC}\left(T\right) = \int_0^1 P_D\left(T\right) dP_{FA}\left(T\right), \tag{2.35}$$

where T indicates specific detector, P_D indicates the probability of detection and P_{FA} is the probability of false alarm. To evaluate the performance curves numerically, we use the algorithm proposed in [27], which calculates the area under curve on the basis of integrating the areas of small trapezoidal bins. That is to say,

$$\text{AUC}\left(T\right) = \sum_{\kappa} \bar{P}_D^{\kappa}\left(T\right) \Delta P_{FA}^{\kappa}\left(T\right), \tag{2.36}$$

where $\Delta P_{FA}^{\kappa}\left(T\right) = P_{FA}^{\kappa+1}\left(T\right) - P_{FA}^{\kappa}\left(T\right)$ and $\bar{P}_D^{\kappa}\left(T\right) = \frac{P_D^{\kappa+1}(T)+P_D^{\kappa}(T)}{2}$.

CHAPTER 3

Multi-sensor Composite Hypothesis Testing

Once the fundamentals of signal detection have been introduced, it is time to move one step further towards the purpose of this thesis, which is mainly related to signal detection in WSN. To do so we will consider in this chapter the problem of multi-sensor hypothesis testing, which is actually the scenario we find in a WSN composed of several sensors, and where a single decision on their observed measurement has to be made. Due to noise and other impairments, the output of a single sensor may be ambiguous and misleading in some cases, but this uncertainty can be reduced using the contributions from multiple sensors. Keeping this in mind, in this chapter we propose centralized multi-sensor detection schemes that are based on the principles of the fundamental signal detection theory and model order selection techniques [29, 30].

Since, the centralized multi-sensor detection can largely resort to classical detection theory, one can find many previous contributions in the existing literature. However, there are still some problems left for centralized detection, especially in the case of composite hypothesis where the lack of information on the sensor/signal behaviour, can be inferred from the measurements of neighboring sensors. Therefore, we present detailed discussion on the multi-sensor composite hypothesis schemes. Apart from that, it is also important to point out that, in most applications, the signal levels received at different areas of the WSN may be non-homogeneous, in the sense that the signal emitted from the source/target to be detected may affect a small subset of sensors. These affected sensors will be normally located close to the event (i.e., target, signal source, fire etc), and the rest of unaffected sensors observe very weak signal power or only noise. If we want to detect efficiently the emergence of such a signal, it is important to use only the observations from the good sensors that contain useful information and to suppress noise from the non-affected sensors. This leads us to say that a key point in hypothesis testing for WSN involves sensor selection, in order to pick the observations of the most relevant sensors for the detection rule. This selection brings us to the concept of rank-reduction or dimension reduction which make possible to improve the output SNR.

Keeping this fact into consideration, the main motivation of this chapter is to present a detailed overview of the fundamentals on composite multi-sensor detection, while introduc-

ing the concept of rank-reduction through model order selection. To do so we will divide the discussion to be conducted in this chapter into two approaches:

Approach 1: Multi-sensor detection scheme for deterministic signals

In the first approach (see Section 3.1) we will assume that the received signals at the sensors are unknown deterministic signals. The deterministic approach is a simple case that simplifies the formulation, allows to get insights on multi-sensor detection, and can be related to some practical cases of interest. For instance, in some applications of WSN, the local sensors receive the physical quantity and they convert this into electrical signal that is just a scalar values and unknown to the fusion center. Taking into account this, we just consider the values as unknown deterministic. However, we consider that the noise has Gaussian distribution without loss of generality. Hence, we discus the detection of unknown deterministic signal in the presence of Gaussian noise both with known and unknown noise powers. Next, in this category we propose detection techniques where the fusion center selects the relevant sensors with help of model order selection techniques [29, 30] that leads us to the concept of rank-reduction.

Approach 2: Multi-sensor detection scheme for stochastic signals

In many applications i.e., telecommunication, it is convenient to model the received signal with some stochastic models. Detection plays a pivotal role in the telecommunication systems specially in the case of spectrum sensing in cognitive radios. Hence, in the second category we consider the case of the detection of signals that are stochastic. Moreover, in this category we also propose detection techniques where the fusion center select important sensors that leads us to the concept of rank-reduction as it is in the previous category.

3.1 Multi-sensor detection scheme for deterministic signals

In this section we present the centralized detection scheme for multiple sensors in WSN that collaboratively detect deterministic signals emitted from an event (i.e., source or target). We assume that the WSN consists of K sensors that are uniformly deployed in a field. Moreover, we assume a single event that emits a signal with amplitude A and the sensors receive attenuated and noisy versions of it. Hence, the received signal at the k-th sensor at time n can be expressed as:

$$\begin{aligned} \mathcal{H}_0 :\ & x_k\left(n\right) = w_k\left(n\right), \\ \mathcal{H}_1 :\ & x_k(n) = A_k + w_k(n), \end{aligned} \tag{3.1}$$

with $k = 1, 2, 3, ...K$ as the sensor indexation. Similarly, the Gaussian distributed noise disturbance is given by $w_k(n) \sim \mathcal{N}\left(0, \sigma_k^2\right)$, and $A_k = A/d_{k,t}$ is the unknown received amplitude at sensor$k$ with $d_{k,t}$ is the distance from the event to this sensor. For the subsequent derivations, a set of N samples will be assumed to be available at the sensor as

$\mathbf{x}_k \triangleq [x_k(0), x_k(1), \ldots, x_k(N-1)]^T$. The fusion center stacks all of the K received vectors as $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_K \end{bmatrix}$ and process them for the detection of the event. In this section we assume that the signal parameters $\{A_k\}_{k=1}^K$ are unknown. Moreover, we further divide the problem of detecting unknown deterministic signal into two cases:

1. **Case Study # 1:-** Detection of unknown deterministic signal in presence of known noise.
2. **Case Study # 2:-** Detection of unknown deterministic signal in presence of unknown noise.

Hence, based on these facts, in Section (3.1.1) we derive the centralized system for the detection of unknown signal $\{A_k\}_{k=1}^K$ with assumption that the noise powers are known. In Section (3.1.2), we repeat the same by considering that the noise powers $\left\{\sigma_k^2\right\}_{k=1}^K$ are unknown.

3.1.1 Case Study #1: Unknown deterministic signal and known noise

In this case, we assume that the signal parameters $\{A_k\}_{k=1}^K$ are unknown and the noise powers $\left\{\sigma_k^2\right\}_{k=1}^K$ are known to the fusion center. The assumption is valid in those cases where the noise powers can be calibrated during the process of deployment of the wireless sensor network and noise powers are not affected by the random environmental effects. Now, in the presence of unknown parameters $\{A_k\}_{k=1}^K$ the hypothesis testing cannot be solved by using the optimal Neyman-Pearson criteria [25]. Therefore, we need to use a detection schemes that jointly perform the estimation and detection of the unknown deterministic signal. Hence, we discuss the GLRT that jointly performs the estimation and detection of the unknown deterministic signal $\{A_k\}_{k=1}^K$. The GLRT uses maximum likelihood estimate (MLE) to find the unknown parameters, and for the detection problem (3.1) with unknown $\{A_k\}_{k=1}^K$ it can be formulated as:

$$\mathcal{L}_{\text{G,Kn}}(\mathbf{X}) = \sum_{k=1}^{K} \log \left\{ \frac{\max\limits_{A_k} f_{\mathbf{x}}\left(\mathbf{x}_k; A_k, \sigma_k^2, \mathcal{H}_1\right)}{f_{\mathbf{x}}\left(\mathbf{x}_k; \sigma_k^2, \mathcal{H}_0\right)} \right\} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma, \tag{3.2}$$

where subscript "Kn" is used to distinguish the known noise case, $f_{\mathbf{x}}\left(\mathbf{x}_k; A_k, \sigma_k^2, \mathcal{H}_1\right)$ and $f_{\mathbf{x}}\left(\mathbf{x}_k; \sigma_k^2, \mathcal{H}_0\right)$ are the parametrized likelihood functions under hypotheses $\mathcal{H}_1$ and $\mathcal{H}_0$, respectively. The threshold γ is used by assuring a given constant probability of false alarm, P_{FA}. By considering the fact that the noise is Gaussian and independent across the time and the space dimensions, then the likelihood function under the hypothesis $\mathcal{H}_1$, $f_{\mathbf{x}}\left(\mathbf{x}_k; A_k, \sigma_k^2, \mathcal{H}_1\right)$ can be written as:

$$f_{\mathbf{x}}\left(\mathbf{x}_k; A_k, \sigma_k^2, \mathcal{H}_1\right) = \frac{1}{\sqrt{\left(2\pi\sigma_k^2\right)^N}} \exp\left\{ -\frac{1}{2\sigma_k^2} \sum_{n=1}^{N} \left(x_k(n) - A_k\right)^2 \right\} \tag{3.3}$$

Once we know the likelihood function (3.3), the MLE of the unknown parameter A_k can be found by differentiating $\log f_{\mathbf{x}}\left(\mathbf{x}_k; A_k, \sigma_k^2, \mathcal{H}_1\right)$ with respect to A_k and equating it to zero. Doing so we can find the MLE of A_k as: $\hat{A}_k = \frac{1}{N}\sum_{n=1}^{N} x_k(n)$. On the other hand the likelihood function under the hypothesis $\mathcal{H}_0$, $f_{\mathbf{x}}\left(\mathbf{x}_k; \sigma_k^2, \mathcal{H}_0\right)$ can be written as:

$$f_{\mathbf{x}}\left(\mathbf{x}_k; \sigma_k^2, \mathcal{H}_0\right) = \frac{1}{\sqrt{\left(2\pi\sigma_k^2\right)^N}} \exp\left\{-\frac{1}{2\sigma_k^2}\sum_{n=1}^{N} x_k^2(n)\right\} \tag{3.4}$$

It is to be noted that the noise is assumed to be known, therefore, we have no unknown parameter under the hypothesis $\mathcal{H}_0$. Based on the above facts, the final expression of the GLRT (3.2) becomes:

$$\mathcal{L}_{\text{G,Kn}}(\mathbf{X}) = \sum_{k=1}^{K} \frac{\bar{x}_k^2}{\sigma_k^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{3.5}$$

where $\bar{x}_k = \frac{1}{N}\sum_{n=1}^{N} x_k(n)$.

Now by following the above steps one can observe that in order to derive the detector (3.5), the fusion center needs to have all of the information from the local sensors. However, a major hurdle is the communication bandwidth constraint. Because in centralized scheme each local node transmits whole vector of samples $\{\mathbf{x}_k\}_{k=1}^{K}$ to the fusion center. In order to relax this requirement the k^{th} sensor should perform some local processing and instead of sending the whole vector it should send a summary of the information to the fusion center. In order to do so the detection scheme (3.5) can be implemented by using the following two steps:

1. The k-th sensor calculates the mean $\bar{x}_k$ which is sufficient statistic in this case and send it to the fusion center.
2. The fusion center calculates $\sum_{k=1}^{K}\left(\bar{x}_k/\sigma_k^2\right)^2$ for the final decision based on the received $\{\bar{x}_k\}_{k=1}^{K}$.

The process is further summarized in the form of an algorithm in 3.1 and shown in the Fig. 3.1

Algorithm 3.1 Implementation of $\mathcal{L}_{\text{G,Kn}}(\mathbf{x})$ as a decentralized detection with two steps method

1. Processing at the local sensors:
   - At the k-th sensor we have N samples of the receive signal as: $\mathbf{x}_k \triangleq [x_k(0), x_k(1), \ldots, x_k(N-1)]^T$.
   - Sensor k calculates $\bar{x}_k$ which also sufficient statistic in this case and sends it to the fusion center.
2. Processing at the fusion center:
   - The fusion center finds $\sum_{k=1}^{K} \bar{x}_k^2/\sigma_k^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma$.

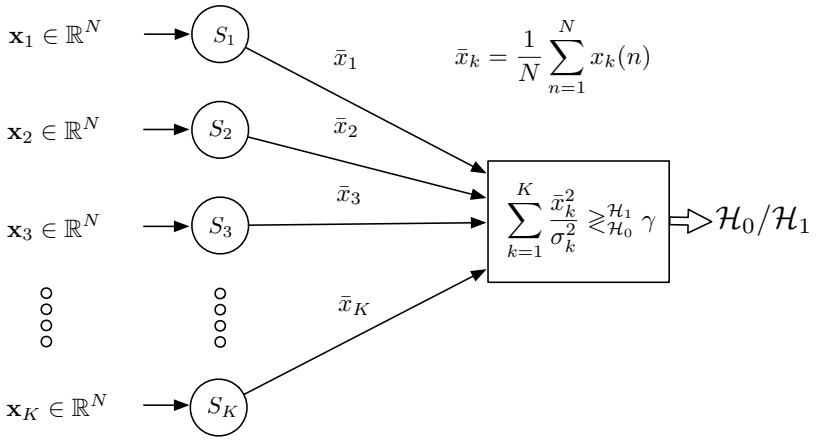


Figure 3.1: Two steps implementation of the GLRT for multi-sensor detection: Unknown deterministic signal and known noise

3.1.2 Case Study #2: Unknown deterministic signal and unknown noise

In the composite hypothesis problems some of the parameters appearing in a hypothesis may be unique to the hypothesis, such as amplitude or phase of a signal to be detected, whereas other parameters, such as those of interference signals and noise may be common to both hypotheses. Parameters which are common to all hypotheses are usually called *nuisance* parameters. As their name indicates, parameters of this type are an inconvenience since they do not play any role in determining which hypothesis is true.

In this chapter the example of nuisance parameters is noise powers as these are present in both hypotheses. In Section 3.1.1, we assumed that the noise powers are known to the fusion center. However, noise may change instantaneously due to the unknown environmental impairments (i.e., due to interference). It is because in practice the noise powers at local sensors not only depend on the circuitry rather they also depend upon the environment and they may instantaneously change. Moreover, the noise level also varies with presence and absence of signal [31]. This variation could be due, e.g., to the receiver electronics. For instance, when automatic gain control (AGC) is used, the noise factor depends on the signal amplitude. This effect is a well-known problem for digital cameras where signal-dependent noise is always present. More generally, any non-linearity in the electronics can modify the noise power by creating products between the noise part and the signal part [31]. Hence, in this type of situations, the assumption of known noise power or a-priori perfect estimation will indeed degrade the detection performance. In such situations one must resort to joint estimation of both noise and the unknown signal during the process of detection. Hence, in the remaining of the present section we revisit the above detection systems while considering unknown Gaussian noise. Keeping this in mind, in Section 3.1.2.1 we formulate multi-sensor detection scheme based on the GLRT but without the assumption of known noise powers. In the sequel, then in Section 3.1.2.2 we derive the multi-sensor detection scheme based on the Rao and Wald formulations for the the case of unknown signal parameters and unknown

noise powers.

3.1.2.1 GLRT

In this section, we discuss the problem of centralized detection of the unknown signal in the presence of Gaussian noise with unknown powers $\left\{\sigma_k^2\right\}_{k=1}^K$ at the local sensors. For this case the GLRT expression(3.2) becomes:

$$\mathcal{L}_{\text{G,Un}}\left(\mathbf{X}\right)=\sum_{k=1}^{K}\log\left\{\frac{\max\limits_{\sigma_{1,k}^2,A_k} f_{\mathbf{x}}\left(\mathbf{x}_k;A_k,\sigma_{1,k}^2,\mathcal{H}_1\right)}{\max\limits_{\sigma_{0,k}^2} f_{\mathbf{x}}\left(\mathbf{x}_k;\sigma_{0,k}^2,\mathcal{H}_0\right)}\right\}\gtrless_{\mathcal{H}_0}^{\mathcal{H}_1}\gamma, \tag{3.6}$$

where the subscript "Un" stands for the unknown noise power. Since, in practice the noise powers at the sensor k are different in the presence and absence of the target signal, therefore, $\sigma_{1,k}^2$, $\sigma_{0,k}^2$ for $k=1,2,...,K$ are the noise powers under hypotheses $\mathcal{H}_1$ and $\mathcal{H}_0$, respectively. As it was in the previous section, the estimates of the unknown parameters under the hypothesis $\mathcal{H}_1$ can be found by differentiating $\log f_{\mathbf{x}}\left(\mathbf{x}_k;A_k,\sigma_{1,k}^2,\mathcal{H}_1\right)$ with respect to A_k and $\sigma_{1,k}^2$ and equating it to zero. In the same way we can find the estimate of the unknown $\sigma_{0,k}^2$ under hypothesis $\mathcal{H}_0$ by differentiating $\log f_{\mathbf{x}}\left(\mathbf{x}_k;\sigma_{0,k}^2,\mathcal{H}_0\right)$ with respect to $\sigma_{0,k}^2$ and equating it to zero. Doing so the final expression of the GLRT (3.6) becomes

$$\mathcal{L}_{\text{G,Un}}\left(\mathbf{X}\right)=\sum_{k=1}^{K}\log\left(\frac{\hat{\sigma}_{0,k}^2}{\hat{\sigma}_{1,k}^2}\right)\gtrless_{\mathcal{H}_0}^{\mathcal{H}_1}\gamma, \tag{3.7}$$

where $\hat{\sigma}_{0,k}^2=\frac{1}{N}\sum_{n=1}^{N}x_k^2\left(n\right)$ and $\hat{\sigma}_{1,k}^2=\frac{1}{N}\sum_{n=1}^{N}\left(x_k^2\left(n\right)-\bar{\mathbf{x}}_k^2\right)$. By putting these values in (3.7) we can achieve the following expression

$$\begin{aligned}\mathcal{L}_{\text{G,Un}}\left(\mathbf{X}\right)&=\sum_{k=1}^{K}\log\left(1+\frac{\bar{x}_k^2}{\hat{\sigma}_{1,k}^2}\right)\gtrless_{\mathcal{H}_0}^{\mathcal{H}_1}\gamma,\\&=\sum_{k=1}^{K}\mathcal{L}_{\text{G,Un}}^{k}\left(\mathbf{x}_k\right)\gtrless_{\mathcal{H}_0}^{\mathcal{H}_1}\gamma,\end{aligned} \tag{3.8}$$

where $\mathcal{L}_{\text{G,Un}}^{k}\left(\mathbf{x}_k\right)=\log\left(1+\frac{\bar{x}_k^2}{\hat{\sigma}_{1,k}^2}\right)$ for $k=1,2,\cdots,K$ is the local GLRT at sensor k. This detector has the advantage that it does not need the knowledge of noise and it also considers the fact that the noise power under $\mathcal{H}_1$ is different from the noise power under $\mathcal{H}_0$. In the same way as we discussed in section **??**, the decentralized detection strategy can be established as follow:

1. Every sensor node e.g k-th for $k=1,2,.....,K$, calculate $\mathcal{L}_{\text{G,Un}}^{k}\left(\mathbf{x}_k\right)$.
2. All K sensors send $\mathcal{L}_{\text{G,Un}}^{k}\left(\mathbf{x}_k\right)$ to the fusion center and the fusion center calculates (3.8) for the final decision.

In order to implement the two steps, we must assume that the local sensors have capability to estimate their corresponding noise powers based on the N received samples to find $\left\{\mathcal{L}^k_{\text{G,Un}}\left(\mathbf{x}_k\right)\right\}_{k=1}^{K}$ in (3.8). The process of the two step method is summarized in algorithm 3.2 and is also given in Fig.3.2. The two steps method for (3.8) can be considered more decentralized compared to the case of known noise power since herein the local sensors need to estimate unknown noise powers locally.

Algorithm 3.2 Implementation of $\mathcal{L}_{\text{G,Un}}(\mathbf{x})$ as a decentralized detection with two steps method

1. Processing at the local sensors:
   - The k-th sensor receives N samples as:
   $$\mathbf{x}_k \triangleq [x_k(0), x_k(1), \ldots, x_k(N-1)]^T .$$
   - Every sensor calculates $\left\{\frac{\bar{x}_k^2}{\hat{\sigma}^2_{1,k}}\right\}_{k=1}^{K}$ and sends it to the fusion center.
2. Processing at the fusion center:
   - The fusion center finds $\sum_{k=1}^{K}\log\left(1+\frac{\bar{x}_k^2}{\hat{\sigma}^2_{1,k}}\right)\underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}}\gamma$.

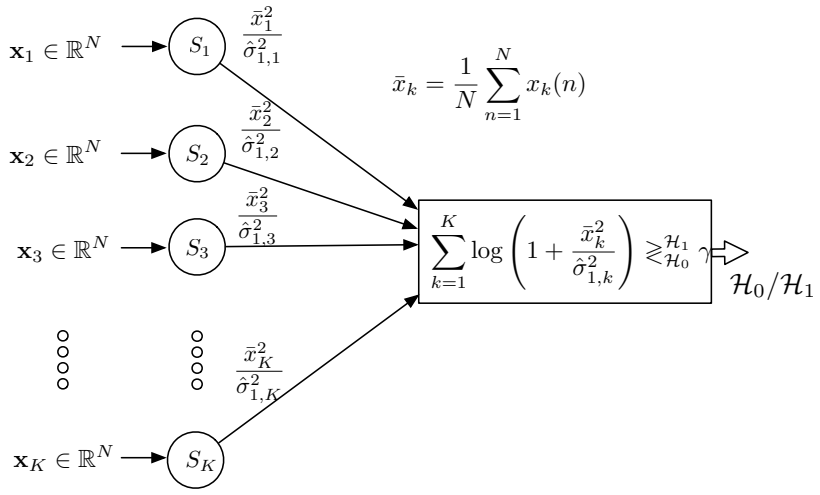


Figure 3.2: Two steps implementation of the GLRT for multi-sensor detection: Unknown deterministic signal and unknown noise

3.1.2.2 Rao and Wald test statistics

The formulation of Rao test for our problem can be expressed as:

$$\mathcal{L}_{\text{R,Un}}\left(\mathbf{X}\right) = \sum_{k=1}^{K}\frac{\partial \log f_{\mathbf{x}}\left(\mathbf{x}_k; A_k, \sigma^2_{1,k}, \mathcal{H}_1\right)}{A_k}\Big|^2_{A_k=0,\sigma^2_{1,k}=\sigma^2_{0,k}}\left[\mathbf{I}^{-1}\left(\tilde{\theta}\right)\right]_{A_k,A_k} \tag{3.9}$$

where $\tilde{\theta} = \left[A_k = 0, \hat{\sigma}^2_{0,k}\right]^T$ are the MLE of the unknown parameters under the hypothesis, $\mathcal{H}_0$. In order to solve (3.9) we can find that

$$\begin{aligned}\frac{\partial \log f_{\mathbf{x}}\left(\mathbf{x}_k; A_k, \sigma^2_{1,k}, \mathcal{H}_1\right)}{A_k} \mid_{A_k=0,\sigma^2_k=\sigma^2_{0,k}} &= \sum_{k=1}^{K} \frac{N}{\sigma^2_{1,k}} \left(x_k - A_k\right) \mid_{A_k=0,\sigma^2_k=\sigma^2_{0,k}} \\ &= \frac{N\bar{x}_k}{\hat{\sigma}^2_{0,k}}\end{aligned} \tag{3.10}$$

where $\hat{\sigma}^2_{0,k} = \frac{1}{N}\left\|\mathbf{x}_k\right\|^2$. Similarly, we can also write [25]

$$\left[\mathbf{I}^{-1}\left(\tilde{\theta}\right)\right]_{A_k A_k} = \left\{I_{A_k,A_k}\left(\tilde{\theta}\right) - I_{A_k,\sigma^2_{1,k}}\left(\tilde{\theta}\right) I_{\sigma^2_{1,k},\sigma^2_{1,k}}\left(\tilde{\theta}\right) I_{\sigma^2_{1,k},A_k}\left(\tilde{\theta}\right)\right\}^{-1} \tag{3.11}$$

where we can easily find $I_{A_k,A_k}\left(\tilde{\theta}\right)$ by using the expression [32, Chap. 3]:

$$I_{A_k,A_k}\left(\tilde{\theta}\right) = -\mathrm{E}\left[\frac{\partial^2 \log f_{\mathbf{x}}\left(\mathbf{x}_k; A_k, \sigma^2_{1,k}, \mathcal{H}_1\right)}{\partial A_k \partial A_k}\right] = \frac{N}{\hat{\sigma}^2_{0,k}}. \tag{3.12}$$

Similarly, we can easily verify that $I_{\sigma^2_k,A_k}\left(\tilde{\theta}\right) = I_{A_k\sigma^2_k}\left(\tilde{\theta}\right) = 0$, therefore we get

$$\left[\mathbf{I}^{-1}\left(\tilde{\theta}\right)\right]_{A_k A_k} = \left\{\frac{N}{\hat{\sigma}^2_{0,k}}\right\}^{-1} \tag{3.13}$$

By putting (3.10) and (3.13) in (3.9) we can write the decision rule based on the Rao test (3.9) as:

$$\mathcal{L}_{\mathrm{R,Un}}\left(\mathbf{X}\right) = \sum_{k=1}^{K} \frac{\bar{x}^2_k}{\hat{\sigma}^2_{0,k}} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma. \tag{3.14}$$

By comparing the Rao statistic (3.14) with the GLRT (3.8), we can see that in the case of the unknown noise powers at the local sensor the two statistics are different and the difference lies in the denominator. However, we will see that this difference exist only in the case of higher SNR. For low SNR the two test statistics can be proved to be equal. In order to prove this we recall from the GLRT expression in (3.8) that

$$\begin{aligned}2\mathcal{L}^k_{\mathrm{G,Un}}\left(\mathbf{x}_k\right) &= \log\left(1 + \frac{\bar{x}^2_k}{\hat{\sigma}^2_{1,k}}\right) \\ &= \log\left(\frac{1}{1 - \frac{\bar{x}^2_k}{\hat{\sigma}^2_{0,k}}}\right).\end{aligned} \tag{3.15}$$

When $\frac{\bar{x}_k^2}{\hat{\sigma}_{0,k}^2} \ll 1$, then $\log\left(\frac{1}{1-\frac{\bar{x}_k^2}{\hat{\sigma}_{0,k}^2}}\right) \backsimeq \frac{\bar{x}_k^2}{\hat{\sigma}_{0,k}^2}$ [25, Example 6.10]. Hence, in the case $\frac{\bar{x}_k^2}{\hat{\sigma}_{0,k}^2} \ll 1$ we can develop the relationship $\mathcal{L}_{\text{R,Un}}(\mathbf{x}) \simeq \mathcal{L}_{\text{G,Un}}(\mathbf{x})$. On the other hand for the same detection problem, the Wald test decides $\mathcal{H}_1$ if [25]

$$\mathcal{L}_{\text{W,Un}}(\mathbf{X}) = \sum_{k=1}^{K} \hat{A}_k^2 \left[\mathbf{I}^{-1}\left(\hat{\theta}_1\right)\right]_{A_k,A_k} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma. \tag{3.16}$$

where $\hat{\theta}_1$ is the estimate of the unknown parameters under the hypothesis $\mathcal{H}_1$. Solving (3.16) by using (3.13), we get

$$\mathcal{L}_{\text{W,Un}}(\mathbf{X}) = \sum_{k=1}^{K} \frac{\bar{x}_k^2}{\hat{\sigma}_{1,k}^2} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma. \tag{3.17}$$

where $\hat{\sigma}_{1,k}^2 = \frac{1}{N}\sum_{n=1}^{N}\left(x_k^2(n) - \bar{x}_k^2\right)$, $\hat{A}_k^2 = \bar{x}_k^2$. In order to facilitate the reader, we summarize the final expressions of the detectors in table 3.1.

Name of detector	Expression
GLRT (known noise)	$\mathcal{L}_{\text{G,Kn}}(\mathbf{X}) = \sum_{k=1}^{K} \frac{\bar{x}_k^2}{\sigma_k^2} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma$
GLRT (unknown noise)	$\mathcal{L}_{\text{G,Un}}(\mathbf{X}) = \sum_{k=1}^{K} \log\left(1 + \frac{\bar{x}_k^2}{\hat{\sigma}_{1,k}^2}\right) \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma$
Rao (unknown noise)	$\mathcal{L}_{\text{R,Un}}(\mathbf{X}) = \sum_{k=1}^{K} \frac{\bar{x}_k^2}{\hat{\sigma}_{0,k}^2} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma$
Wald (unknown noise)	$\mathcal{L}_{\text{W,Un}}(\mathbf{X}) = \sum_{k=1}^{K} \frac{\bar{x}_k^2}{\hat{\sigma}_{1,k}^2} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma$

Table 3.1: Summary of detectors

3.1.2.3 Numerical example

In this section we will compare the detection schemes discussed above with a simple numerical example by using Matlab simulations. As an example we consider a scenario with $K = 20$ sensors that are randomly distributed in a square field. Regarding the assessment of the detectors we will analyze their performance through the use of ROC. Herein, we use the average SNR of all radio as: $\bar{\kappa} = \frac{1}{K}\sum_{k=1}^{K}\kappa_i$ and the SNR of $k-$ th sensor is $\kappa_k = \zeta^{k-1}\bar{\kappa}$, for $k = 1, 2, ..., K$. For a given average SNR $\bar{\kappa}$ and SNR gap ζ, we can generate the SNRs of sensors. The match of the SNRs to the sensors can be random. In order to compare the detection performance of the three detectors, we conducted three different experiments by considering $\bar{\kappa} = -12$dB, $\bar{\kappa} = -18$dB., $\bar{\kappa} = -22$dB.

In the first experiment we fix the average SNR $\bar{\kappa}$ at: $\bar{\kappa} = -12$dB and plot the ROC curves for two different values of the sample size N: as $N = 3$ and $N = 6$. The results are shown in Fig. 3.3 that clearly show that the GLRT performs better compared to the Rao and Wald tests for this specific case. Apart from this an interesting point to be noted is the performance of the Rao test at different values of N, as we can see that at a very small N, its performance is worst than the performance of Wald test.

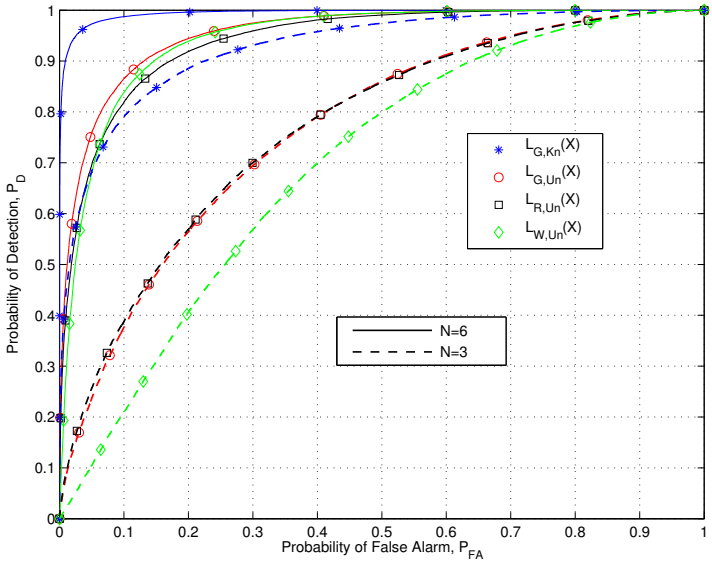


Figure 3.3: ROC Curves to compare the performance of the GLRT, Rao Test and Wald Test for the case of unknown deterministic signal and unknown noise: Average SNR $\bar{\kappa} = -12$dB.

In order to further analyze the three detectors, in the second experiment we fix the average SNR $\bar{\kappa}$ at: $\bar{\kappa} = -18$dB and plot the ROC curves for $N = 5$, $N = 10$ that are shown in Fig. 3.4. From the result in this figure and Fig. 3.3 we can easily conclude that the performance of the Rao test varies for different values of N at a specific value of average SNR. From both figures, we can observe that with increasing N, the detection performance of the Rao test reaches to the performance level of GLRT. Furthermore, in Fig. 3.4 it is also interesting to see that the Rao and the GLRT perform better than the Wald test at $N = 10$, while at $N = 5$ the performance of the Wald test is at par of the GLRT.

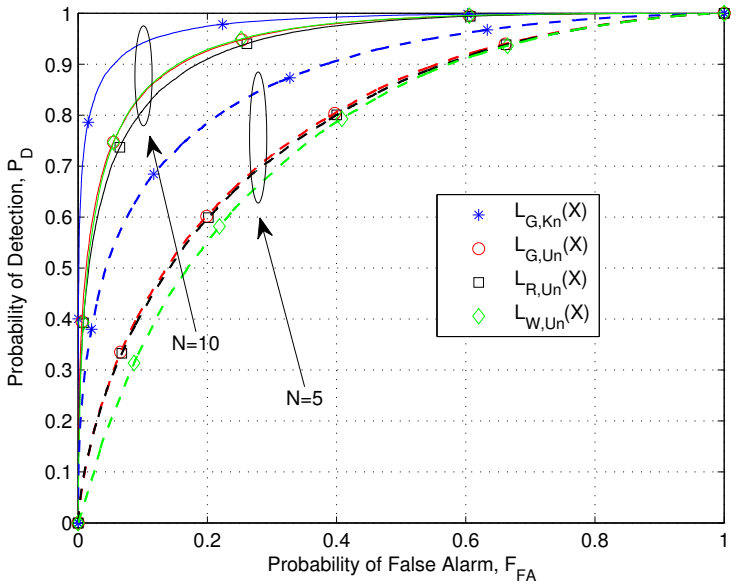


Figure 3.4: ROC Curves to compare the performance of the GLRT, Rao test and Wald test for the case of unknown deterministic signal and unknown noise: Average SNR $\bar{\kappa} = -18$dB.

For the plots in Fig. 3.5, we repeat the above experiment by lowering the average SNR further at at: $\bar{\kappa} = -22$dB. In this experiment the detection performance of the three detectors is shown for two different sample size $N = 30$ and $N = 60$. As expected the results in this experiment show that for higher N and small average SNR, the three detection schemes yield about the same performance and reaches to the level of the known noise GLRT $\mathcal{L}_{\mathrm{G,Kn}}(\mathbf{X})$. The results also show that at very low SNR ($\bar{\kappa} = -22$dB), the performance of the Rao is very much at par of the GLRT, which confirms the analytical reasoning in Section 3.1.2.2. Overall, the results obtained in these experiments show that as $N \to \infty$, the performance of the Rao and Wald detectors coincides with that of the GLRT, but for the finite data records, the performance of these detection schemes is not the same (sometimes the Rao detector approaches the GLRT and sometimes it is the Wald detector the one approaching the GLRT). In the end, their main advantage compared to the GLRT is that these can be easily derived.

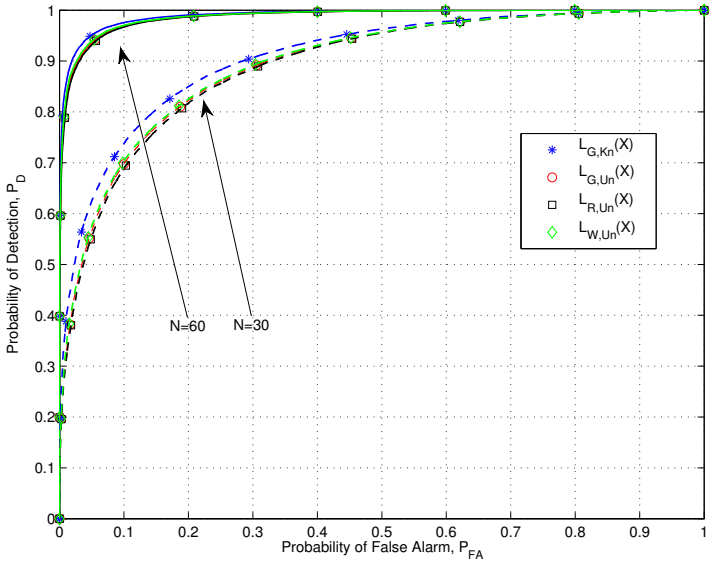


Figure 3.5: ROC Curves to compare the performance of the GLRT, Rao test and Wald test for the case of unknown deterministic signal and unknown noise: Average SNR $\bar{\kappa} = -22$dB.

3.1.3 Rank-reduction approach

In some applications of a sensor network, the signal may be a local disturbance that only affects a small subset of sensors. Due to limitations in the sensing range, only some sensors would observe that signal and the rest would receive small signal or no signal at all, hence, observing mostly noise. This is particularly valid in the detection of events in sensor network spread over a large area, since the emitted signal power decays isotropically as a function of distance, and it is affected by a significant path loss attenuation due to the large area being covered by the network, as well as by fading/shadowing effects. As a consequence, only a subset of sensors will be able to receive enough power levels so as to easily detect the presence of the event with a given detection performance [33]. The rest of sensors will typically receive extremely weak power levels. This observation allows us to distinguish between the so-called *active* and *inactive* sensors, respectively. If we want to detect efficiently the emergence of an event, it is important to use only the observations that contain useful information and to suppress noise from the non-affected sensors. Therefore, the performance of the signal detection technique can be improved by identifying the useful observation samples among the set of all received measurements [30]. This process of important sensor selection will lead us to the rank-reduced version of the original signal space. Hence, it will allow us to benefit from an equivalent SNR gain due the reduced dimensionality subspace.

3.1.3.1 Rank-reduced signal model

In order to describe the signal model, we assume the case where sensors are not just intercept or dumb sensors rather they do some local processing and send summary of the observations to the fusion center. In other words, after receiving the N $(n = 1, 2, ..., N)$ samples of $x_k(n) = A_k + w_k(n)$ for $k = 1, 2, ..., K$, each sensor calculates the mean as: $\bar{x}_k = \frac{1}{N}\sum_{n=1}^{N} x_k(n)$ and dispatches it to the fusion center. The fusion center collects the measurements of K sensors in a vector $\mathbf{x} = [\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_K]^T$ and the signal model at the fusion center becomes

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{x} = \mathbf{w}, \\ \mathcal{H}_1 &: \mathbf{x} = \mathbf{s} + \mathbf{w}, \end{aligned} \tag{3.18}$$

where $\mathbf{s} = [A_1, A_2, \cdots, A_K]^T$, $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_w)$ with $\mathbf{\Sigma}_w = \text{diag}\left[\sigma_1^2/N, \sigma_2^2/N, \cdots, \sigma_K^2/N\right]$, as the noise is considered to be independent with distinct noise powers at different sensors. Hence, under $\mathcal{H}_0$ $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_w)$ and under $\mathcal{H}_1$ $\mathbf{x} \sim \mathcal{N}(\mathbf{s}, \mathbf{\Sigma}_w)$. The detection problem (4.1) is nothing but to differentiate $\mathbf{s} \neq \mathbf{0}$ from $\mathbf{s} = \mathbf{0}$. The vector $\mathbf{s}$ contains the received unknown deterministic amplitudes of received signals at the K sensors. In this vector the samples corresponding to active sensors will be non-zero and rest of the samples will have extremely weak values that will hardly contribute constructively in the detection performance. To be more specific we are indeed in front of situation where vector $\mathbf{s}$ is a sparse vector. Hence, in the following we take into effect this sparsity to model the detection scheme. If we assume that there are $\mathcal{K}$ non-zero samples (samples corresponding to the non-active sensors) in $\mathbf{s}$ then the model order of (4.1) is $\mathcal{K}$ and $\mathcal{K}$ can be any number in the range $1 \leq \mathcal{K} \leq K$, and it can be found by using model order selection technique [29]. Once we know $\mathcal{K}$, then the signal model can be expressed as:

$$\mathbf{x} = \begin{bmatrix} \mathbf{I}_{\mathcal{K}} \\ \mathbf{0}_{(K-\mathcal{K})\times\mathcal{K}} \end{bmatrix} \mathbf{s}_{\mathcal{K}} + \mathbf{w} \tag{3.19}$$

where $\mathbf{s}_{\mathcal{K}}$ is a vector containing only the $\mathcal{K}$ non-zero signals. It is important to mention that the formulation in (3.19) assumes that the non-zero signal samples have been ordered. We remark that this assumption is not valid in practice and we will relax it later-on in the discussion. Similarly, it is clear by observing signal model (3.19) that it resembles the well known classical linear model. It can also be inferred from (3.19) that it is a problem of rank-reduction because the number of useful signal samples is smaller than the total number of samples. Then for the rank-reduced version of the signal model (4.1) can be written as:

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{x} = \mathbf{w}, \\ \mathcal{H}_1 &: \mathbf{x} = \mathbf{T}_{\mathcal{K}}\mathbf{s}_{\mathcal{K}} + \mathbf{w} \end{aligned} \tag{3.20}$$

where $\mathbf{T}_{\mathcal{K}} = \begin{bmatrix} \mathbf{I}_{\mathcal{K}} & \mathbf{0}_{(K-\mathcal{K})\times\mathcal{K}} \end{bmatrix}^T$. The detection problem to be solved must cope with the presence of a set of unknown parameters of unknown dimension $\mathcal{K}$ (i.e. unknown length). The traditional GLRT will always implement the test statistic based on the maximum

order K, and thus it always includes observation samples that contain only noise [30] when indeed these observations should be discarded. This occurs because the multiple alternative probability density functions (PDF) are a set of nested PDF families. The net effect is that the test statistic will always overestimate the actual model order, thus including dimensions of the signal subspace where almost no signal, but only noise is present. This results in a reduction of the power of the detector to produce a desired result [30]. This problem will occur whenever the number of signal components is unknown and that is why the GLRT posses limitation in this type of detection problems. That is why $\mathcal{K}$ should first be estimated and then GLRT is conducted based on the estimated model $\hat{\mathcal{K}}$.

3.1.3.2 Multifamily Likelihood Ratio Test

Estimation of the true model order $\mathcal{K}$ and GLRT can be found jointly by using the multifamily likelihood ratio test (MFLRT). MFLRT is proposed in [30] and is given by:

$$T_{\text{MFLRT}}(x) = \max_{1\leq i\leq K} f\left(\bar{L}_{G_i}(x)\right) \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma. \tag{3.21}$$

where $f\left(\bar{L}_{G_i}(x)\right)$ is a key transforming function used to accommodate different signal models, with $\bar{L}_{G_i}(\mathbf{x}) = 2\ln L_{G_i}(\mathbf{x})$ and $L_{G_i}(\mathbf{x})$ as generalized likelihood ratio (GLR) while considering i as the size of the vector of true signals. The GLR $L_{G_i}(\mathbf{x})$ is found in section 3.1.3.3. In (3.21) γ is the threshold. Function $f\left(\bar{L}_{G_i}(x)\right)$ is given as [30]:

$$f\left(\bar{L}_{G_i}(x)\right) = \left[\bar{L}_{G_i}(x) - i\left(\ln\left(G_i(x)\right)+1\right)\right] u\left(G_i(x)-1\right) \tag{3.22}$$

where $G_i(x) = \frac{\bar{L}_{G_i}(x)}{i}$ and $u\left(x\right)$ is the unit step function. It is interesting to note that the MFLRT extends the GLRT to allow testing with multiple alternative model orders.

3.1.3.3 Rank-Reduced Generalized Likelihood Ratio

Before introducing the proposed rank reduced GLRT to be used in (3.22), we present here the test statistic based on the classical linear model, $\mathbf{y} = \mathbf{B}\theta + \mathbf{v}$ with θ as a $p \times 1$ (i.e. $p \leq K$) vector of unknown parameters, $\mathbf{v}$ as a vector that contains errors with PDF $\mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}_v\right)$ and $\mathbf{B}$ as a known $K \times p$ observation matrix. This will facilitate us in finding $\bar{L}_{G_i}(x)$. The likelihood function when $\theta \neq 0\left(\mathcal{H}_1\right)$ can be written as:

$$f_{\mathbf{y}}\left(\mathbf{Y};\theta,\mathcal{H}_1\right) = \frac{1}{\sqrt{2\pi\mathbf{\Sigma}_v}}\exp\left[-\frac{1}{2}\left(\mathbf{y}-\mathbf{B}\theta\right)^T\mathbf{\Sigma}_v^{-1}\left(\mathbf{y}-\mathbf{B}\theta\right)\right] \tag{3.23}$$

and in the case when $\theta \neq 0\left(\mathcal{H}_1\right)$, the likelihood function can be written as:

$$f_{\mathbf{y}}\left(\mathbf{Y};\mathcal{H}_0\right) = \frac{1}{\sqrt{2\pi\mathbf{\Sigma}_v}}\exp\left[-\frac{1}{2}\mathbf{y}^T\mathbf{\Sigma}_v^{-1}\mathbf{y}\right] \tag{3.24}$$

Using this model for detecting the presence of $\theta \neq 0\,(\mathcal{H}_1)$ against the case $\theta = 0\,(\mathcal{H}_0)$ results in the test statistic as [25, Theorem 7.1,p. 274]:

$$L_G(\mathbf{y}) = \mathbf{y}^T \boldsymbol{\Sigma}_v^{-1} \mathbf{P}_B \mathbf{y} - \frac{1}{2} \mathbf{y}^T \mathbf{P}_B \boldsymbol{\Sigma}_v^{-1} \mathbf{P}_B \mathbf{y} \tag{3.25}$$

where we used the fact that $\hat{\theta} = \left(\mathbf{B}^H \mathbf{B}\right)^{-1} \mathbf{B}^H \mathbf{y}$ is the ML-estimate of θ with $\mathbf{P}_B = \mathbf{B}\left(\mathbf{B}^H \mathbf{B}\right)^{-1} \mathbf{B}^T$ as the projection matrix.

Based on (3.25) for the signal model (3.20) the expression for the GLR, $L_{G_i}(\mathbf{x})$ to be used in (3.21), can be written as:

$$\begin{aligned} L_{G_i}(\mathbf{x}) &= \mathbf{x}_{\text{sort}}^T \boldsymbol{\Sigma}_{w,\text{sort}}^{-1} \mathbf{P}_{T_i} \mathbf{x}_{\text{sort}} - \frac{1}{2} \mathbf{x}_{\text{sort}}^T \mathbf{P}_{T_i} \boldsymbol{\Sigma}_{w,\text{sort}}^{-1} \mathbf{P}_{T_i} \mathbf{x}_{\text{sort}} \\ &= \frac{1}{2} \mathbf{x}_{\text{sort}}^T \mathbf{P}_{T_i} \boldsymbol{\Sigma}_{w,\text{sort}}^{-1} \mathbf{P}_{T_i} \mathbf{x}_{\text{sort}} \end{aligned} \tag{3.26}$$

where $\boldsymbol{\Sigma}_{w,\text{sort}} = \text{diag}\left(\text{sort}\left[\sigma_1^2/N, \sigma_2^2/N, \cdots, \sigma_K^2/N\right]\right)$ and we used the fact that while considering that $\boldsymbol{\Sigma}_{w,\text{sort}}^{-1}$ is diagonal matrix then $\mathbf{P}_{T_i} \boldsymbol{\Sigma}_{w,\text{sort}}^{-1} \mathbf{P}_{T_i} = \boldsymbol{\Sigma}_{w,\text{sort}}^{-1} \mathbf{P}_{T_i}$ and

$$\mathbf{P}_{T_i} = \mathbf{T}_i \left(\mathbf{T}_i^T \mathbf{T}_i\right)^{-1} \mathbf{T}_i^T = \begin{bmatrix} \mathbf{I}_i & \mathbf{0}_{i\times(K-i)} \\ \mathbf{0}_{(K-i\times i)} & \mathbf{0}_{(K-i)\times(K-i)} \end{bmatrix}.$$

Note that $\mathbf{x}_i$ is the vector containing first i samples of the sorted observation vector $\mathbf{x}_{\text{sort}}$ (i.e. $\mathbf{x}_{\text{sort}}$=sort($\mathbf{x}$)). In practice we don't have knowledge of the ordering and thus first $\mathbf{s}$ is estimated as $\hat{\mathbf{s}} = \left(\mathbf{T}^T \mathbf{T}\right)^{-1} \mathbf{T}^T \mathbf{x}$, then the magnitudes of the elements in $\hat{\mathbf{s}}$ are sorted in descending order and after that $\mathbf{x}$ is ordered according to ordered $\hat{\mathbf{s}}$. The reason for this sorting is that MFLRT assumes ordered true signal vector $\mathbf{s}$. Based on the indexation of $\mathbf{x}_{\text{sort}}$ then we find $\boldsymbol{\Sigma}_{w,\text{sort}}$ from $\boldsymbol{\Sigma}_w$. With this in mind, we can also write (3.26) as:

$$\begin{aligned} L_{G_i}(\mathbf{x}) &= \frac{1}{2} \mathbf{x}_{\text{sort}}^T \mathbf{P}_{T_i} \boldsymbol{\Sigma}_{w,\text{sort}}^{-1} \mathbf{P}_{T_i} \mathbf{x}_{\text{sort}} \\ &= \frac{1}{2} \mathbf{x}_{\text{P,sort}}^T \boldsymbol{\Sigma}_{w,\text{sort}}^{-1} \mathbf{x}_{\text{P,sort}} \end{aligned} \tag{3.27}$$

where $\mathbf{x}_{\text{P,sort}} = \mathbf{P}_{T_i} \mathbf{x}_{\text{sort}}$ is the projected version of vector $\mathbf{x}_{\text{sort}}$, to space spanned by the vector that contains the signal sample received at the active sensors. Based on $\mathbf{P}_{T_i}$ we can write the simplified version of the (3.27) as:

$$L_{G_i}(\mathbf{x}) = \frac{1}{2} \sum_{k=1}^{i} \frac{\bar{\mathbf{x}}_k^2}{\sigma_k^2} \tag{3.28}$$

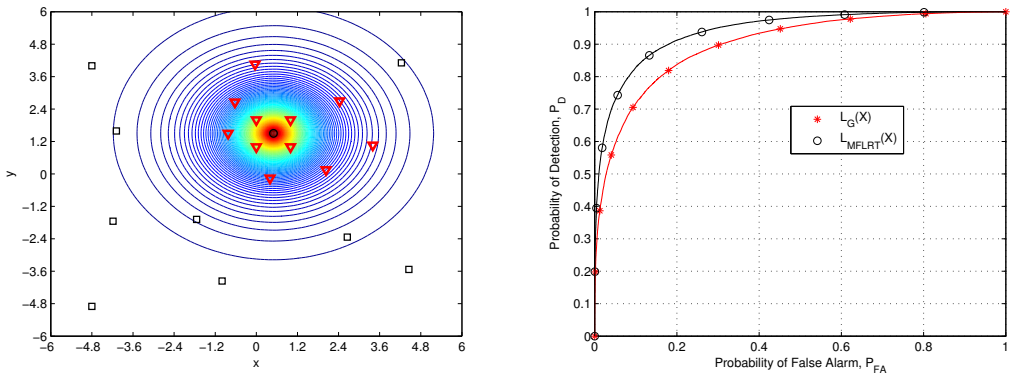
The implementation of the process has been given in the form of algorithm (3.3)

Algorithm 3.3 Implementation process of the proposed detector without spatial information.

1. The received observations from the sensor field are stacked in vector $\mathbf{x}$.
2. Take the absolute values$|\hat{\mathbf{s}}|$ and sort $|\hat{\mathbf{s}}|$ in descending order.
3. Order $\mathbf{x}$ according to sorted $|\hat{\mathbf{s}}|$ to get $\mathbf{x}_{\text{sort}}$.
4. Based on the sorted $\mathbf{x}_{\text{sort}}$ shuffle $\mathbf{n}_u = \left[\sigma_1^2/N, \sigma_2^2/N, \cdots, \sigma_K^2/N\right]$ to get sort $\tilde{\mathbf{n}}_u$.
5. Find $\boldsymbol{\Sigma}_{w,\text{sort}} = \text{diag}\left(\tilde{\mathbf{n}}_u\right)$
6. Implement (3.21) by using (3.28) as:
   - $\mathbf{t} \leftarrow \langle\rangle$
   - while $i \leqq K$ do
     - Calculate$\ln L_{G_i}(\mathbf{x}) = \frac{1}{2}\mathbf{x}_{\text{sort}}^T \mathbf{P}_{T_i} \boldsymbol{\Sigma}_{w,\text{sort}}^{-1} \mathbf{P}_{T_i} \mathbf{x}_{\text{sort}}$
     - Obtain $f_i = f\left(\bar{L}_{G_i}(x)\right)$ as given in (3.22)
     - Push f_i onto $K \times 1$ vector $\mathbf{t}$
     - $i = i + 1$
     - end while
   - $\mathbf{t} = \left[\ f_1, \quad f_2, \quad \cdots, \quad f_K\ \right]$
   - $T_{\text{MFLRT}}(x) = \max(\mathbf{t}) \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma$ as in (3.21) and the index corresponding to the maximum is $\hat{\mathcal{K}}$.

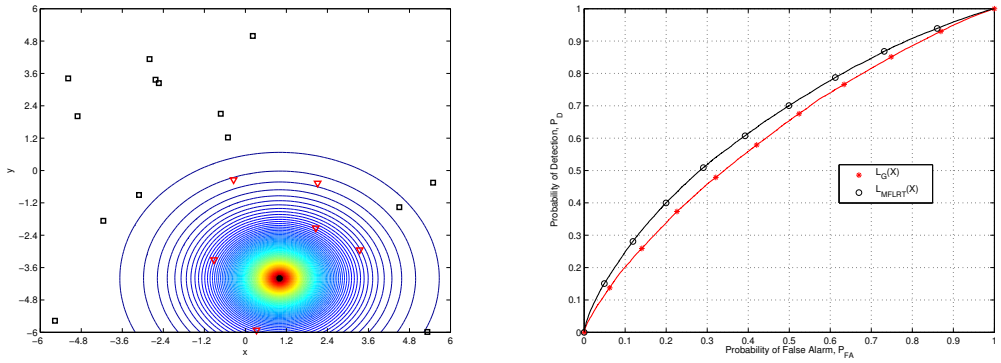
3.1.3.4 Numerical example

In this section, we present results of computer simulations to illustrate the comparison of the detection performance of the rank reduced detector with the traditional GLRT. Experiments are performed with a total number of $K = 20$ sensor nodes, which are randomly distributed in the sensor field of size 12 sq-meters. As an example, two simplified scenarios are depicted in Fig.3.6a and Fig.3.7a. To make the scenarios resemble to more practical cases, we also fix the sensing range for the sensor nodes to $R = 3$ meters. In both figures, the event is represented with thick black circle. The sensors that find the event within their sensing ranges are denoted by "∇", while "$\square$" indicate those sensors that receive zero amplitude from the event. Using procedure similar to Section 3.1.2.3 we use the average SNR of all sensors as: $\bar{\kappa} = -8$dB. On the basis of these scenarios, we analyze the rank-reduced detector with the help of ROC curves given in Fig.3.6b and 3.7b . In both figures subscript “MFLRT” indicates the detector that uses the sensor selection based on MFLRT.



(a) Field scenario#1 **(b)** ROC curves for scenario #1.

Figure 3.6: Experiment #1: ROC curves: Comparison of the MFLRT with the traditional GLRT for unknown deterministic signal: Scenario #1



(a) Field scenario #2 **(b)** ROC curves for scenario #2.

Figure 3.7: Experiment #2: ROC curves: Comparison of the MFLRT with the traditional GLRT for unknown deterministic signal: Scenario #2

Results in Fig.3.6b and 3.7b clearly show that the MFLRT-based detector has a better ROC characteristic compared to the traditional GLRT. The MFLRT selects those sensors that receive observations with better SNR, thus improving the overall SNR. Therefore, this rank-reduced multi-sensor detection (MFLRT-based detector) leads to a better performance, when the signal emitted from the event effects only a subset of sensors. Similarly, comparing Fig.3.7b with Fig.3.6b, we can observe that the far location of the sensors from the target degrades the performance of both detectors in experiment #2. However, the MFLRT-based detector still outperforms the traditional GLRT. This suggests that the proposed algorithm 3.3 can achieve an improved detection performance without taking into account the topology of sensors.

3.2 Multi-sensor detection of stochastic signal

Stochastic processes play a fundamental role in mathematical models of phenomena occurring in nature. In communication systems transmitted signal after random effects or any electromagnetic emission from a physical event are mostly approximated with the stochastic signals such as Gaussian [34]. The Gaussian signal approximation is also widely followed in cooperative spectrum sensing that is based on the concept of multi-sensor detection scheme. The Gaussian approximation in addition to resulting in tractable models and useful detectors, is reasonable if the primary network employs orthogonal frequency division multiplexing (OFDM) as modulation format [34]. Hence, it is important to extend our previous results to the case of stochastic signals, particularly for the Gaussian case.

Keeping this in mind, in this section we assume a scenario where an event emits a Gaussian signal of the form $s_0(n) = \mathcal{N}(0, P_0)$. The event signal received at the k-th sensor can be represented as: $s_k(n) = \mathcal{N}(0, P_k)$, where $P_k = \frac{P_0}{d_{t,k}^2}$ with $\{d_{t,k}\}_{k=1}^K$ as the distance between k-th sensor and the place where the event is happened. To be more specific the k-th sensor receives the noisy observation $\{x_k\}_{k=1}^K$, where $x_k(n) = s_k(n) + w_k(n)$ in presence of the event signal and $x_k(n) = w_k(n)$ when the event signal is not present. Every sensor dispatches x_k to the fusion center. Hence, at time instant n the signal model at the fusion center based on the signals collected through K sensors is given as:

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{x}(n) = \mathbf{w}(n) \\ \mathcal{H}_1 &: \mathbf{x}(n) = \mathbf{s}(n) + \mathbf{w}(n) \end{aligned} \tag{3.29}$$

where $\mathbf{x}(n) = \begin{bmatrix} x_1(n) & x_2(n) & \cdots & x_K(n) \end{bmatrix}^T$ contains the noisy observations, $\mathbf{s}(n) = \begin{bmatrix} s_1(n) & s_2(n) & \cdots & s_K(n) \end{bmatrix}^T$ contains the received signals and vector $\mathbf{w}$ contains the samples of the additive noise. Based on these facts, we consider two cases for the detection problem (3.29) as follow.

1. **Case Study # 1**:- The observations of the local sensors are conditionally independent when conditioned on the hypothesis $\mathcal{H}_1$.
2. **Case Study # 2** :-When the observations are correlated in the case of hypothesis $\mathcal{H}_1$.

3.2.1 Case Study#1: Uncorrelated signal

In the current section we assume that the signal at different sensors are independent from each other, therefore, $\mathbf{s} \sim \mathcal{N}(0, \mathbf{\Sigma}_s)$ with

$$\mathbf{\Sigma}_s = \begin{bmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & P_K \end{bmatrix}. \tag{3.30}$$

The assumption of conditional independence of the sensor observations simplifies the problem and makes it more tractable. Later on in Section 3.2.2 we will relax this assumption. In the sequel, we assume that the noise at different sensors is also independent but nonidentical Gaussian distributed as: $\mathbf{w}(n) \sim \mathcal{N}(0, \mathbf{\Sigma}_w)$ with

$$\mathbf{\Sigma}_w = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_K^2 \end{bmatrix}. \tag{3.31}$$

We remark that, herein, we assume that the noise powers $\{\sigma_k^2\}_{k=1}^K$ are known.

3.2.1.1 GLRT

We already know that in the presence of unknown parameters $\{P_k\}_{k=1}^K$, the hypothesis testing cannot be solved by using the optimal Neyman-Pearson criteria [25]. Therefore, once again we need to use the GLRT. Since we consider independent noise as well as signal at different sensors, therefore, the parametrized log-LRT can be written as:

$$\begin{aligned} \mathcal{L}_G(\mathbf{X}) &= \sum_{k=1}^{K} \log \left\{ \frac{\max\limits_{P_k} f_{\mathbf{x}}(\mathbf{x}_k; P_k, \mathcal{H}_1)}{f_{\mathbf{x}}(\mathbf{x}_k; \mathcal{H}_0)} \right\} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma, \\ &= \sum_{k=1}^{K} \mathcal{L}_k(\mathbf{x}_k) \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma, \end{aligned} \tag{3.32}$$

where $\mathcal{L}_k(\mathbf{x}_k) \triangleq \log \left\{ \frac{\max\limits_{P_k} f_{\mathbf{x}}(\mathbf{x}_k; P_k, \mathcal{H}_1)}{f_{\mathbf{x}}(\mathbf{x}_k; \mathcal{H}_0)} \right\}$ and $\mathbf{x}_k \triangleq \begin{bmatrix} x_k(1) & x_k(2) & \cdots & x_k(N) \end{bmatrix}^T$ is vector that contains the number of samples collected at k-th sensor. Similarly, $\mathbf{X} \triangleq \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_K \end{bmatrix}$. By solving $\mathcal{L}_k(\mathbf{x}_k)$ in (3.32), we can get the expression as:

$$\begin{aligned} \mathcal{L}_k(\mathbf{x}_k) &= -\frac{N}{2} \log\left(\hat{P}_k + \sigma_k^2\right) - \frac{1}{2\left(\hat{P}_k + \sigma_k^2\right)} \sum_{n=1}^{N} x_k^2(n) \\ &\quad + \frac{N}{2} \log\left(\sigma_k^2\right) + \frac{1}{2\sigma_k^2} \sum_{n=1}^{N} x_k^2(n), \\ &= -\frac{N}{2} \log\left(\frac{1}{N} \sum_{n=1}^{N} x_k^2(n)\right) - \frac{N}{2} \\ &\quad + \frac{N}{2} \log\left(\sigma_k^2\right) + \frac{1}{2\sigma_k^2} \sum_{n=1}^{N} x_k^2(n), \\ &= -\frac{N}{2} \left\{ \log\left(\frac{\varepsilon_k}{\sigma_k^2}\right) + 1 - \frac{\varepsilon_k}{\sigma_k^2} \right\}, \end{aligned} \tag{3.33}$$

where in (3.33) we used the fact that the MLE of P_k as: $\hat{P}_k = \frac{1}{N}\sum_{n=1}^{N} x_k^2(n) - \sigma_k^2$ and $\varepsilon_k = \frac{1}{N}\sum_{n=1}^{N} x_k^2(n)$. By putting (3.33) in (3.32), we get the final expression for the GLRT as:

$$\mathcal{L}_G(\mathbf{X}) = \frac{N}{2}\sum_{k=1}^{K}\left\{\frac{\varepsilon_k}{\sigma_k^2} - \log\left(\frac{\varepsilon_k}{\sigma_k^2}\right) - 1\right\} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma \tag{3.34}$$

We can see that the only term that depends upon the observations is $\varepsilon_k = \frac{1}{N}\sum_{n=1}^{N} x_k^2(n)$. In this case it is also sufficient statistics. So instead sending the whole vector $\mathbf{x}_k$ the k-th sensor may send ε_k. Hence, we can implement the detection scheme (3.34) in the form of decentralized detection scheme. In such scheme the local sensors perform some local processing by calculating ε_k. After calculating ε_k each sensor dispatches the result to the fusion center. The fusion center makes the final decision by using (3.34) to decide whether the event signal is present or not. This two step method is summarized in algorithm 3.4.

Algorithm 3.4 Implementation of (3.34) as a decentralized detection with two steps method

1. Processing at the local sensors:
   - The k-th sensor receives N samples as
     $$\mathbf{x}_k \triangleq [x_k(0), x_k(1), \ldots, x_k(N-1)]^T$$
   - Having $\mathbf{x}_k$, sensor $\{k\}_{k=1}^{K}$ calculates $\varepsilon_k = \frac{1}{N}\sum_{n=1}^{N} x_k^2(n)$ and sends it to the fusion center.
2. Processing at the fusion center:
   - The fusion center finds $\frac{N}{2}\sum_{k=1}^{K}\left\{\frac{\varepsilon_k}{\sigma_k^2} - \log\left(\frac{\varepsilon_k}{\sigma_k^2}\right) - 1\right\} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma$.

3.2.1.2 Rank-reduction approach

In Section 3.1.3, we have discussed that the event appears at a random position and it will be surrounded by a given number of $\mathcal{K} \leq K$ active sensors (i.e. the ones which receive detectable power). Hence, we need to give more importance to those sensors that have strongest signal power. In order to do so, we can sort the sensors in descending order based on the the strength of the signal power. The process is similar to algorithm 3.3 by using the estimated $\widehat{\text{SNR}}s$ or $\left\{\frac{\varepsilon_k}{\sigma_k^2}\right\}_{k=1}^{K}$. Let $\mathbf{l}_G = [\mathcal{L}_1(\mathbf{x}_1), \mathcal{L}_2(\mathbf{x}_2), \cdots, \mathcal{L}_K(\mathbf{x}_K)]$ is a vector that contains the local GLRTs, sorted in the descending order based on the estimated $\widehat{\text{SNR}}_k = \frac{\varepsilon_k}{\sigma_k^2}$. Once we know the ordering of the sensor, the problem can be formulated as:

$$\mathcal{L}_{\max}(\mathbf{X}) = \max_{1\leq\mathcal{K}\leq K}\sum_{k=1}^{\mathcal{K}}\mathcal{L}_k(\mathbf{x}_k) \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma \tag{3.35}$$

However, as we discussed in Section 3.1.3.1 that the scheme in (3.35) will always implement the test statistic based on the maximum order K [30]. Therefore, $\mathcal{K}$ should first be estimated and then the GLRT should be conducted based on the estimated model $\hat{\mathcal{K}}$ by using MFLRT. Note that the implementation of the MFLRT is similar as in the case of section 3.1.3 but the GLR $L_{G_i}(\mathbf{x})$ should be replaced by:

$$\mathcal{L}_k(\mathbf{x}_k) = \frac{\varepsilon_k}{\sigma_k^2} - \log\left(\frac{\varepsilon_k}{\sigma_k^2}\right) - 1 \tag{3.36}$$

The implementation process of the model order based detection scheme is given in algorithm 3.5. We remark here that other model order selection techniques such as minimum description length (MDL) and Akaike information criterion (AIC) can used to implement the problem. Similarly, as it was in the case of Section 3.1.3, the detection scheme in algorithm 3.5 does not take into effect the fact the sensors are forming spatial cluster around the event.

Algorithm 3.5 Implementation process of the proposed detector

1. Sensor $\{k\}_{k=1}^{K}$ receives $\mathbf{x}_k \in \mathbb{R}^N$ and it calculates $\varepsilon_k = \frac{1}{N}\sum_{n=1}^{N} x_k^2(n)$.
2. Every sensor sends ε_k to the fusion center thus fusion center has $\mathbf{e} = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_K \end{bmatrix}$.
3. With assumption of known noise power, we can find the SNR vector as: $\mathbf{e}_\sigma = \begin{bmatrix} \frac{\varepsilon_1}{\sigma_1^2} & \frac{\varepsilon_2}{\sigma_2^2} & \cdots & \frac{\varepsilon_K}{\sigma_K^2} \end{bmatrix}$.
4. Order the SNR's in vector $\mathbf{e}_\sigma$ in the descending order, we get $\tilde{\mathbf{e}}_\sigma$.
5. Based on the ordered $\tilde{\mathbf{e}}_\sigma$, find each element of
$$\mathbf{l}_G = \begin{bmatrix} \mathcal{L}_1(\mathbf{x}_1) & \mathcal{L}_2(\mathbf{x}_2) & \cdots & \mathcal{L}_K(\mathbf{x}_K) \end{bmatrix}.$$
6. Based on $\mathbf{l}_G$ the fusion center calculates the following steps:
   - $\mathbf{t} \leftarrow \langle\rangle$
   - while $L \leqq K$ do
     - Find $T_x = \sum_{l=1}^{L} \mathcal{L}_l(\mathbf{x}_k)$.
     - Obtain $f_i = f(T_x)$ as given in (3.22)
     - Push f_i onto $K \times 1$ vector $\mathbf{t}$
     - $i = i + 1$
     - end while
   - $\mathbf{t} = \begin{bmatrix} f_1, & f_2, & \cdots, & f_K \end{bmatrix}$
   - $T_{\text{MFLRT}}(x) = \max(\mathbf{t}) \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma$ as in (3.21) and the index corresponding to the maximum is $\hat{\mathcal{K}}$.

3.2.1.3 Numerical example

In this section, we present results based on computer simulations to illustrate the performance of the GLRT $\mathcal{L}_G(\mathbf{X})$ in (3.34) and detection scheme based on the MFLRT that is presented in presented in the algorithm 3.5. For the analysis to be conducted herein, experiments are performed with a total number of $K = 20$ sensor nodes, which are randomly distributed in the sensor field of size 12 sq-meters. In order to analyze the performance of detectors we use the average SNR of all receivers as: $\bar{\kappa} = -8$dB.

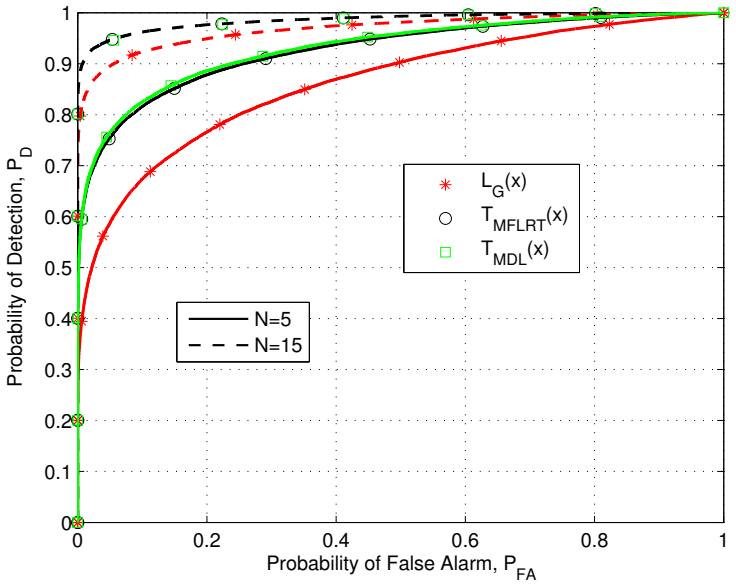


Figure 3.8: ROC curves: Comparison of the MFLRT with the traditional GLRT for unknown stochastic signal: Average SNR $\bar{\kappa} = -8$dB

In this example the ROC curves are obtained for two different values of N as: $N = \{5, 15\}$. The results are plotted in Fig. 3.8 that show that the selection of the sensors with the help of model order selection indeed improves the performance of the GLRT. We can also see that for this specific detection problem, the performance of MFLRT is similar to the GLRT implementation with the help minimum description length (MDL). Furthermore, it can be seen that though performance of the detectors improves but the behavior of the the detection schemes remains the same for two different values of N. From the results it can be concluded that as it was in the case of deterministic signal, when the effect of target signal is non-uniform over the widely spread sensor field, then the rank-reduced multi-sensor detection scheme clearly out-performs the traditional GLRT.

3.2.2 Case Study #2: Correlated signal

The assumption of conditional independence is not generally valid in practical situations where the proximity of the sensors to one another will result in correlated observations [14] [17]. This spatial correlation is a feature that can be used for detection since the noise

processes at different sensors can be safely assumed statistically independent. Hence, in this section we formulate the detection schemes for the sensor network that exploit the spatial covariance matrix of the received observations.

When the signal is present then at time instant n the received signal vector at the fusion center (based on the signals collected through K sensors) is given as :$\mathbf{x}(n) = \mathbf{s}(n) + \mathbf{w}(n)$. In the absence of signal we have $\mathbf{x}(n) = \mathbf{w}(n)$. Therefore, the signal model can be represented as:

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{x} \sim \mathcal{N}(0, \mathbf{\Sigma}_0) \\ \mathcal{H}_1 &: \mathbf{x} \sim \mathcal{N}(0, \mathbf{\Sigma}_1) \end{aligned} \tag{3.37}$$

Contrary to the signal model in Section 3.2.1, herein, we consider to relax the assumption of the spatial independence. We assume that the signal at different sensors are spatially correlated, therefore, $\mathbf{s} \sim \mathcal{N}(0, \mathbf{\Sigma}_s)$ and $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_s + \mathbf{\Sigma}_0$. The noise at different sensors is assumed to be independent but nonidentical Gaussian distributed as: $\mathbf{w} \sim \mathcal{N}(0, \mathbf{\Sigma}_0)$ with $\mathbf{\Sigma}_0 \triangleq \mathbf{\Sigma}_w = \text{diag}\left(\begin{bmatrix} \sigma_1^2 & \sigma_2^2 & \cdots & \sigma_K^2 \end{bmatrix}\right)$. Based on these facts we can easily infer that the detection problem (3.37) is expected to focus on the exploitation of the inner structure of the covariance matrix of the vector-valued $\{\mathbf{x}(n), n = 1, 2, ...N\}$.

3.2.2.1 GLRT

The detection problem in (3.37) not only decides based on the presence or absence of the signal energy but also takes into account whether the off-diagonal elements of the sample covariance matrix are zero or not. In order to test this phenomena a number of tests have been presented in the field of statistics. Comprehensive details about multivariate tests of this nature can be found in [35, 36]. The studies presented in these references are based the multivariate statistics that deal with multivariate hypothesis testing for covariance matrices. In the case when the covariance matrices are unknown, the most popular scheme that quantifies such hypothesis is based on different variations of GLRT schemes. For the signal model (3.37), the GLRT decides in favor of $\mathcal{H}_1$ only if

$$\Lambda_G(\mathbf{X}) = \frac{\max\limits_{\mathbf{\Sigma}_1} f_{\mathbf{x}}(\mathbf{X}, \mathbf{\Sigma}_1)}{\max\limits_{\mathbf{\Sigma}_0} f_{\mathbf{x}}(\mathbf{X}, \mathbf{\Sigma}_0)} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \gamma, \tag{3.38}$$

where $f_{\mathbf{x}}(\mathbf{X}, \mathbf{\Sigma}_1)$ and $f_{\mathbf{x}}(\mathbf{X}, \mathbf{\Sigma}_1)$ are the likelihood function under hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively, that can be represented as:

$$f_{\mathbf{x}}(\mathbf{X}, \mathbf{\Sigma}_1) = \frac{1}{(2\pi |\mathbf{\Sigma}_1|)^{\frac{N}{2}}} \exp\left\{ -\frac{1}{2} \sum_{n=1}^{N} \mathbf{x}^T(n) \mathbf{\Sigma}_1^{-1} \mathbf{x}(n) \right\} \tag{3.39}$$

and

$$f_{\mathbf{x}}(\mathbf{X}, \mathbf{\Sigma}_0) = \frac{1}{(2\pi |\mathbf{\Sigma}_0|)^{\frac{N}{2}}} \exp\left\{ -\frac{1}{2} \sum_{n=1}^{N} \mathbf{x}^T(n) \mathbf{\Sigma}_0^{-1} \mathbf{x}(n) \right\}, \tag{3.40}$$

where $|.|$ represents determinant of a matrix. The solution of (3.38) involves finding MLE of $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_0$, that can be found by taking differentiation of $\log f\left(\mathbf{X}, \mathbf{\Sigma}_1\right)$ and $\log f\left(\mathbf{X}, \mathbf{\Sigma}_0\right)$, respectively. The results are then equated to zero to achieve the MLE $\hat{\mathbf{\Sigma}}_1 = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}\left(n\right)\mathbf{x}^T\left(n\right)$ is a sample covariance matrix. As we assume that the N snapshots of the observation vector $\mathbf{x}\left(n\right)$ are statistically independent in time, then the maximum likelihood estimator is equal to the sample covariance matrix $\hat{\mathbf{\Sigma}}_1$. Similarly, $\hat{\mathbf{\Sigma}}_0 = \text{diag}\left(\frac{1}{N}\sum_{n=1}^{N}\mathbf{x}\left(n\right)\mathbf{x}^T\left(n\right)\right) = \hat{\mathbf{\Sigma}}_1 \odot \mathbf{I}_K$. Based on these results the solution of (3.38) becomes

$$\Lambda_G\left(\mathbf{X}\right) = \frac{\left|\hat{\mathbf{\Sigma}}_1\right|}{\left|\hat{\mathbf{\Sigma}}_1 \odot \mathbf{I}_K\right|} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \gamma. \tag{3.41}$$

In practice, the GLRT is used based on the assumption that the sample size is large while the sample dimension is small as when $N \ggg K$, it is an optimal detector. However, in case of the small sample support the GLRT indeed suffers from the ill-conditioning and singularity issue of the sample covariance matrix. To cope with this problem, we present some alternative tests that are ad-hoc in the sense that they are not derived based on the traditional detection theory or hypothesis testing rules. These test are rather derived based on the underlying principle that the statistical covariance matrices of signal and noise are generally different. This difference is used in the proposed methods to differentiate the signal component from background noise.

3.2.2.2 Covariance absolute value detector

We have argued in the previous section that one way to mitigate the repercussions of a singular sample covariance matrix, is to design detectors that avoid those matrix operations that are vulnerable to the singularity of this matrix for instance, the determinant. Taking into account this fact, the Covariance Absolute Value (CAV) detector proposed in [11] can be used to avoid such problems. The CAV detector is a ratio between the sum of elements of the sample covariance matrix and the sum of diagonal elements of that matrix as:

$$\Lambda_{\text{CAV}}\left(\mathbf{X}\right) = \frac{\sum_i^K \sum_j^K |r_{ij}|}{\sum_i^K \text{abs}\,|r_{ii}|} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \gamma. \tag{3.42}$$

Here $|.|$ is for absolute value, whereas, r_{ij} is the i, j^{th} element of sample covariance matrix. The detector in (3.42) does not need any prior information of the signal, the channel nor the noise power. The test statistic in (3.42) is robust against the high dimensionality and it does not assume any prior information about the signal and noise distribution.

3.2.2.3 Detector based on normalized Frobenius norm

The Frobenius norm of an arbitrary matrix $\mathbf{A}$ can be represented as:

$$\|\mathbf{A}\|_F = \sqrt{\sum_i \sum_j [\mathbf{A}]_{i,j}^2}. \tag{3.43}$$

Based on the Frobenius norm of the sample covariance matrix, we adopt the following test to compare it with the asymptotically optimal GLRT (3.41). The test can be represented as:

$$\Lambda_{\text{Frob}}(\mathbf{X}) = \frac{\left\|\hat{\mathbf{\Sigma}}_1 - \hat{\mathbf{\Sigma}}_0\right\|_F}{\left\|\hat{\mathbf{\Sigma}}_1\right\|_F} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \gamma. \tag{3.44}$$

The detector (3.44) is also robust in front of the high dimensionality and small sample support, and it does not assume any prior information about the signal and noise distribution.

3.2.2.4 Numerical example

In this section, the above mentioned covariance based detection schemes are compared with each other by using numerical simulations. In order to analyze the performance of these detectors we use $K = 20$ sensors uniformly distributed in the field. We use the average SNR of all receivers similar to the Section 3.1.2.3. With these parameters we plot ROC curves in Fig. 3.9 and Fig. 3.10. In order to see the effect of sample size N, in Fig. 3.9 we plot the ROC curves for $\Lambda_{\text{CAV}}(\mathbf{X})$, $\Lambda_{\text{Frob}}(\mathbf{X})$ and $\Lambda_G(\mathbf{X})$ while considering two cases. In the first case we use $N = 20$ and in the second case we use $N = 60$. Similarly, in Fig. 3.10 we repeat the same experiment with ROC curves by introducing the effect noise power uncertainty. In the literature, there are quite a few methods to add uncertainty to noise powers. Herein, in this experiment we model the noise power uncertainty by generating the noise power at the k^{th} sensor as $\sigma_k^2 \sim \mathcal{U}\left(\frac{\sigma_{n,k}^2}{\alpha_{nu}}, \alpha_{nu}\sigma_{n,k}^2\right)$, where $\alpha_{nu} \geq 1$ and $\alpha_{nu} = 1$ means no noise uncertainty [15]. In this plot we also compare the performance $\Lambda_{\text{CAV}}(\mathbf{X})$, $\Lambda_{\text{Frob}}(\mathbf{X})$ and $\Lambda_G(\mathbf{X})$ with the energy detector

$$\Lambda_{\text{Eng}}(\mathbf{X}) = \frac{1}{NK} \sum_{k=1}^{K} \sum_{n=1}^{N} x_k^2(n). \tag{3.45}$$

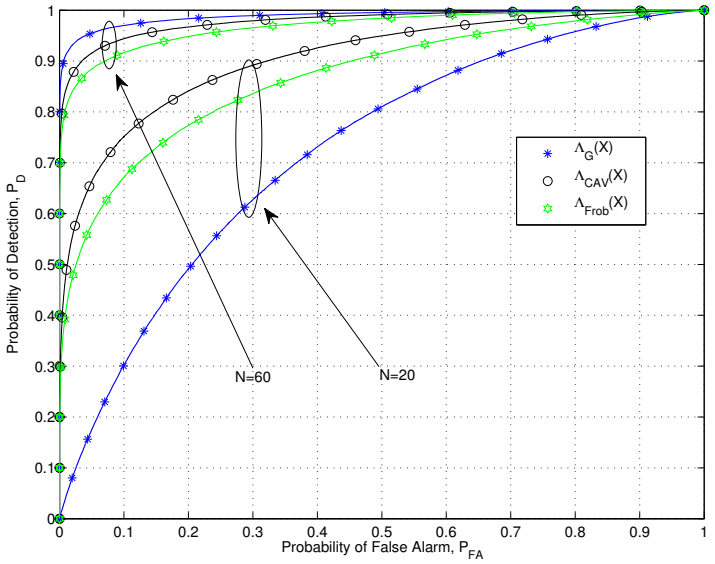


Figure 3.9: ROC curves: Comparison of the GLRT with CAV and Frobenius detector for total number of sensors $K = 20$ and average SNR $\bar{\kappa} = -15$dB.

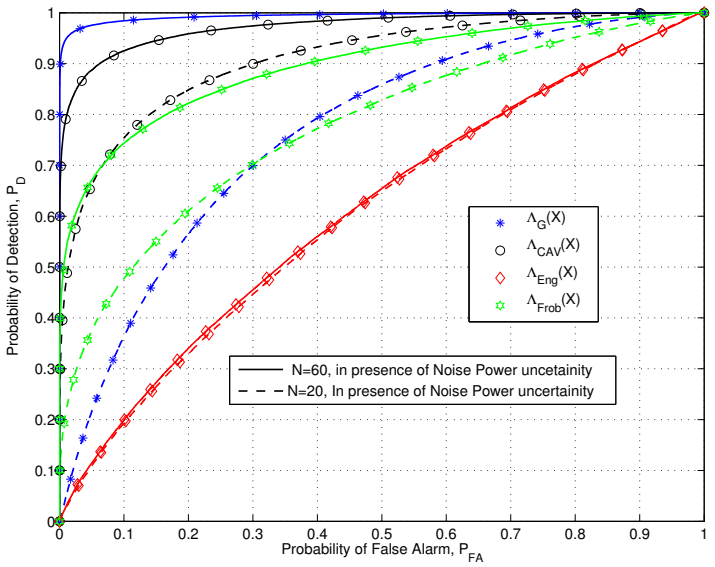


Figure 3.10: ROC curves: Comparison of the GLRT, CAV and Frobenius detector with the energy detector in the presence of uncertain noise, total number of sensors $K = 20$ and average SNR $\bar{\kappa} = -15$dB.

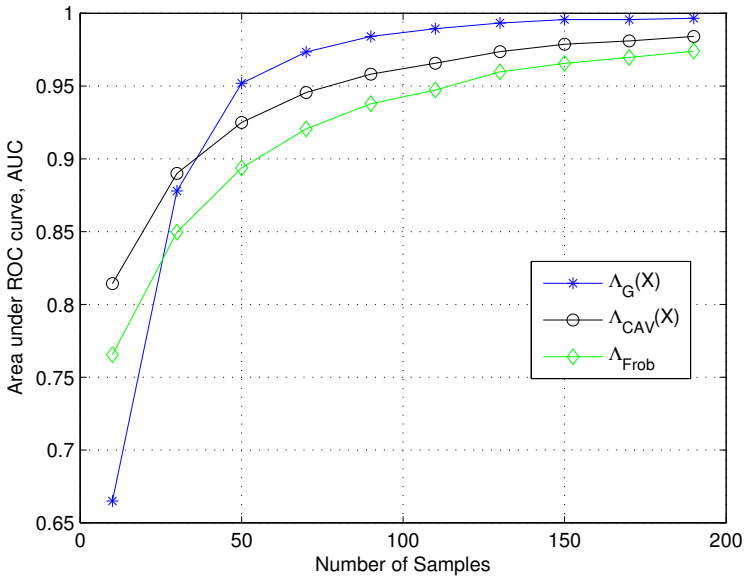


Figure 3.11: AUC curves: Comparison of the GLRT with CAV and Frobenius detector for changing sample size (N), total number of sensors $K = 20$ and average SNR $\bar{\kappa} = -15$dB.

The results in these two experiments clearly reveal that the GLRT $\Lambda_G(\mathbf{X})$ outperforms, $\Lambda_{\text{CAV}}(\mathbf{X})$ and $\Lambda_{\text{Frob}}(\mathbf{X})$ only if the sample size is much higher than the number of sensors $K = 20$. Similarly, results in 3.10 confirm that energy detector cannot perform in the presence of severe noise power uncertainty even we increase the N. This is due the effect called "SNRwall", which is the SNR below which robust detection is impossible for the given detector [15]. On the contrary, it can be seen that in case of the other three detection schemes increasing N results in increase in the performance in the presence of noise power uncertainty. Hence they are robust against the noise uncertainty compared to the well known energy detector. It is because these schemes use correlation as a side information jointly with the signal energy present in the received observations. In order to further analyze the effects of the sample size N, in 3.11 we plot the AUC plots. The results in this figure clearly show and confirm that the GLRT performs very poorly when the sample support is small. However, with increasing the sample support N the GLRT starts to outperform the other detectors. Hence, we conclude that for the GLRT to perform good, the sample support should be larger than the dimension of the signal vector. In other words N should be larger than the number of sensors K, when this is not the case, the GLRT degenerates due to ill-conditioned sample covariance matrix. Herein, we would like to remark that in the proceeding chapters, we will address the issue of ill-conditioned sample covariance matrix and small sample support, comprehensively. Apart from this, in Section 5 and 6, we will also analyze the effects of noise power uncertainty in detail.

CHAPTER 4

Exploitation of Spatial Proximity through Spatial Signatures

In the previous chapter we discussed the problem of finding a set of relevant sensors, in such a way that a rank reduced approach could be implemented, thus leading to an SNR gain. In the present chapter we are moving one step further by actually exploiting the proximity between neighboring sensors, since this a-priori information is also expected to provide a more robust and improved and detection performance. Therefore, the main focus of this chapter is to study this possible improvement in the detection performance of the multi-sensor detection. In order to achieve this improvement, we assume that the positions of the sensor nodes are known to the fusion center. This assumption is quite reasonable in practical scenarios, because the nodes can easily learn its position information at the time of initial calibration or with the help of the Global Positioning System (GPS) i.e., GPS-enabled wireless sensor network [37][38]. By considering known positions, we can assume that the sensor-to-sensor distances are known to the fusion center. We exploit this information (sensor-to-sensor distances) as a side information to be incorporated in the detection rule with the aim of improving the performance. Keeping this mind, we divide the content of this chapter into two categories.

1. Exploitation of spatial correlation between neighboring sensors using spatial signatures.
2. Multi-sensor quickest detection by exploiting spatial proximity.

4.1 Multi-sensor detection based on spatial signatures

We know from the previous discussions that a WSN typically consists of a large number of inexpensive sensor nodes that are distributed over a large area. In largely deployed WSN, the observations are often correlated in the space domain due to then spatial proximity among sensors [17]. In spite of this, most existing detection techniques are based on energy detection, thus ignoring this important cross-sensor correlation information between closely located sensors [39]. This feature is relevant because the region where the events typically happen within a dense WSN usually spans across an area comprising just a subset of all the

sensor nodes. Those sensors far away from the event are typically unable to receive the signal emitted from the event due to limited sensing ranges. In contrast, sensors closer to the event will often be closely spaced, thus forming a cluster with highly correlated observations [17]. The presence of this structure within the received samples at the fusion center may be used to further improve the event detection performance. Previously, there have been some attempts to incorporate correlated measurements into the formulation of signal detection problems. However, many of these studies focus on the discrimination between correlated and independent observations [35, Ch. 9-10]. However, these detectors typically focus on detecting the presence of correlated data, as a possible indication that an event may be present within the data. They do not focus, instead, on exploiting the actual correlation structure that impinges onto the sensor field when an emitting target is present.

In the previous chapter we exploited the fact that in order to detect efficiently the emergence of an event, it is important to use only the observations that contain useful information and to suppress noise from the non-affected sensors. We have shown that the performance of the signal detection technique can be improved by identifying the useful observation samples among the set of all received measurements. Keeping this in mind, herein, we move one step ahead to exploit the fact that the affected sensors will be located close to the event as well as close to each other in the form of a spatial cluster. This spatial proximity present between the affected sensors can be used to further improve the detection performance of the multi-sensor detection schemes. Hence, in this Section a novel detector is presented. Not only it exploits the selection of the useful set of samples in order to reject noise, but it also takes advantage of the signal correlation occurring with an emitting target is present, based on the novel concept of *signatures*. The concept of signatures would be somehow equivalent to *steering vectors* in the field of array signal processing. Signatures are adopted herein as a way to capture the structure of spatially correlated measurements between neighboring sensor nodes. Each sensor node will have a signature vector representing its physical-layer connectivity with the rest of the sensors in the field. As it will be shown later on, this approach is found to significantly improve the detection performance compared to traditional approaches.

4.1.1 Signal model

In order to describe the signal model, we assume the case where sensors are doing some processing locally and they send summary of the observations to the fusion center. In other words, after receiving the N samples of $x_k(n) = A_k + w_k(n)$, the k-th, sensor calculates the mean as: $\bar{x}_k = \frac{1}{N}\sum_{n=1}^{N} x_k(n)$ and dispatches it to the fusion center. The fusion center collects the measurements of K sensors in a vector $\mathbf{x} = [\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_K]$ and the signal model at the fusion center becomes

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{x} = \mathbf{w}, \\ \mathcal{H}_1 &: \mathbf{x} = \mathbf{s} + \mathbf{w}, \end{aligned} \tag{4.1}$$

where $\mathbf{s} = [A_1, A_2, \cdots, A_K]$, $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_w)$ with $\mathbf{\Sigma}_w = \text{diag}\left[\sigma_1^2/N, \sigma_2^2/N, \cdots, \sigma_K^2/N\right]$. Hence, under $\mathcal{H}_0$ $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_w)$ and under $\mathcal{H}_1$ $\mathbf{x} \sim \mathcal{N}(\mathbf{s}, \mathbf{\Sigma}_w)$. The detection problem (4.1) is nothing but to differentiate $\mathbf{s} \neq \mathbf{0}$ from $\mathbf{s} = \mathbf{0}$. The vector $\mathbf{s}$ contains the received unknown deterministic amplitudes of received at the K sensors. In this vector the samples corresponding to active sensors will be non-zeros and rest of the samples will have extremely weak values that will hardly contribute constructively in the detection performance. To be more specific we are indeed in front of situation where vector $\mathbf{s}$ is a sparse vector. Hence, in the following we take into effect this sparsity to model the detection scheme.

4.1.1.1 Signal model with spatial signatures

The active sensors are not only located close to the event but also located close to each other and thus result in a spatial cluster. Due to this neigbouring or proximity, there will exist some correlation structure. The main focus of this paper is to exploit this correlation structure and design a signal detector at the fusion center based on the principle of GLRT. In order to do so, a structure signal model is proposed on the concept of *signatures*. For the case of the i-th sensor, its signature is a vector that contains the attenuation terms to all the K sensors, as if the signal source was located at the i-th sensor position. Thus, the k-th signature is a $(K \times 1)$ vector $\mathbf{h}_k$ as follows,

$$\mathbf{h}_k \triangleq \left[h(d_{1,k}), \ldots, h(d_{k-1,k}), 1, h(d_{k+1,k}), \ldots, h(d_{K,k})\right]^T. \tag{4.2}$$

where $h(d_{k,j}) = e^{-\beta d_{k,j}}$, takes into account the attenuation loss due to the distance between k-th and the j-th sensor locations, with β as the known path loss exponent. We assume herein that the fusion center has complete knowledge of the positions of the sensors in the network. Therefore, all of the K signatures present in the matrix $\mathbf{H} \triangleq [\mathbf{h}_1, \mathbf{h}_2 \ldots, \mathbf{h}_K]$ are known to the fusion center. Matrix $\mathbf{H}$ is assumed to be a full-rank $K \times K$ matrix with K signatures as columns and we call it "*signature matrix*". This spatial structure is reflected in the signatures matrix $\mathbf{H}$, and that the signal vector $\mathbf{s}$ must be a linear combination of these signatures as $\mathbf{s} = \mathbf{Ha}$. With such a spatial knowledge the following so called structured signal model applies,

$$\mathbf{x} = \mathbf{Ha} + \mathbf{w} \tag{4.3}$$

where $\mathbf{a}$ is the $K \times 1$ vector containing weights of each signature onto the received signal. In other words, these weights a_k , $k = 1, 2, \cdots\cdots, K$ can be understood as a kind of virtual amplitudes that when linearly combined with the sensors signatures, they reproduce the signal strength field captured by the whole WSN in the presence of an emitting source. That is to say, the elements within $\mathbf{a}$ quantify the importance of each of the sensors signatures in the reconstruction of the signal field emitted by the event. Therefore, by selecting the largest weights, we are actually choosing the most relevant sensors on the basis of their physical proximity to the event. The more the sensor is located closer to the event, the more its signature vector will be aligned with $\mathbf{s}$, and thus the larger the weight assigned to this signature. By using this linear combination of signatures in (4.3), we are taking into

account both the distances between neighboring sensors and the location of the sensors with respect to the event, thus fully exploiting the spatial information contained within the received signals.

In practice the event will typically appear in a random position within the sensor field, and of course it will be surrounded by a given number of $\mathcal{K}$ active sensors (i.e. the ones which receive non-zero amplitudes). It means that the important signatures are those which are related to the signals from active sensors and rest of the $K - \mathcal{K}$ signatures could be ignored. Thus we are in-front of detection problem where it is convenient to use rank-reduced version of the signal model (4.3) as we have been doing in the previous chapter. As we need to select the $\mathcal{K}$ signatures of active sensors with their corresponding unknown weights, $\mathcal{K}$ needs to be estimated by model order selection. The $\mathcal{K}$ signatures are then stacked into matrix $\mathbf{H}_{\mathcal{K}}$, which is a reduced version of $\mathbf{H}$. Similarly, the unknown weights are stacked in $\mathbf{a}_{\mathcal{K}}$, the reduced version of vector $\mathbf{a}$.

4.1.2 Structured GLRT based on spatial signatures

Based on the results obtained in the previous chapter, in the case of spatial signal model (4.3), assuming model order $\mathcal{K} = K$, the GLR can be written as:

$$\begin{aligned} L_G(\mathbf{x}) &= \mathbf{x}^T \boldsymbol{\Sigma}_w^{-1} \mathbf{P}_H \mathbf{x} - \frac{1}{2} \mathbf{x}^T \mathbf{P}_H \boldsymbol{\Sigma}_w^{-1} \mathbf{P}_H \mathbf{x} \\ &= \frac{1}{2} \mathbf{x}^T \mathbf{P}_H \boldsymbol{\Sigma}_w^{-1} \mathbf{P}_H \mathbf{x} \end{aligned} \tag{4.4}$$

where $\mathbf{P}_H = \mathbf{H} \left(\mathbf{H}^H \mathbf{H} \right)^{-1} \mathbf{H}^T$ is the projection matrix onto the space spanned by all of the K signatures. The correlation-aware detector (4.4) matches the received measurements with the expected correlation pattern implicitly contained within the projection matrix onto the signatures space, $\mathbf{P}_H$. In the problem under study, only the signatures related to the $\mathcal{K}$ active sensors are of importance and signatures of sensors that receive only noise should be discarded. That is why we use reduced-rank version of signal model (4.3) to find GLR and we use MFLRT to cope with issue of unknown $\mathcal{K}$. To do so we first estimate the $K \times 1$ vector of signature weights $\mathbf{a}$ as $\hat{\mathbf{a}} = \left(\mathbf{H}^H \mathbf{H} \right)^{-1} \mathbf{H}^H \mathbf{x}$ then we sort the magnitudes of elements in $\hat{\mathbf{a}}$ in descending order. After that signatures in $\mathbf{H}$ are ordered according to the sorted $\hat{\mathbf{a}}$. On the basis of ordered $\mathbf{H}$ and ordered $\hat{\mathbf{a}}$, MFLRT is thus implemented. For MFLRT the GLR, after assuming order size equal to i, can be formulated as [25]:

$$L_{G_i}(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{P}_{H_i} \boldsymbol{\Sigma}_w^{-1} \mathbf{P}_{H_i} \mathbf{x} \tag{4.5}$$

where $\mathbf{P}_{H_i} = \mathbf{H}_i \left(\mathbf{H}_i^H \mathbf{H}_i \right)^{-1} \mathbf{H}_i^T$ is the reduced-rank projection matrix onto the subspace spanned by the i selected signatures. Here the resulting detector matches the received measurements with the expected correlation pattern implicitly contained within the projection matrix onto the subspace of active signatures, $\mathbf{P}_{\mathbf{H}_i}$. The energy of the resulting projection is the one which is then compared to a threshold for determining whether the expected signal

Algorithm 4.1 Implementation process of the proposed detector

1. The received observations from the sensor field are stacked in vector $\mathbf{x}$.
2. Using already known full signature matrix $\mathbf{H}$, we define $\mathbf{x} = \mathbf{H}\mathbf{a} + \mathbf{w}$.
3. Estimate the the full virtual weights vector $\mathbf{a}$ as $\hat{\mathbf{a}} = \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{x}$
4. Take the absolute values$|\hat{\mathbf{a}}|$ and sort $|\hat{\mathbf{a}}|$ in descending order resulting in $\tilde{\mathbf{a}} = \text{sort}\left(|\hat{\mathbf{a}}|\right)$
5. Shuffle the signature vectors in $\mathbf{H}$ according to the sorting of $\tilde{\mathbf{a}}$ as $\tilde{\mathbf{H}} = \text{sort}\left(\mathbf{H}\right)$.
6. Implement MFLRT by using (4.5) as:
   - $\mathbf{t} \leftarrow \langle\rangle$
   - while $i \leqq K$ do
     - Find $\mathbf{H}_i$=$\tilde{\mathbf{H}}(1:i)$
     - Calculate $\mathbf{P}_{\mathbf{H}_i} = \mathbf{H}_i\left(\mathbf{H}_i^T\mathbf{H}_i\right)^{-1}\mathbf{H}_i^T$
     - Calculate$\ln L_{G_i}\left(\mathbf{x}\right) = \frac{1}{2}\mathbf{x}^T\mathbf{P}_{H_i}\mathbf{\Sigma}_w^{-1}\mathbf{P}_{H_i}\mathbf{x}$
     - Obtain $f_i = f\left(\bar{L}_{G_i}(x)\right)$ as given in section 3.1.3.
     - Push f_i onto $K \times 1$ vector $\mathbf{t}$
     - $i = i + 1$
     - end while
   - $\mathbf{t} = \left[\ f_1, \quad f_2, \quad \cdots, \quad f_K\ \right]$
   - $T_{\text{MFLRT}}(x) = \max\left(\mathbf{t}\right) \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma$ and the index corresponding to the maximum is $\hat{\mathcal{K}}$.

structure was contained on the data or not. Moreover, let us define $\mathbf{x}_P \triangleq \mathbf{P}_{H_i}\mathbf{x}$ then we can write (4.5) as:

$$L_{G_i}(\mathbf{x}) = \frac{1}{2}\mathbf{x}_P^T\mathbf{\Sigma}_w^{-1}\mathbf{x}_P \tag{4.6}$$

where $\mathbf{x}_P$ is the projected version of the received observation vector $\mathbf{x}$ to the subspace spanned by the signatures of the active sensors. This will indeed result in an SNR gain due to the projection of the observation vector onto a reduced dimensionality subspace. The proposed detector with the reduced-rank spatial signal model aims improvement in the detection performance of the GLRT by including the spatial information in the form of signatures. The implementation process of the proposed detector is summarized in Algorithm 4.1

4.1.3 Performance analysis

We find the performance of the proposed detector in terms of probability of detection (P_D) and probability of false alarm (P_{FA}). For the model order $\mathcal{K}$ and using properties of the projection matrix $\mathbf{P}_{H_\mathcal{K}}$, we can write the proposed detector (4.5) using the results for linear

detectors in [25, Ch. 13]:

$$T_{\mathcal{K}}(\mathbf{x}) = \mathbf{x}^T \left(\mathbf{H}_{\mathcal{K}}^{\#}\right)^T \mathbf{H}_{\mathcal{K}}^T \boldsymbol{\Sigma}_w^{-1} \mathbf{H}_{\mathcal{K}} \mathbf{H}_{\mathcal{K}}^{\#} \mathbf{x} \tag{4.7}$$

where $\mathbf{H}_{\mathcal{K}}^{\#} = \left(\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}}\right)^{-1} \mathbf{H}_{\mathcal{K}}^T$ is a $(\mathcal{K} \times K)$ matrix corresponding to the Moore-Penrose pseudoinverse matrix of $\mathbf{H}_{\mathcal{K}}$. The expression in (4.7) can be further rearranged by defining the $(\mathcal{K} \times 1)$ vector, $\mathbf{z} \triangleq \mathbf{H}_{\mathcal{K}}^{\#} \mathbf{x}$. In this way, the detector can be expressed as:

$$T_{\mathcal{K}}(\mathbf{x}) = \mathbf{z}^T \mathbf{H}_{\mathcal{K}}^T \boldsymbol{\Sigma}_w^{-1} \mathbf{H}_{\mathcal{K}} \mathbf{z} \tag{4.8}$$

Now $\mathbf{z}$ has Gaussian distribution as $\mathbf{z} \sim \mathcal{N}\left(\mathbf{H}_{\mathcal{K}}^{\#}\mathbf{H}\mathbf{a}, \left(\mathbf{H}_{\mathcal{K}}^T \boldsymbol{\Sigma}_w^{-1} \mathbf{H}_{\mathcal{K}}\right)^{-1}\right)$. Let us define $\boldsymbol{\mu}_z = \mathbf{H}_{\mathcal{K}}^{\#}\mathbf{H}\mathbf{a}$ and $\mathbf{C}_z = \left(\mathbf{H}_{\mathcal{K}}^T \boldsymbol{\Sigma}_w^{-1} \mathbf{H}_{\mathcal{K}}\right)^{-1}$ then the statistics of the proposed detector are:

$$T_{\mathcal{K}}(\mathbf{x}) = \mathbf{z}^T \mathbf{C}_z^{-1} \mathbf{z} \sim \begin{cases} \chi^2_{\mathcal{K}_{\text{sp}}} & \text{for } \mathcal{H}_0 \\ \chi'^2_{\mathcal{K}_{\text{sp}}}(\lambda_{sp}) & \text{for } \mathcal{H}_1 \end{cases}. \tag{4.9}$$

with $\lambda_{sp} = \mathbf{a}^T \mathbf{H}^T \mathbf{P}_{\mathbf{H}_{\mathcal{K}_{\text{sp}}}} \boldsymbol{\Sigma}_w^{-1} \mathbf{H}\mathbf{a} = \mathbf{s}^T \mathbf{P}_{\mathbf{H}_{\mathcal{K}_{\text{sp}}}} \boldsymbol{\Sigma}_w^{-1} \mathbf{s}$ is the corresponding non-centrality parameter. We have clearly indicated the estimated order by $\mathcal{K}_{\text{sp}}$, indicating that the rank-reduced dimension $\mathcal{K}_{\text{sp}}$ has been estimated by performing a model order selection technique onto the full estimate of $\mathbf{a}$. Indeed, $\mathcal{K}_{\text{sp}}$ with the subindex "sp" is related to the fact that spatial information is being exploited. With the estimated model order $\mathcal{K}_{sp}$, the expression for the probability of false alarm and probability of detection becomes:

$$\begin{aligned} P_{FA}(\mathcal{K}_{\text{sp}}) &= \mathsf{Q}_{\chi^2_{\mathcal{K}_{\text{sp}}}}(\gamma), \\ P_D(\mathcal{K}_{\text{sp}}) &= \mathsf{Q}_{\chi'^2_{\mathcal{K}_{\text{sp}}}(\lambda_{sp})}(\gamma). \end{aligned} \tag{4.10}$$

One of the problems that appears in the calculation of the performance is that it depends upon the model order $\mathcal{K}_{\text{sp}}$, which is estimated by model order selection techniques based on noisy input measurements. As a result, the estimated model order becomes a random variable, which makes it very difficult to analyze the performance in close-form. This problem is also aggravated by the fact that the distribution of this random variable is unknown and that is why it is not so trivial to find close-form analytical expression for the performance. In [40], the asymptotic distribution of the model order for Akaike's information criterion (AIC) is obtained but the paper does not demonstrate the statistical optimality for practical cases. On the other hand, in the case of the MFLRT there are no such results available for the distribution of the estimated model order. The only way to proceed is to perform an empirical analysis using histograms, which can be used to obtain the estimated distribution of $\mathcal{K}_{\text{sp}}$, herein refereed as $f_{\mathcal{K}_{\text{sp}}}(l)$. Using $f_{\mathcal{K}_{\text{sp}}}(l)$ the performance of MFLRT with (4.5) in term of average probability of false alarm $\overline{P}_{FA}$ and average probability of detection $\overline{P}_D$ can be

expressed as:

$$\begin{aligned}\overline{P}_{FA} &\triangleq \mathrm{E}_{\mathcal{K}_{\mathrm{sp}}}\left[P_{FA}(\mathcal{K}_{\mathrm{sp}})\right] = \sum_{l=1}^{\mathcal{K}_{\mathrm{msx}}} P_{FA}(l) f_{\mathcal{K}_{\mathrm{sp}}}(l), \\ \overline{P}_{D} &\triangleq \mathrm{E}_{\mathcal{K}_{\mathrm{sp}}}\left[P_{D}(\mathcal{K}_{\mathrm{sp}})\right] = \sum_{l=1}^{\mathcal{K}_{msx}} P_{D}(l) f_{\mathcal{K}_{\mathrm{sp}}}(l).\end{aligned} \tag{4.11}$$

The performance of the proposed detector is also compared with the detector that uses no spatial information that has been discussed in the previous chapter. The PDF of the detector without spatial information is also chi-square and the expression for the probability of false alarm and probability of detection is given as:

$$\begin{aligned}P_{FA}(\mathcal{K}_{\mathrm{ns}}) &= \mathsf{Q}_{\chi^2_{\mathcal{K}_{\mathrm{ns}}}}(\gamma), \\ P_{D}(\mathcal{K}_{\mathrm{ns}}) &= \mathsf{Q}_{\chi'^2_{\mathcal{K}_{\mathrm{ns}}}(\lambda_{ns})}(\gamma),\end{aligned} \tag{4.12}$$

where the non-centrality parameter $\lambda_{ns} = \mathbf{s}^H_{\mathcal{K}_{\mathrm{ns}}} \mathbf{\Sigma}_w^{-1} \mathbf{s}_{\mathcal{K}_{\mathrm{ns}}}$. In the same way as explained before, the performance in term of average probability of false alarm $\overline{P}_{FA}$ and average probability of detection $\overline{P}_D$ is:

$$\begin{aligned}\overline{P}_{FA} &\triangleq \mathrm{E}_{\mathcal{K}_{\mathrm{ns}}}\left[P_{FA}\left(\mathcal{K}_{\mathrm{ns}}\right)\right] = \sum_{l=1}^{\mathcal{K}_{max}} P_{FA}(l) f_{\mathcal{K}_{\mathrm{ns}}}(l), \\ \overline{P}_{D} &\triangleq \mathrm{E}_{\mathcal{K}_{\mathrm{ns}}}\left[P_{D}(\mathcal{K}_{\mathrm{ns}})\right] = \sum_{l=1}^{\mathcal{K}_{max}} P_{D}(l) f_{\mathcal{K}_{\mathrm{ns}}}(l),\end{aligned} \tag{4.13}$$

where $f_{\mathcal{K}_{\mathrm{ns}}}(l)$ is the empirical probability distribution of estimated model order for a fixed value of signal to noise ratio.

4.1.4 Numerical example

In this section, we present the results of computer simulations to illustrate the performance of the detection schemes presented in the above sections. The performance of the proposed detector is also compared with the detection techniques based on the signal model with no spatial information (presented in the previous chapter). Experiments are performed with a total number of $K = 20$ sensor nodes, which are randomly distributed in the sensor field of size 12 sq-meters. This simplified scenario is depicted in Fig.4.1. To make the scenario resemble to more practical cases, we also fix the sensing range for the sensor nodes to $R = 3$ meters. Looking at Fig.4.1, the event is represented with the circle. The sensors that find the event within their sensing ranges are denoted by "∇", while "$\square$" indicate those sensors that receive zero amplitude from the event. On the basis of this scenario, we analyze the proposed detector with the help of the receiver operating characteristics (ROC) curves and the curves showing probability of detection (PD) vs SNR(dB). For the assessment of the detectors being considered in this chapter, the SNRs at the different sensors are allocated randomly. In order to do so, we use the average SNR of all antennas as: $\bar{\kappa} = \frac{1}{K}\sum_{i=1}^{L} \kappa_i$ and the SNR of k^{th}

sensors is $\kappa_i = \zeta^{i-1}\kappa_{\min}$ where $\kappa_{\min}$ is the minimum SNR among those of the K sensors. For a given average SNR $\bar{\kappa}$ and SNR gap ζ, we can generate the SNRs of the sensors, though the match of the SNRs to the sensors can be random.

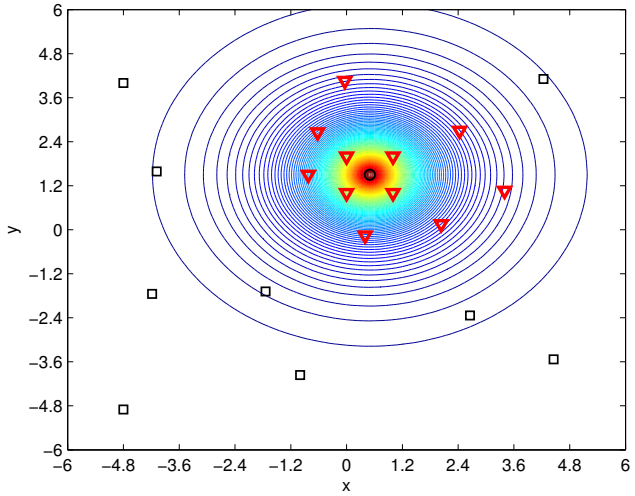


Figure 4.1: Field scenario showing a specific topology of sensors.

In Fig.4.2 as well in Fig.4.3 subscript "ns" indicates the unstructured MFLRT detector that uses no spatial information and subscript "sp" represents the proposed structured MFLRT detector which uses spatial information. It can be seen in Fig.4.2 that the proposed detector has better ROC characteristic compared to the detector that uses no spatial information. Similarly, we can also see that the performance with sensor selection improves with and without spatial information compare to GLRT detector. The plot for GLRT is represented "*" and blue color in the figure. In Fig.4.3, the detection schemes are further analyzed with the help of average probability of detection ($\overline{P}_D$) plotted against different SNR values in dB. These plots also show that detector with spatial information is superior to the detector which is using no spatial information. Hence, if the sensor-to-sensor distances are known, then in order to get improved performance this a-priori information should be used.

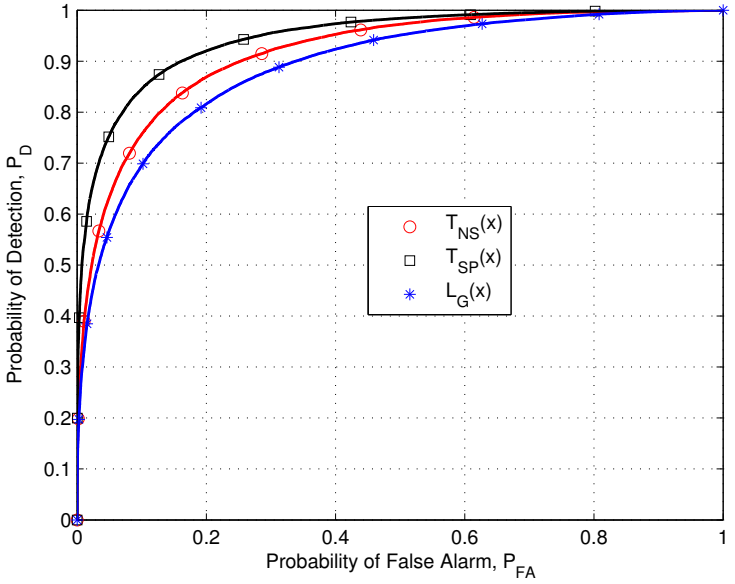


Figure 4.2: ROC Curves: Comparison of the structured MFLRT with the unstructured MFLRT and the traditional GLRT.

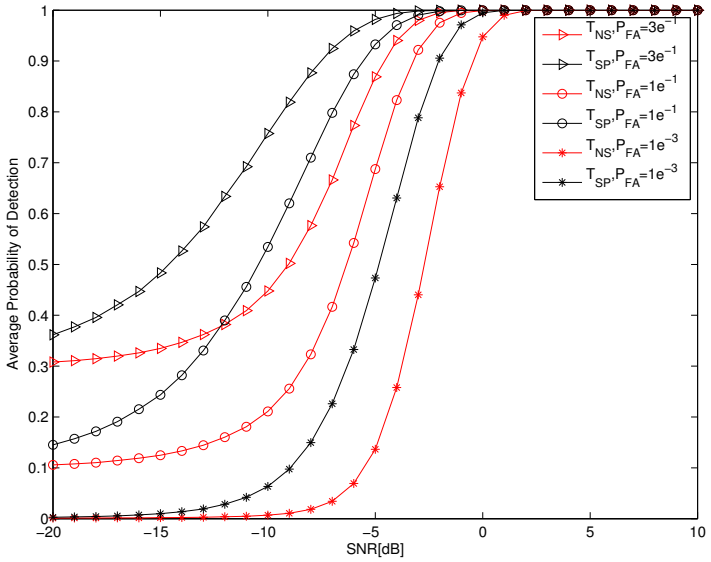


Figure 4.3: Probability of detection (PD) vs. SNR[dB]: Comparison of the structured MFLRT with the unstructured MFLRT.

4.2 Multi-sensor quickest detection

In Section 4.1, we proposed a new GLRT based detector for a centralized detection system. The aim was to achieve an improvement in the detection performance by exploiting correlation among neighboring sensor nodes. Prior information on the basis of sensor's positions has been incorporated by proposing a signal model based on signature matrix, which captures

the correlation among different sensors. On the contrary, here in this section we exploit the prior information of the sensor positions in the context of *quickest detection* [41, 42, 43].

Most of the detection approaches discussed above and in the previous chapters are block-based in the sense that the sensor nodes take a block of samples to decide on the activity state of an event. The main goal of these block-based detection schemes is to maximize the detection probability subject to constraints on the false alarm probability. However, while detecting abrupt changes due to an event, apart from maximizing the probability of detection, the WSN should be able to detect the changes as quickly as possible. For example in cognitive radio, the secondary users need to detect and quit the frequency band as quickly as possible if the corresponding primary radio emerges [41]. In such dynamic applications, the multi-sensor should continuously observe the region to detect the event quickly as possible based on the possible change in the received observations. Therefore, a new detection framework is required that focuses on minimization of the detection delay for a certain level of false alarm.

The above discussion leads us to the concept of multi-sensor sequential change detection [44, 45] and a variant known as quickest detection as discussed in [41, 42, 43]. In sequential change detection methods, the well-known Page's cumulative sum (CUSUM) algorithm has been shown to be optimal in the sense of minimizing the detection delay while maintaining an acceptable level of false alarm [41, 44]. Moreover, the CUSUM algorithm has been previously adopted for the collaborative spectrum sensing by assuming that the observations from multiple sensors are independent and identically distributed both in time as well as in space [41, 45, 46]. To the best of our knowledge the previous work on multi-sensor quickest detection either ignores the presence of the mutual correlation in the received observations or consider it as a deleterious effect [47]. However, this spatial correlation is a feature that can be used for detection as already mentioned in the previous section, since the noise processes at different nodes can be safely assumed statistically independent.

Taking into account the above discussions, we formulate the Multivariate Cumulative SUM(MCUSUM) algorithm by exploiting the spatial correlation structure. Consequently, the proposed scheme also considers the fact that due to the limited coverage of the signal emitted from a weak intensity event, just a subset of sensors (in the form of a spatial cluster) are able to receive power levels enough for reliable detection [48]. While considering the spatial correlation, it is required that the information regarding the spatially structured covariance matrix should be available. The MCUSUM is a recursive method that operates with single sample, hence, the estimation of the unknown covariance matrix is not feasible [44]. In order to overcome this hurdle we propose to model the spatial covariance structure by using the a-priori spatial information based on the known sensor-to-sensor distances. Similarly, to observe the change point and ignore noisy observations from the unaffected sensors (i.e. sensors that receive the event signal with negligible power), the proposed detection method selects observations from the set of relevant sensors only. Finally, to asses the performance of the proposed scheme, we analyze the expression of the mean detection delay by adopting the asymptotic results given in [49]. Simulations results are included to verify these analytical expressions.

4.2.1 Problem formulation and assumptions

We consider an infrastructure-based WSN where in order to observe the event, K sensors continuously measure their total received power (in dBm)[1]. They normalize the measurements by subtracting the respective mean noise powers (in dBm, assumed to be known) as in [50], and relay the normalized measurements $\{x_i\}_{i=1}^{K}$ (in dBm) to the fusion center. Under $\mathcal{H}_0$, the power at the output of the i-th local energy detector is simply the sum of noise and interference powers. Consequently, due to the effect of the interferences, the total noise power can be modeled as a independent log-normally distributed random variable with variance $\sigma_{n_i}^2$ (assumed to be known) [50]. Hence, under the null hypothesis $\{x_i\}_{i=1}^{K}$ (in dBm) are independent Gaussian distributed random variables with zero means and variances $\left\{\sigma_{n_i}^2\right\}_{i=1}^{K}$. Under $\mathcal{H}_1$, the total power at i-th sensor is the sum of the event signal power (modeled with log-normal random variable having variance $\sigma_{s_i}^2$, that quantifies shadow fading) and the noise power[2]. Given this, the total received power under $\mathcal{H}_1$ is equal to the sum of two independent log-normally distributed random variables and the sum can be approximated as log-normal random variable [51]. Based on this information, given $\mathcal{H}_1$, the measurements $\{x_i\}_{i=1}^{K}$ (in dBm) are assumed to be Gaussian distributed with means $\mu_i \triangleq \mathrm{E}\left[10\log_{10}\left(1+\mathrm{SNR}_i\right)\right]$ and covariances $\sigma_{s_i}\sigma_{s_j}\rho_{i,j}+\sigma_{n_i}^2$ where $\rho_{i,j}$ quantifies correlation of shadowing between sensor i and j [50]. Having said this, at time instant n, the observation vector received at the fusion center can be represented as $\mathbf{x}\left(n\right) \triangleq \left[\begin{array}{cccc} x_1\left(n\right) & x_2\left(n\right) & \cdots & x_K\left(n\right) \end{array}\right]^T$ and the signal model can be written as:

$$\begin{aligned} \mathcal{H}_0 :&\mathbf{x}\left(n\right) = \mathbf{w}\left(n\right), \\ \mathcal{H}_1 :&\mathbf{x}\left(n\right) = \mathbf{s}(n) + \mathbf{w}\left(n\right), \end{aligned} \tag{4.14}$$

where $\mathbf{s}(n) \sim \mathcal{N}\left(\boldsymbol{\mu}_s\left(n\right), \boldsymbol{\Sigma}_s\right)$ with i-th element of the mean vector $\boldsymbol{\mu}_s\left(n\right)$ is μ_i and i,j-th element of $\boldsymbol{\Sigma}_s$ is $\sigma_{s_i}\sigma_{s_j}\rho_{i,j}$. Similarly, $\mathbf{w}\left(n\right) \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_0\right)$ contains noise with $\boldsymbol{\Sigma}_0 = \mathrm{diag}\left\{\begin{array}{cccc} \sigma_{n_1}^2 & \sigma_{n_2}^2 & \cdots & \sigma_{n_K}^2 \end{array}\right\}$. Hence, under $\mathcal{H}_0$, $\mathbf{x}\left(n\right) \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_0\right)$ and under $\mathcal{H}_1$, $\mathbf{x}\left(n\right) \sim \mathcal{N}\left(\boldsymbol{\mu}_s(n), \boldsymbol{\Sigma}_1\right)$ with $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_s + \boldsymbol{\Sigma}_0$.

4.2.1.1 Selection of active sensors

We assume that the event will appear at an unknown random position and the signal power emitted by the event decays isotropically and only a subset of sensors are affected due to it [48]. This observation allows us to distinguish between the so-called *active* (affected) and *inactive* (unaffected) sensors, respectively. For an improved detection performance intuition suggests that the detection rule should rely on the observations of the $\mathcal{K} \leq K$ active sensors, thus discarding the observations from the rest of inactive sensors. Keeping this in mind we

[1] In order to exploit the experimentally obtained shadowing correlation model based on a-priori sensor-to-sensor distances, instead of random amplitudes, herein, we use signal model based on power in dBs.

[2] We are implicitly assuming that the noise includes interference and modeled as a independent log-normally distributed.

are in front of a detection problem where it is convenient to use a rank-reduced version of the signal model in (6.4). To do so, we need to select the observations of the $\mathcal{K}$ most relevant sensors. Herein, to simplify the mathematical exposition, we assume that the received signal vector $\mathbf{x}(n)$ is already ordered such that the first $\mathcal{K}$ samples of $\mathbf{x}(n)$ correspond to the $\mathcal{K}$ active sensors and the remaining $\mathcal{K}^c$ samples consist of only noise correspond to the inactive sensors [30]. We believe that in practice, this assumption can be easily relaxed. For example the signal samples (in dBm) in $\mathbf{x}(n)$ can be sorted by ordering the vector $\mathbf{x}(n)$ in descending order. Further details about such a process can be found in [47, 30]. Similarly, another way to relax the assumption is that the local sensors weight themselves locally based on their received powers and the fusion center sorts the observations based the provided weights. Adding further, the sensors can even perform local selection by comparing their received powers with some predefined threshold and the qualified sensors declare themselves as active sensors. Keeping these facts into considerations the rank-reduced signal model can be written as:

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{x}(n) = \mathbf{w}(n), \\ \mathcal{H}_1 &: \mathbf{x}(n) = \mathbf{s}_{\mathcal{K}}(n) + \mathbf{w}(n), \end{aligned} \tag{4.15}$$

where $\mathbf{s}_{\mathcal{K}}(n) = \mathbf{T}_{\mathcal{K}}\mathbf{s}(n)$ is an ordered vector that has the first $\mathcal{K}$ non-zero signal elements corresponding to the received signals at the $\mathcal{K}$ active sensors and

$$\mathbf{T}_{\mathcal{K}} = \begin{bmatrix} \mathbf{I}_{\mathcal{K}} & \mathbf{0}_{\mathcal{K}\times(K-\mathcal{K})} \\ \mathbf{0}_{(K-\mathcal{K})\times\mathcal{K}} & \mathbf{0}_{(K-\mathcal{K})\times(K-\mathcal{K})} \end{bmatrix} \tag{4.16}$$

Hence, we have $\mathbf{s}_{\mathcal{K}}(n) = \mathcal{N}\left(\mathbf{T}_{\mathcal{K}}\boldsymbol{\mu}_s(n), \mathbf{T}_{\mathcal{K}}\boldsymbol{\Sigma}_s\mathbf{T}_{\mathcal{K}}^T\right)$. For simplicity and without loss of generality, we assume variables in $\mathbf{w}(n)$ are I.I.D, therefore $\sigma_{n_i}^2 = \sigma_{n_j}^2 = \sigma_0^2$, so under $\mathcal{H}_0$ $\mathbf{x}(n) = \mathcal{N}\left(\mathbf{0}, \sigma_0^2\mathbf{I}\right)$. Then under $\mathcal{H}_1$ we have $\mathbf{x}(n) \sim \mathcal{N}\left(\mathbf{T}_{\mathcal{K}}\boldsymbol{\mu}_s(n), \mathbf{T}_{\mathcal{K}}\boldsymbol{\Sigma}_s\mathbf{T}_{\mathcal{K}} + \sigma_0^2\mathbf{I}\right)$. The detection problem (4.15) says that due to the emergence of an event signal, changes occur both in the mean vector and covariance structure. Therefore, the quickest detection technique should be capable of simultaneously monitoring the mean as well as the covariance matrix. Furthermore, the proposed signal model also takes into account the rank-reduction of the signal covariance matrix by introducing $\mathbf{T}_{\mathcal{K}}$ in (4.15). Therefore, in addition to the spatial proximity information, the rank-reduced signal model (4.15) will indeed allow us to benefit from an equivalent signal to noise ratio gain due to selection of the active sensors.

4.2.2 Proposed quickest detection approach

In this section we present the proposed quickest detection scheme. In order to proceed we define a time $n \triangleq N$ as the the discrete-time at which the change (ideally, it occurs at time N_d) is detected. If $N > N_d$ then the detection delay is $\delta = N - N_d$. Similarly, if $N < N_d$, a false alarm has occurred with the average time between the false alarm is being $T_{FA} = \mathrm{E}_0[N]$, where E_0 denotes the average time taken before the change occurs (i.e. event

emerges). Similarly, the worst case detection delay is defined as:

$$\mathrm{E}_1[N] = \underset{N_d \geq 1}{sup} \left(esssup\, \mathrm{E}\left[N - N_d \mid N \geq N_d, \{\mathbf{x}(n)\}_{n=1}^{N_d} \right] \right), \tag{4.17}$$

where $esssup$ denotes essential supremum. Under the Lorden's criterion, the objective is to find the stopping rule that minimizes the worst-case delay while maintaining the time between the false alarms larger than a certain threshold, $T_{FA} \geq \lambda$ [41]. An alternative approach is to minimize the average detection delay,

$$T_\Delta = \sup_{N_d \geq 1} \mathrm{E}_{N_d}[N - N_d \mid N \geq N_d], \tag{4.18}$$

which is asymptotically equivalent to the worst case delay [41]. Now, the algorithm that finds the minimum T_Δ in this problem is the Page's CUSUM Test (PCT) [41, 46, 45]. The stopping time of the PCT is determined as:

$$T(q) = \inf\{n : \mathcal{C}_n \geq q\}, \tag{4.19}$$

where q is a pre-determined threshold [44]. The cumulative statistic $\mathcal{C}_n$ can be recursively calculated for $n \geq 1$ as follow:

$$\mathcal{C}_n = \max(\mathcal{C}_{n-1}, 0) + \mathcal{L}_n(\mathbf{x}), \tag{4.20}$$

where $\mathcal{C}_0 = 0$, $\mathcal{L}_n(\mathbf{x}) \triangleq \log_e \frac{f_1(\mathbf{x}(n))}{f_0(\mathbf{x}(n))}$ and $f_h(\mathbf{x}(n))$ with $h = 0, 1$ are likelihood functions under the alternate hypotheses. Observing (4.20) we can see that the CUSUM algorithm finds the first n for which $C_n \geq q > 0$.

In the proposed MCUSUM based on the signal model (4.15), not only N_d but also K_a is unknown, hence, it involves a nested likelihood ratio. Taking into effect these consideration, the Log-likelihood ratio statistic to be used in (4.20) is given as:

$$\mathcal{L}_n(\mathbf{x}) = \max_{1 \leq L \leq K} \sum_{l=1}^{K_a} \mathrm{Log} \left\{ \mathcal{D}_l \frac{\exp\left(-\frac{1}{2}\tilde{\mathbf{x}}^T(n)\mathbf{\Phi}_l\tilde{\mathbf{x}}(n)\right)}{\exp\left(-\frac{1}{2}\mathbf{x}^T(n)\mathbf{\Sigma}_0^{-1}\mathbf{x}(n)\right)} \right\}, \tag{4.21}$$

where $\mathcal{D}_l = \sqrt{\det\mathbf{\Sigma}_0}/\sqrt{\det\mathbf{\Sigma}_{l,1}}$, $\mathbf{\Phi}_l = \mathbf{\Sigma}_{l,1}^{-1}$ with $\mathbf{\Sigma}_{l,1} = \mathbf{T}_l\mathbf{\Sigma}_s\mathbf{T}_l^T + \sigma_0^2\mathbf{I}$, and $\tilde{\mathbf{x}}(n) = \mathbf{x}(n) - \boldsymbol{\mu}_{l,s}(n)$ with $\boldsymbol{\mu}_{l,s}(n) = \mathbf{T}_l\boldsymbol{\mu}_s(n)$.

Now the recursive test (4.20) needs input in the form of $\mathcal{L}_n(\mathbf{x})$ (4.21) at every time instant n, the estimation of the unknown covariance matrix $\mathbf{\Sigma}_{l,1}$ is not feasible. Therefore, we present a mechanism to find the unknown covariance matrix $\mathbf{\Sigma}_{l,1}$ by assuming locations of nodes are a-priori known [50]. In order to do so, first we assume the shadowing effects at different sensors have similar variances as $\sigma_{s_i}^2 = \sigma_{s_j}^2 = \sigma_s^2$, then we can write $\mathbf{T}_l\mathbf{\Sigma}_s\mathbf{T}_l = \sigma_s^2\mathbf{T}_l\mathbf{\Psi}\mathbf{T}_l^T$ with $\mathbf{\Psi}$ as the correlation matrix having elements $\rho_{i,j}$ for $i, j = 1, 2, ...K$. Similarly, the shadowing parameter σ_s^2 is an experimentally obtained parameter that is dependent on the propagation environment and it is assumed to be known [50, 52]. Moreover, in

(4.21) the matrix $\mathbf{\Phi}_l$ can be considered as a concentration matrix or a precision matrix. The elements of the precision matrix can be interpreted in terms of partial correlations and partial variances. Partial correlation measures the degree of association between two random variables. In our problem, in order to exploit this association we model the correlation between two receivers with the help of the correlation model in [52] by exploiting the fact that nodes locations are known. Hence, the correlation between i^{th} and j^{th} sensor can be modeled as $\rho_{i,j} = e^{-ad_{i,j}}$ [50, 52]. It is mentioned in [52], that for urban areas, at 1700 MHz $a \approx 0.12$ and for suburban areas at 900MHz $a \approx 0.002$. Moreover, in (4.21) we can write $\tilde{\mathbf{x}}^T(n)\mathbf{\Phi}_l\tilde{\mathbf{x}}(n) = \text{Tr}[\mathbf{\Phi}_l\tilde{\mathbf{x}}(n)\tilde{\mathbf{x}}^T(n)]$. Keeping all these into effect, we can deduce that the statistic (4.21) indeed matches the expected covariance matrix $\mathbf{\Sigma}_{l,1}$ with the received covariance matrix $\tilde{\mathbf{x}}(n)\tilde{\mathbf{x}}^T(n)$ as an additional detection metric jointly with the signal energy. In other words, it takes into account both correlation and the reliability of information by selecting the active sensors.

Finally, the proposed MCUSUM statistic is achieved by putting $\mathcal{L}_n(\mathbf{x})$ from (4.21) in (4.20). The proposed quickest detection can be interpreted as a two dimensional MCUSUM as it operates with the "max" operation both in the space and time. At every received observation vector it selects the active sensors and exploits the spatial structure of the received observations by matching the received covariance matrix with the modeled one, as explained in the previous paragraph.

4.2.3 Performance analysis

The performance of the quickest detector is normally defined by T_{FA} and T_Δ [45, 49]. For a particular detector larger T_{FA} and smaller T_Δ means better detection performance. Consequently, to explain the advantages of sensor selection and exploitation of spatial structure, herein, we assume the case where the active sensors have been perfectly identified. Keeping this in mind, based on the well known result by Lorden [49], the expressions for the asymptotic T_{FA} and T_Δ are given as [45]:

$$T_{FA} = \text{E}_0\left[N\right] \geq e^q \text{ as } q \to \infty \tag{4.22}$$

and

$$T_\Delta = \text{E}_1\left[N\right] \sim \frac{q}{\mathcal{D}\left(f_{\mathcal{H}_1} \parallel f_{\mathcal{H}_0}\right)} \text{ as } q \to \infty, \tag{4.23}$$

where for a specific L, $f_{\mathcal{H}_1}(\mathbf{x})$ and $f_{\mathcal{H}_0}(\mathbf{x})$ are the likelihood functions of the signal model (4.15) under $\mathcal{H}_1$ and under $\mathcal{H}_0$, respectively. Similarly, $\mathcal{D}\left(f_{\mathcal{H}_1} \parallel f_{\mathcal{H}_0}\right) = \text{E}_1\left[\mathcal{L}_n(\mathbf{x})\right]$ is the Kullback-Leibler Divergence (KLD) between the two densities. The KLD between the two densities is defined as:

$$\mathcal{D}\left(f_{\mathcal{H}_1} \parallel f_{\mathcal{H}_0}\right) = \int f_{\mathcal{H}_1}(\mathbf{x}) \log \frac{f_{\mathcal{H}_1}(\mathbf{x})}{f_{\mathcal{H}_0}(\mathbf{x})} d\mathbf{x}. \tag{4.24}$$

Note that $\mathrm{E}_1\left[\mathcal{L}_n\left(\mathbf{x}\right)\right]$ is always positive and $\mathrm{E}_0\left[\mathcal{L}_n\left(\mathbf{x}\right)\right]$ is always negative [44]. The asymptotic results from Lorden's approach can be interpreted as for a constant T_{FA} or q, T_Δ is inversely proportional to the KLD. It means that the higher is the value of $\mathcal{D}\left(f_{\mathcal{H}_1} \parallel f_{\mathcal{H}_0}\right)$, the smaller will be T_Δ. In order to analyze $\mathcal{D}\left(f_{\mathcal{H}_1} \parallel f_{\mathcal{H}_0}\right)$ of the proposed scheme, let us suppose the exact number of the active sensors is $\mathcal{K}'$, then the solution to the integral (4.24) is [53, Chapter 10]:

$$\begin{aligned}\mathcal{D}\left(f_{\mathcal{H}_1} \parallel f_{\mathcal{H}_0}\right) = \frac{1}{2}\log\frac{\det\left(\mathbf{\Sigma}_0\right)}{\det\left(\mathbf{\Sigma}_{\mathcal{K}',1}\right)} + \frac{1}{2}\mathrm{Tr}\left(\mathbf{\Sigma}_{\mathcal{K}',1}\mathbf{\Sigma}_0^{-1}\right) \\ -\frac{1}{2}\mathrm{Tr}\left(\mathbf{\Sigma}_{\mathcal{K}',1}\mathbf{\Sigma}_{\mathcal{K},1}^{-1}\right) + \frac{1}{2}\mathrm{Tr}\left(\mathbf{\Sigma}_0^{-1}\boldsymbol{\mu}_{\mathcal{K}',s}(n)\boldsymbol{\mu}_{\mathcal{K}',s}^T(n)\right),\end{aligned} \tag{4.25}$$

where $\mathbf{\Sigma}_{\mathcal{K}',1} = \mathbf{T}_{\mathcal{K}'}\mathbf{\Sigma}_s\mathbf{T}_{\mathcal{K}'}^T + \sigma_0^2\mathbf{I}$ and $\mathbf{\Sigma}_{\mathcal{K},1} = \mathbf{T}_{\mathcal{K}}\mathbf{\Sigma}_s\mathbf{T}_{\mathcal{K}}^T + \sigma_0^2\mathbf{I}$. Similarly, based on the assumption that the number of active sensors $\mathcal{K}'$ is perfectly known, we have the following two cases.

Case 1. $\mathcal{K} = \mathcal{K}'$: Number of active sensors are perfectly selected,

$$\begin{aligned}\mathcal{D}\left(f_{\mathcal{H}_1} \parallel f_{\mathcal{H}_0}\right) = \frac{1}{2}\log\frac{\det\left(\mathbf{\Sigma}_0\right)}{\det\left(\mathbf{\Sigma}_{\mathcal{K}',1}\right)} - \frac{K}{2} \\ + \frac{\mathcal{K}'}{2}\frac{\sigma_s^2}{\sigma_0^2} + \left(\frac{1}{2\sigma_0^2}\sum_{i=1}^{\mathcal{K}'}\mu_{i,s}^2\right).\end{aligned} \tag{4.26}$$

Case 2. $\mathcal{K} = K$, all sensors (active and inactive) are selected

$$\begin{aligned}\mathcal{D}^{'}\left(f_{\mathcal{H}_1} \parallel f_{\mathcal{H}_0}\right) =& \frac{1}{2}\log\frac{\det\left(\mathbf{\Sigma}_0\right)}{\det\left(\mathbf{\Sigma}_{K,1}\right)} + \frac{\mathcal{K}'}{2}\frac{\sigma_s^2}{\sigma_0^2} \\ &- \frac{1}{2}\mathrm{Tr}\left(\mathbf{\Sigma}_{\mathcal{K}',1}\mathbf{\Sigma}_{K,1}^{-1}\right) + \left(\frac{1}{2\sigma_0^2}\sum_{i=1}^{\mathcal{K}'}\mu_{i,s}^2\right),\end{aligned} \tag{4.27}$$

where $\mathbf{\Sigma}_{K,1} = \mathbf{\Sigma}_s + \sigma_0^2\mathbf{I}$.

Lemma 4.1. *$\mathcal{D}^{'}\left(f_{\mathcal{H}_1} \parallel f_{\mathcal{H}_0}\right) \leq \mathcal{D}\left(f_{\mathcal{H}_1} \parallel f_{\mathcal{H}_0}\right)$ and equality prevails when $\mathcal{K}' = K$.*

Proof. To compare (4.26) and (4.27) and analyze the impact of only sensor selection, we suppose $\mathbf{\Sigma}_s = \sigma_s^2\mathbf{I}$ in both of the equations. Therefore, $\mathrm{Tr}\left(\left(\mathbf{T}_{\mathcal{K}'}\mathbf{\Sigma}_s\mathbf{T}_{\mathcal{K}'}^T + \sigma_0^2\mathbf{I}\right)\mathbf{\Sigma}_{K,1}^{-1}\right) = \mathcal{K}'\sigma_s^2/\left(\sigma_s^2+\sigma_0^2\right) + K\sigma_0^2/\left(\sigma_s^2+\sigma_0^2\right)$, $\log\left\{\det\left(\mathbf{\Sigma}_0\right)/\det\left(\mathbf{T}_{K_a'}\mathbf{\Sigma}_s\mathbf{T}_{\mathcal{K}'}^T + \sigma_0^2\mathbf{I}\right)\right\} = \mathcal{K}'\log\left(\frac{\sigma_0^2}{\sigma_s^2+\sigma_0^2}\right)$ and $\log\left\{\det\left(\mathbf{\Sigma}_0\right)/\det\left(\mathbf{\Sigma}_{K,1}\right)\right\} = K\log\left(\frac{\sigma_0^2}{\sigma_s^2+\sigma_0^2}\right)$. Putting these values in (4.26) and (4.27) we can easily deduce that $\mathcal{D}^{'}\left(f_{\mathcal{H}_1} \parallel f_{\mathcal{H}_0}\right) \leq \mathcal{D}\left(f_{\mathcal{H}_1} \parallel f_{\mathcal{H}_0}\right)$ for the case $K \geq \mathcal{K}'$. □

By using the results of this lemma and the Lorden's asymptotic approximation (4.23), it is clear that by using sensor selection jointly with a-prior spatial information indeed enhances the performance of the MCUSUM.

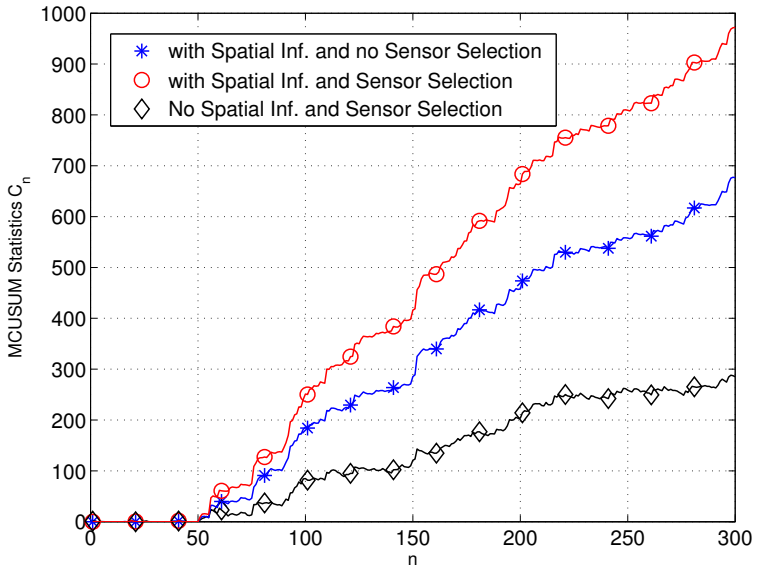


Figure 4.4: Typical behaviour of the MCUSUM statistic $\mathcal{C}_n$.

4.2.4 Numerical results

For the analysis we use $K = 30$ sensors uniformly distributed in a square field. The event appears at a random location that emits some power. To comply with the discussions in Section 3, we set the coverage of the event's signal such that an unknown subset of K sensors receive non-zero powers from the event. Consequently, we fix $\sigma_0^2 = 2$, $\sigma_s^2 = 1.5$. The behaviour of the adaptive multivariate CUSUM statistic $\mathcal{C}_n$ for the three schemes shown in Figure 4.4. These results show the change in the log-likelihood ratios for the three schemes before and after the emergence of the event. For all three cases the curves of MCUSUM statistics show sudden increase right after the change in the distributions occurs at $n = 50$. We can also see that for a certain value of q, the algorithm that uses jointly spatial information and sensor selection has the quickest response to the change created by the event's signal.

In Figure 4.5, with the help of monte-carlo simulations we present the normalized histograms of the detection delays in responding to the change occurs. These histograms also shows that the proposed sequential detection scheme based on the spatial information performs better.

The results in Figure 4.6 confirm the above numerical results where we plot the average detection delay versus the changing value of the threshold q. As asymptotically T_{FA} is exponential function of the threshold q, increasing q also means increase in T_{FA}. In Figure 4.6, we can see that increase in q (or T_{FA}) results in increase in the average detection delay times of the three schemes. Similarly, it can be seen that the increase in the response time of the proposed detection scheme is less than the scheme that assume no spatial structure. Moreover, using observations of only active sensors further increase the performance.

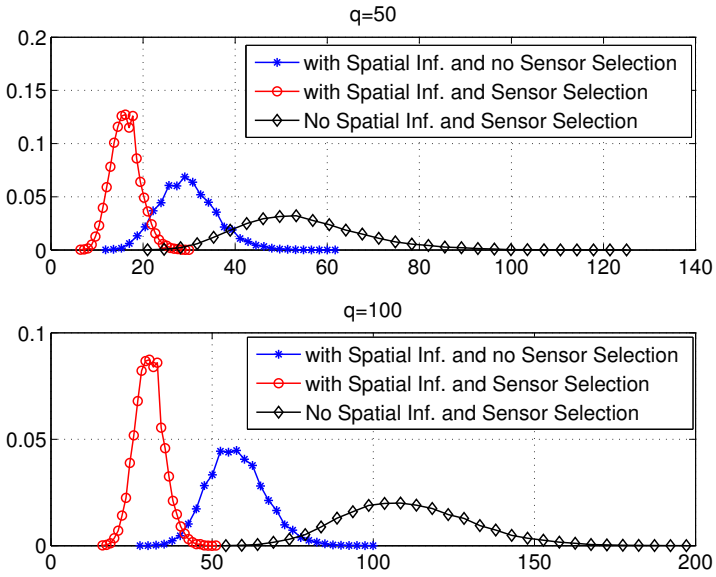


Figure 4.5: Normalized histograms of 20000 points of the detection delays for thresholds q = 50 and q=100.

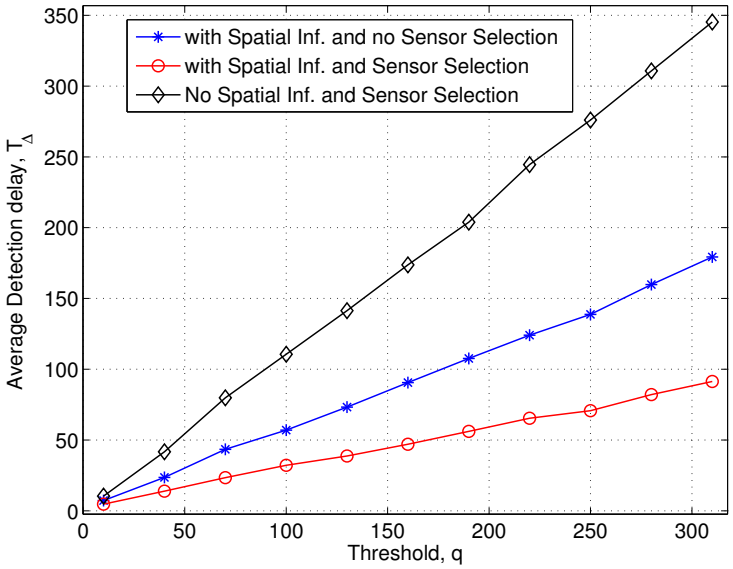


Figure 4.6: Average detection delay time T_Δ vs threshold q.

CHAPTER 5

Collaborative Spectrum Sensing with Spatial Signatures

In the previous chapter we presented a detailed discussion on rank-reduced multi-sensor detection schemes based on model order selection and exploitation of spatial correlation. In the present chapter we adopt those concepts to applying them to the application of multi-sensor based spectrum sensing to detect the presence of licensed user transmission. Due to the rapid growth in the field of radio communication, most of the available spectrum has already become congested and the assignment of frequencies to new services is currently a critical problem. On the other hand, studies show that assigned frequencies are not occupied all the time, implying that the traditional way of spectrum allocation has resulted in under utilization of such a precious resource. In that sense, cognitive radio (CR) has the potential to become the solution to the spectrum under utilization problem. The CR paradigm is based upon the coexistence of both licensed and unlicensed users within the same frequency band, in such a way that the latter are allowed to utilize the free spectrum holes left by former, in a dynamic and opportunistic manner [54, 55]. This technology is being very actively researched, and even regulatory and standard bodies have also begun to support the idea of spectrum reuse [56, 57]. Among the various functions of a CR system, reliable sensing of the spectrum of licensed or *primary* users (PU) is certainly of paramount importance. Such a spectrum sensing is performed by unlicensed or *secondary* users (SU), either following a single-sensor or a multi-sensor approach. The process of spectrum sensing with a single sensor is fundamentally limited by local impairments, such as the noise level, the SNR wall [15] and radio propagation effects such as path loss and fading experienced by this sensor, which significantly deteriorate its sensing performance [58]. In contrast, collaborative spectrum sensing relies on the combination of measurements coming from multiple neighboring sensors [59]. Therefore, collaborative approaches are able to circumvent most of the propagation impairments of single-sensor spectrum sensing, due to the presence of diversity in the set of measurements being processed at the fusion center [60]. It has to be taken into account that for the case of large-scale sensor networks, the signal of the PU will only reach to a subset of sensors (i.e. those sensors located close to the PU), which will typically be closely spaced, thus forming a cluster having highly correlated observations [17].

Motivated by the above facts, we propose a modified GLRT-based detector that achieves the regularization of the unknown covariance matrix with the help of *spatial signatures*,

the concept was preliminary introduced in the previous chapter. Herein, we extended it to the problem of detecting Gaussian random waveforms emitted from a PU with unknown covariance matrix. This is in contrast to the deterministic approach considered in the previous chapter, where an unknown but constant waveform was assumed to be transmitted by the PU, and where ideal propagation conditions as well as perfect knowledge on the signal parameters was also assumed. In that sense, the present contribution offers a much more realistic approach by assuming the emission of Gaussian random waveforms, by including the presence of shadowing and noise power uncertainty, and by taking into account the practical problems that may arise in large-scale networks when estimating the unknown covariance matrix of the PU. The latter is indeed related to the required number of observations to avoid ill-conditioning in the estimation of this matrix, which is typically on the order of the number of sensors [61]. Therefore, detection algorithms requiring the inverse or the determinant of this matrix can no longer be applied for short observation periods. To cope with this practical problem, the present work incorporates the concept of *shrinkage* estimation, a method that is found to improve the stability of estimated covariance matrices with short data records [61]. Simulation results have been obtained to compare the proposed detection schemes with and without spatial signatures, as well as with and without shrinkage estimation, showing that the introduction of spatial structures and shrinkage estimation significantly improves the overall detection performance.

The remaining of the chapter is organized as follows. In 5.1, the problem statement and details about the signal model are presented. Section 5.2 presents the structured signal model based on the concept of spatial signatures, and 5.3 introduces the proposed detection algorithm. In 5.4 we briefly discuss the shrinkage method for the estimation of the covariance matrix. Finally, we present the simulation results in 5.5.

5.1 Problem statement

We consider herein a large cognitive radio network (CRN) where both primary and secondary users coexist in the same geographical area. We assume an infrastructure-based secondary network [16], where each cell consists of a single base station (BS) working as a fusion center, and K SUs working as sensors. We also assume that sensors are deployed in the region following a uniform distribution and that the sensors and PU remain stationary in their position during the observation interval. The signal power emitted by the PU decays isotropically as a function of distance, and it is affected by a significant path loss attenuation due to the large area being covered by the network, as well as by fading/shadowing effects. As a consequence, only a subset of sensors will be able to receive enough power levels so as to easily detect the presence of the PU with a given detection performance [33, 62]. The rest of sensors will typically receive extremely weak power levels, and this observation allows us to distinguish between the so-called *active* and *inactive* sensors, respectively. In the process of collaborative spectrum sensing, the BS coordinates the opportunistic spectrum access of SUs in its cell, by directing sensors to perform spectrum sensing periodically. At the end of

each sensing period, all the sensors report their measurements to the BS, which makes the final decision about the presence or absence of the PU [60]. Once the final decision is made at the BS, it is broadcast back to SUs within the cell, in order to inform them about the presence or an absence of the PU. Similarly to [37], we further assume that the BS knows the location of the SUs, either through the use of positioning techniques, or through some calibration process.

5.1.1 Signal model and test statistics at the SU

In the collaborative sensing system considered herein, we assume that sensors simply measure the PU signal power on a target frequency band using an energy detector, and they report their sensing results to the BS [63]. This is a simplistic interpretation of collaborative sensing, which indeed covers a much wider area [64], but it allows us to concentrate on the specific problem of energy detection. Indeed, the energy detector is the simplest detector that can be constructed in practice. It uses very limited a-priori information regarding the signal, since the detection is based only on the signal power. In the sequel, we will consider an observation interval of $n = 1, 2, \ldots, N$ sensing periods. During the n-th sensing period, every SU captures a snapshot of $m = 1, 2, \ldots, M$ received signal samples in order to estimate the received signal power. At the i-th sensor (i.e. SU), two hypotheses arise regarding the existence of a PU signal on a given channel:

$$\begin{aligned} &\mathcal{H}_0 : \; y_i\left(m;n\right) = z_i\left(m;n\right) \\ &\mathcal{H}_1 : \; y_i\left(m;n\right) = g_i\left(m;n\right) + z_i\left(m;n\right) \end{aligned} \qquad i = 1, 2, \ldots, K \tag{5.1}$$

where $z_i(m;n) \sim \mathcal{N}(0, \sigma^2_{\varepsilon,i})$ are the i.i.d. samples encompassing the aggregate of random disturbances affecting each sensor, whereas $g_i(m;n)$ are the received signal samples corresponding to the random waveform emitted from the PU. Based on these samples, the energy detector at sensor i for the n-th sensing period is given by:

$$T_i(n) = \frac{1}{M}\mathbf{y}_i^T(n)\mathbf{y}_i(n) = \frac{1}{M}\sum_{m=1}^{M} y_i^2\left(m;n\right) \tag{5.2}$$

where $\mathbf{y}_i(n) \triangleq [y_i(1;n), y_i(2;n), \ldots, y_i(M;n)]^T$.

It is interesting to note that the energy detector in (5.2) can fairly be approximated by a Gaussian distribution in virtue of the central limit theorem (CLT), provided that M is sufficiently large[1]. Moreover, it should also be taken into account that the overall noise in a wireless receiver is often considered to be an ensemble of various effects, including not only the thermal noise contribution, but also other degradations such as the presence of interference signals from distant PUs or from other opportunistic SUs. Due to this fact, it is

[1]For instance, in the case of IEEE 802.22 WRANs, sensors measure the entire 6 MHz DTV channel at the Nyquist rate during observation intervals of 1 ms, and thus a total of $M = 6 \cdot 10^3$ samples are typically processed per snapshot [63].

very difficult to determine the exact noise powers $\sigma^2_{\varepsilon,i}$, even if we calibrate the system [65]. In some situations, this noise power uncertainty can cause the rise of the SNR wall, which can be understood as the minimum SNR below which signal cannot be detected, thus hindering the overall detection process [15]. Consequently, and from a practical point of view, it is of interest to assume that $\sigma^2_{\epsilon,i}$ is unknown. A similar statement can be made for the power being received from the PU at sensor i, which is referred herein as $P_{\varepsilon,i} \triangleq \mathrm{E}\left[|g_i(m;n)|^2\right]$, and is also considered to be unknown due to the unknown location of the PU and the presence of shadowing/fading that may alter the actual received power from its nominal value.

With the above considerations, the test statistics for the energy detector at sensor i in (5.2) can be modeled by the following Gaussian distribution [63]:

$$\begin{aligned} \mathcal{H}_0 :\ & T_i(n) \sim \mathcal{N}\left(\sigma^2_{\varepsilon,i}, \frac{2\sigma^4_{\varepsilon,i}}{M}\right) \\ \mathcal{H}_1 :\ & T_i(n) \sim \mathcal{N}\left(P_{\varepsilon,i} + \sigma^2_{\varepsilon,i}, \frac{2\left(P^2_{\varepsilon,i} + 2P_{\varepsilon,i}\sigma^2_{\varepsilon,i}\right)}{M} + \frac{2\sigma^4_{\varepsilon,i}}{M}\right) \end{aligned} \tag{5.3}$$

where both $\sigma^2_{\varepsilon,i}$ and $P_{\varepsilon,i}$ are assumed to remain constant during the whole observation interval of N sensing periods, and thus, they can be treated as unknown deterministic parameters herein.

5.1.2 Signal model and test statistics at the BS

Every sensor calculates an estimate of its received power level according to (5.2), and transmits this power estimate to the BS through a reporting channel. At the BS, the power estimates received from the K sensors at the n-th sensing period are stacked into the $(K \times 1)$ vector $\mathbf{x}(n) \triangleq [x_1(n), \ldots, x_K(n)]^T$, where $x_i(n)$ stands for the noisy and attenuated version of $T_i(n)$ after propagation through the reporting channel from sensor i to the BS [66]. Although the actual propagation effects of the K reporting channels are assumed to be unknown herein, they are considered to remain constant within the observation interval of N sensing periods. Therefore, the received signal model at the BS can still be expressed as some noisy contribution under $\mathcal{H}_0$ and some signal plus noise contribution under $\mathcal{H}_1$, as follows:

$$\begin{aligned} \mathcal{H}_0 :\ & \mathbf{x}\,(n) = \mathbf{w}\,(n) \\ \mathcal{H}_1 :\ & \mathbf{x}\,(n) = \mathbf{s}(n) + \mathbf{w}\,(n) \end{aligned} \tag{5.4}$$

for $n = 1, \ldots, N$ and with $\mathbf{w}(n) \sim \mathcal{N}(\boldsymbol{\mu}_w, \boldsymbol{\Sigma}_{\mathbf{w}})$ a $(K \times 1)$ vector containing the reported noise power levels at each sensor when no PU is present, whereas $\mathbf{s}(n) \sim \mathcal{N}(\boldsymbol{\mu}_s; \boldsymbol{\Sigma}_{\mathbf{s}})$ is a $(K \times 1)$ vector with the PU power levels. It is important to recall here that both $\mathbf{w}(n)$ and $\mathbf{s}(n)$ are power measurements, and thus, $\{\boldsymbol{\mu}_w, \boldsymbol{\mu}_s\}$ contain the mean noise powers and the mean signal powers at each sensor, respectively, whereas $\{\boldsymbol{\Sigma}_{\mathbf{w}}, \boldsymbol{\Sigma}_{\mathbf{s}}\}$ represent the variability of the corresponding power estimates being reported by the sensors. At the BS, and because of the disturbances that may appear due to propagation through the reporting channel, we

will assume that power measurements under $\mathcal{H}_0$ are i.i.d. with some common variability σ_w^2, in such a way that $\mathbf{\Sigma_w} = \sigma_w^2 \mathbf{I}_K$ for some unknown σ_w^2, and $\mathbf{I}_K$ the $(K \times K)$ identity matrix. Note also that both $\boldsymbol{\mu}_s$and $\mathbf{\Sigma_s}$ depend on the characteristics of the random waveform being emitted by the PU and its position with respect to the K sensors. Finally, and for the sake of clarity, we can express the signal model at the BS to be Gaussian distributed as follows:

$$\begin{aligned} \mathcal{H}_0 :&\ \mathbf{x}(n) \sim \mathcal{N}(\boldsymbol{\mu}_0, \mathbf{\Sigma}_0) \\ \mathcal{H}_1 :&\ \mathbf{x}(n) \sim \mathcal{N}(\boldsymbol{\mu}_1, \mathbf{\Sigma}_1) \end{aligned} \tag{5.5}$$

where $\boldsymbol{\mu}_0 \triangleq \boldsymbol{\mu}_w$ and $\mathbf{\Sigma}_0 \triangleq \mathbf{\Sigma_w}$ under the $\mathcal{H}_0$ hypothesis, whereas $\boldsymbol{\mu}_1 \triangleq \boldsymbol{\mu}_s + \boldsymbol{\mu}_w$ and $\mathbf{\Sigma}_1 \triangleq \mathbf{\Sigma_s} + \mathbf{\Sigma_w}$ under the $\mathcal{H}_1$ hypothesis. In practice, $\mathbf{\Sigma_s}$ will depart from a diagonal matrix, and the correlation represented by non-diagonal elements in $\mathbf{\Sigma}_s$ will typically indicate the presence of correlated shadowing effects in the received PU signal strengths [22].

5.2 Signatures-based problem formulation

5.2.1 Preliminaries

For an improved sensing performance, intuition suggests that the detection rule should rely on the observations of active sensors, thus discarding observations from the rest of inactive sensors. This approach can be understood as a kind of rank reduction method, in which the removal of the most noisy dimensions of the received signal subspace is known to substantially improve the overall signal-to-noise ratio. Moreover, since active sensors are typically located close to the PU, and also close to each other forming a spatial cluster, this side information should also be considered in the design of the detector. It is for this reason that one of the key points of this thesis is the identification of the set of active sensors, a purpose that will be achieved through the help of model order selection techniques and the spatial structure of the neighboring sensors. To this end, we propose a structured signal model based on the concept of *spatial signatures* that was introduced in Section 4.1.1 of the previous chapter.

5.2.2 Full structured signal model

By making explicit the role played by the spatial structure of the network, through the dependence on spatial signature matrix $\mathbf{H}$ introduced in introduced in Section 4.1.1, the values of $\mathbf{s}(n)$ in (5.4) can be expressed as $\mathbf{s}(n) = \mathbf{H}\mathbf{a}(n)$. Then, the signal model in (5.4) can be rewritten as:

$$\begin{aligned} \mathcal{H}_0 :&\ \mathbf{x}(n) = \mathbf{w}(n) \\ \mathcal{H}_1 :&\ \mathbf{x}(n) = \mathbf{H}\mathbf{a}(n) + \mathbf{w}(n) \end{aligned} \tag{5.6}$$

where $\mathbf{a}(n) \sim \mathcal{N}(\mu_a, \mathbf{\Sigma_a})$ is a $(K \times 1)$ vector containing the random weights of each signature onto the received signal. That is to say, the elements within $\mathbf{a}(n)$ quantify the

importance of each of the sensors signatures in the reconstruction of the signal field emitted by the PU. Therefore, by selecting the largest weights, we are actually choosing the most relevant sensors on the basis of their physical proximity to the PU. The more the sensor is located closer to the PU, the more its signature vector will be aligned with $\mathbf{s}(n)$, and thus the larger the weight assigned to this signature. By using this linear combination of signatures in (5.6), we are taking into account both the distances between neighboring sensors and the location of the sensors with respect to the PU, thus fully exploiting the spatial information contained within the received signals. From a statistical point of view, the only difference with respect to the unstructured signal model in (5.4)-(5.5) is that now, a specific structure is imposed onto both μ_1 and $\mathbf{\Sigma}_1$, with $\mu_1 = \mathbf{H}\mu_a$ and $\mathbf{\Sigma}_1 = \mathbf{H}\mathbf{\Sigma}_{\mathbf{a}}\mathbf{H}^T + \sigma_w^2\mathbf{I}_K$. Finally, once we have the signal model with the embedded spatial structure, the next step will be to select the relevant signatures contributing to the received signal, which is discussed in Section 5.2.3.

5.2.3 Rank-reduced structured signal model

The PU will typically appear at an unknown and random position, and it will be surrounded by a given number of $\mathcal{K} \leq K$ active sensors. We need to select the relevant signatures of the active sensors so that the rest of $K - \mathcal{K}$ signatures can reasonably be ignored. In some sense, we are in front of a detection problem where it is convenient to use a rank-reduced version of the signal model in (5.6). To do so, we will select the $\mathcal{K}$ most relevant signatures by using model order selection techniques [29]. Once we select the set of $\mathcal{K}$ active sensors, then their signatures will be stacked into a truncated $(K \times \mathcal{K})$ matrix $\mathbf{H}_{\mathcal{K}}$. Similarly, the selected weights will be stacked into a $(\mathcal{K} \times 1)$ vector $\mathbf{a}_{\mathcal{K}}(n)$, which is the reduced version of vector $\mathbf{a}(n)$ in (5.6). The resulting rank-reduced signal model can be written as:

$$\begin{aligned} \mathcal{H}_0 &: \ \mathbf{x}\,(n) = \mathbf{w}\,(n) \\ \mathcal{H}_1 &: \ \mathbf{x}\,(n) = \mathbf{H}_{\mathcal{K}}\mathbf{a}_{\mathcal{K}}(n) + \mathbf{w}\,(n) \end{aligned} \tag{5.7}$$

where the random weights $\mathbf{a}_{\mathcal{K}}(n)$ are assumed to be Gaussian distributed with $\mathbf{a}_{\mathcal{K}}(n) \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{a}_{\mathcal{K}}}, \mathbf{\Sigma}_{\mathbf{a}_{\mathcal{K}}})$. Therefore, the difference with respect to the full structured model in (5.6) is that $\boldsymbol{\mu}_1 = \mathbf{H}_{\mathcal{K}}\boldsymbol{\mu}_{\mathbf{a}_{\mathcal{K}}}$ and $\mathbf{\Sigma}_1 = \mathbf{H}_{\mathcal{K}}\mathbf{\Sigma}_{\mathbf{a}_{\mathcal{K}}}\mathbf{H}_{\mathcal{K}}^T + \sigma_w^2\mathbf{I}_K$, both depending on the unknown model order $\mathcal{K}$. In addition to the spatial information provided by the use of spatial signatures, the rank-reduced version of matrix $\mathbf{H}$ will indeed allow us to benefit from an equivalent SNR gain due the projection of the received signal onto a reduced dimensionality subspace.

5.3 Detection algorithms

In our detection problem, there are unknown parameters under both hypotheses that prevent us from adopting the well-known Neyman-Pearson detector. This obstacle is typically circumvented by adopting the GLRT approach, since it usually results in simple detectors with good performance. The main principal drawback, however, occurs when the dimension of

the unknown signal vector (i.e. the model order) is unknown [30]. In the signal model (5.7), the model order (i.e. the number of $\mathcal{K}$ active sensors) is unknown, and thus we cannot use the GLRT in straightforward manner. Instead, we need to modify the GLRT by using model order selection techniques [29] in order to determine the appropriate value for $\mathcal{K}$ to be used.

5.3.1 Unstructured GLRT

Before discussing the structured detector that incorporates the MDL criterion and spatial information into the traditional GLRT, herein, we derive the unstructured traditional GLRT for the signal model 5.4 using the discussion in Section 3.2.2. In the detection problem to be solved indicated in (5.5), we need to estimate the unknowns $\{\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0\}$ under $\mathcal{H}_0$, as well as $\{\boldsymbol{\mu}_1, \boldsymbol{\Sigma_1}\}$ under $\mathcal{H}_1$. To do so, we will assume that the BS has available the measurements of K sensors for N consecutive sensing periods, which are stacked into the $(K \times N)$ matrix $\mathbf{X} \triangleq [\mathbf{x}(1), \ldots, \mathbf{x}(n)]$. In these circumstances, the expression for the traditional or unstructured-GLRT can be written as:

$$\Lambda_{\text{UG}}(\mathbf{X}) = \frac{\max\limits_{\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0} f_{\mathbf{x}}(\mathbf{X}; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)}{\max\limits_{\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1} f_{\mathbf{x}}(\mathbf{X}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)} \gtrless_{\mathcal{H}_1}^{\mathcal{H}_0} \gamma \tag{5.8}$$

where γ is a threshold that determines a given probability of false alarm. The estimates of unknown parameters required in (5.8) are typically found by using maximum likelihood estimation (MLE), because it is asymptotically an unbiased and efficient estimator [35, Chap. 3]. Regarding the MLE of the unknown mean vector μ_1, it can easily be found as:

$$\hat{\boldsymbol{\mu}}_1 = \bar{\mathbf{x}} \triangleq \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}(n). \tag{5.9}$$

By using [35, Lemma 3.2.1], the MLE of the covariance matrix $\boldsymbol{\Sigma}_1$ can be written as:

$$\hat{\boldsymbol{\Sigma}}_1 = \hat{\boldsymbol{\Sigma}}_{\mathbf{x}} = \hat{\mathbf{R}}_{\mathbf{x}} - \bar{\mathbf{x}}\bar{\mathbf{x}}^T \tag{5.10}$$

where $\hat{\mathbf{R}}_x \triangleq \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}(n)\mathbf{x}^T(n)$. The sample covariance matrix $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}$ of the received samples at the BS and their mean vector $\bar{\mathbf{x}}$ are the sufficient statistics under hypothesis $\mathcal{H}_1$. Similarly under hypothesis $\mathcal{H}_0$, the ML estimate of μ_0 can be obtained as $\hat{\boldsymbol{\mu}}_0 = \bar{\mathbf{x}}$, and the ML estimate of σ_w^2 can be found as:

$$\hat{\sigma}_w^2 = \frac{1}{K} \text{Tr}(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}) \tag{5.11}$$

where $\text{Tr}(\cdot)$ is the trace operator. Replacing all of the unknowns with their estimates and after some mathematical manipulations, the final expression for (5.8) turns out to be :

$$\Lambda_{\text{UG}}(\mathbf{X}) = \frac{|\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}|}{\left[\frac{1}{K}\text{Tr}(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}})\right]^K} \gtrless_{\mathcal{H}_1}^{\mathcal{H}_0} \gamma. \tag{5.12}$$

The test statistic in (5.12), which does not take into account any spatial information, is nothing but the traditional Mauchly's *Sphericity test* [35, Chap 10]. The detector operates with the full sample covariance matrix $\hat{\mathbf{\Sigma}}_{\mathbf{x}}$, and it neither considers the relevance of the active sensors nor the spatial structure as a side information A modified Sphericity test has been proposed in [67, 68], by exploiting the fact that the signal covariance matrix may be of low-rank dimensionality. However, the performance of this unstructured Sphericity test-based detector can further be improved by incorporating spatial structure as a side information and using only the particular observations sent by the active sensors. This fact will be introduced later on in section 5.3.2.

5.3.2 Structured GLRT with spatial information

As we have already mentioned in Section 5.2.3, the key point in the proposed rank-reduced signal model in (5.7) is the determination of the spatial model order L. Since $L \leq K$, the detector with spatial information can operate with a reduced signal subspace by rejecting those dimensions (i.e. those spatial signatures) where the PU signal contribution is almost negligible. Therefore, some performance gain is expected compared to traditional unstructured signal detectors. The process of determining the optimal L is typically coupled with the one of signal detection. In particular, the GLRT approach to cope with an unknown model order decides that the signal is present if ,

$$\Lambda'_{\mathrm{SG}}\left(\mathbf{X}\right) = \max_{1\leq l\leq K} \Lambda_{\mathrm{SG},l}(\mathbf{X}) = \max_{1\leq l\leq K} \left\{ \frac{\max\limits_{\boldsymbol{\mu}_0,\sigma^2_{w_0}} f_{\mathbf{x}}\left(\mathbf{X};\boldsymbol{\mu}_0,\sigma^2_{w_0}\right)}{\max\limits_{\boldsymbol{\mu}_{\mathbf{a}_l},\mathbf{\Sigma}_{\mathbf{a}_l},\sigma^2_{w_1}} f_{\mathbf{x}}\left(\mathbf{X};\boldsymbol{\mu}_{\mathbf{a}_l},\mathbf{\Sigma}_{\mathbf{a}_l},\sigma^2_{w_1}\right)} \right\} \gtrless_{\mathcal{H}_1}^{\mathcal{H}_0} \gamma \quad (5.13)$$

where, we use $\sigma^2_{w_0}$ and $\sigma^2_{w_1}$ for the variance of power estimates being received at the BS under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. The denominator in (5.13) is the likelihood function of the observation under $\mathcal{H}_1$, which includes the unknown model order $\mathcal{K}$ as an additional parameter to be determined by searching the maximum of the GLRT from $l = 1, \ldots, K$. As we know from the discussions in Chapter 4, that the problem with (5.13) is that it will always implement the test statistic with maximum model order $l = K$. To cope with this problem, we will consider the well-known minimum description length (MDL) criterion. With the help of this selection technique, both the estimation of the true model order $\mathcal{K}$ and the evaluation of the GLRT can be done jointly. The detector combining the structured-GLRT and the MDL can be expressed as:

$$\Lambda_{\mathrm{SG}}\left(\mathbf{X}\right) = \min_{1\leq l\leq K} \{ l \log K - 2 \log \Lambda_{\mathrm{SG},l}(\mathbf{X}) \} \gtrless_{\mathcal{H}_1}^{\mathcal{H}_0} \gamma \quad (5.14)$$

where $l \log K$ is a penalty function preventing the GLRT statistic to monotonically increase with increasing model orders. In (5.14), $\Lambda_{\mathrm{SG},l}(\mathbf{X})$ stands for the structured-GLRT statistic while considering $\hat{\mathcal{K}} = l$ as the tentative model order, whose likelihood functions under both $\mathcal{H}_1$ and $\mathcal{H}_0$ will be derived in Section 5.3.2.1. Finally, in Section 5.3.2.2, we will propose

an algorithm to evaluate the structured-GLRT in (5.14) and take advantage of the available spatial information embedded onto the signatures matrix.

5.3.2.1 Derivation of the structured-GLRT for the tentative model order $\hat{\mathcal{K}} = l$

In this section, we will derive the expression for the structured-GLRT $\Lambda_{\mathrm{SG},l}(\mathbf{X})$ required in (5.14), which assumes a tentative model order $\hat{\mathcal{K}} = l$ for the parameters $\{\boldsymbol{\mu}_{\mathbf{a}_l}, \boldsymbol{\Sigma}_{\mathbf{a}_l}\}$ in the likelihood function under $\mathcal{H}_1$. Bearing in mind the Gaussian nature of the received measurements at the BS, as already introduced in Section 5.2.3, the likelihood function in the denominator of $\Lambda_{\mathrm{SG},l}(\mathbf{X})$ in (5.13) is given by

$$f_{\mathbf{x}}\left(\mathbf{X}; \boldsymbol{\mu}_{\mathbf{a}_l}, \boldsymbol{\Sigma}_{\mathbf{a}_l}, \sigma^2_{w_1}\right) = |\boldsymbol{\Sigma}_{1,l}|^{-\frac{N}{2}} \exp\left[-\frac{N}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}_{1,l}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}\right)\right] \tag{5.15}$$

where the unknown parameter $\boldsymbol{\Sigma}_{\mathbf{a}_l}$ is embedded into the covariance matrix $\boldsymbol{\Sigma}_{1,l}$, since $\boldsymbol{\Sigma}_{1,l} = \mathbf{H}_l\boldsymbol{\Sigma}_{\mathbf{a}_l}\mathbf{H}_l^T + \sigma^2_{w_1}\mathbf{I}$. The sample covariance matrix $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}$ has the expression:

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^{N}\left(\mathbf{x}\left(n\right) - \mathbf{H}_l\boldsymbol{\mu}_{\mathbf{a}_l}\right)\left(\mathbf{x}\left(n\right) - \mathbf{H}_l\boldsymbol{\mu}_{\mathbf{a}_l}\right)^T. \tag{5.16}$$

The ML estimate of the unknown mean vector $\boldsymbol{\mu}_{\mathbf{a}_l}$ is $\hat{\boldsymbol{\mu}}_{\mathbf{a}_l} = \left(\mathbf{H}_l^T\mathbf{H}_l\right)^{-1}\mathbf{H}_l^T\bar{\mathbf{x}}$, where $\bar{\mathbf{x}} \triangleq (1/N)\sum_{n=1}^{N}\mathbf{x}\left(n\right)$. Substituting $\hat{\boldsymbol{\mu}}_{\mathbf{a}_l}$ into (5.16) and using [35, Lemma 3.2.1], we can write (5.16) as:

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{x}} = \hat{\mathbf{R}}_x - \bar{\mathbf{x}}\bar{\mathbf{x}}^T + \left(\bar{\mathbf{x}} - \bar{\mathbf{x}}_p\right)\left(\bar{\mathbf{x}} - \bar{\mathbf{x}}_p\right)^T \tag{5.17}$$

where $\bar{\mathbf{x}}_p = \mathbf{P}_{\mathbf{H}_l}\bar{\mathbf{x}}$ and $\mathbf{P}_{\mathbf{H}_1} \triangleq \mathbf{H}_l\left(\mathbf{H}_l^T\mathbf{H}_l\right)^{-1}\mathbf{H}_l^T$ is the projection matrix onto the l-dimensional subspace spanned by $\mathbf{H}_l$. The vector $\bar{\mathbf{x}}_p$ is therefore the projected version of the mean vector $\bar{\mathbf{x}}$ onto the subspace spanned by the columns (i.e. signatures) of $\mathbf{H}_l$. Next, in order to find the ML estimate of $\hat{\boldsymbol{\Sigma}}_{\mathbf{a}_l}$ we can apply the logarithm on both sides of (5.15), take the derivative w.r.t. $\boldsymbol{\Sigma}_{\mathbf{a}_l}$ and equate to zero. By doing so, we get [69, Sec. 8.5],

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{a}_l} = \mathbf{H}_l^{\dagger}\left[\hat{\boldsymbol{\Sigma}}_{\mathbf{x}} - \hat{\sigma}^2_{w_1}\mathbf{I}\right]\mathbf{H}_l^{\dagger T} \tag{5.18}$$

where $\mathbf{H}_l^{\dagger} \triangleq \left(\mathbf{H}_l^T\mathbf{H}_l\right)^{-1}\mathbf{H}_l^T$ is the Moore-Penrose pseudoinverse and $\hat{\sigma}^2_{w_1}$ is given by

$$\hat{\sigma}^2_{w_1} = \frac{\mathrm{Tr}\left[\mathbf{P}^{\perp}_{H_l}\hat{\boldsymbol{\Sigma}}_x\right]}{K - l + 1} \tag{5.19}$$

with $\mathbf{P}^{\perp}_{\mathbf{H}_l} \triangleq \mathbf{I} - \mathbf{P}_{\mathbf{H}_l}$ the orthogonal projection matrix of $\mathbf{P}_{\mathbf{H}_l}$. In (5.19), the variance $\sigma^2_{w_1}$ of power estimates at the BS is estimated by using a vector that is the projected version of the observation vector $\mathbf{x}\left(n\right)$ onto the noise subspace, since $\mathbf{P}^{\perp}_{H_l}\hat{\boldsymbol{\Sigma}}_x = \mathbf{P}^{\perp}_{H_l}\left(\hat{\mathbf{R}}_x - \bar{\mathbf{x}}\bar{\mathbf{x}}^T\right) = \frac{1}{N}\sum_{n=1}^{N}\left[\mathbf{P}^{\perp}_{H_l}\mathbf{x}\left(n\right)\right]\left[\mathbf{P}^{\perp}_{H_l}\mathbf{x}\left(n\right)\right]^T - \mathbf{P}^{\perp}_{H_l}\bar{\mathbf{x}}\left(\mathbf{P}^{\perp}_{H_l}\bar{\mathbf{x}}\right)^T$. Consequently, for the overall covari-

ance matrix $\mathbf{\Sigma}_{1,l}$,

$$\hat{\mathbf{\Sigma}}_{1,l} = \mathbf{P}_{\mathbf{H}_l}\left[\hat{\mathbf{\Sigma}}_{\mathbf{x}} - \hat{\sigma}^2_{w_1}\mathbf{I}\right]\mathbf{P}_{\mathbf{H}_l} + \hat{\sigma}^2_{w_1}\mathbf{I} \tag{5.20}$$

which can also be written as,

$$\hat{\mathbf{\Sigma}}_{1,l} = \mathbf{P}_{\mathbf{H}_l}\hat{\mathbf{\Sigma}}_{\mathbf{x}}\mathbf{P}_{\mathbf{H}_l} + \frac{1}{K-l+1}\mathrm{Tr}\left[\mathbf{P}^{\perp}_{\mathbf{H}_l}\hat{\mathbf{\Sigma}}_x\right]\mathbf{P}^{\perp}_{\mathbf{H}_l}. \tag{5.21}$$

On the other hand, under hypothesis $\mathcal{H}_0$, we need to determine the unknown parameters $\left\{\mu_0, \sigma^2_{w_0}\right\}$ required by likelihood function in the numerator of $\Lambda_{\mathrm{SG},l}(\mathbf{X})$ in (5.13). Regarding the the ML estimate of $\sigma^2_{w_0}$, it can be obtained as $\hat{\sigma}^2_{w_0} = \frac{1}{K}\mathrm{Tr}\left(\hat{\mathbf{R}}_{\mathbf{x}} - \bar{\mathbf{x}}\bar{\mathbf{x}}^T\right) \simeq \frac{1}{K}\mathrm{Tr}\left(\hat{\mathbf{\Sigma}}_{\mathbf{x}}\right)$, where we already used the fact that $\hat{\mu}_0 = \bar{\mathbf{x}}$. With these results in mid, we can obtain the expression for the structured-GLRT with tentative model order $\hat{K}_a = l$ as:

$$\Lambda_{\mathrm{SG},l}(\mathbf{X}) \simeq \frac{\left|\hat{\sigma}^2_{w_0}\mathbf{I}\right|^{-\frac{N}{2}}\exp\left[-\frac{N}{2}\right]}{\left|\hat{\mathbf{\Sigma}}_l\right|^{-\frac{N}{2}}\exp\left[-\frac{N}{2}\right]} = \frac{\left|\hat{\mathbf{\Sigma}}_{1,l}\right|^{\frac{N}{2}}}{\hat{\sigma}^{KN}_{w_0}}. \tag{5.22}$$

Substituting $\hat{\mathbf{\Sigma}}_{1,l} = \mathbf{P}_{\mathbf{H}_l}\hat{\mathbf{\Sigma}}_x\mathbf{P}_{\mathbf{H}_l} + \hat{\sigma}^2_{w_1}\mathbf{P}^{\perp}_{\mathbf{H}_l}$ and $\hat{\sigma}^2_{w_0} = \frac{1}{K}\mathrm{Tr}(\hat{\mathbf{\Sigma}}_{\mathbf{x}})$, after some mathematical manipulations, (5.22) becomes

$$\Lambda_{\mathrm{SG},l}(\mathbf{X}) \simeq \frac{\left|\mathbf{P}_{\mathbf{H}_l}\hat{\mathbf{\Sigma}}_{\mathbf{x}}\mathbf{P}_{\mathbf{H}_l} + \frac{1}{K-l+1}\mathrm{Tr}(\mathbf{P}^{\perp}_{\mathbf{H}_l}\hat{\mathbf{\Sigma}}_{\mathbf{x}})\mathbf{P}^{\perp}_{\mathbf{H}_l}\right|^{\frac{N}{2}}}{\left[\frac{1}{K}\mathrm{Tr}(\hat{\mathbf{\Sigma}}_{\mathbf{x}})\right]^{\frac{KN}{2}}}. \tag{5.23}$$

The expression in (5.23) provides a closed-form expression for the structured-GLRT with tentative model order $\hat{L} = l$. The main characteristic of this expression is that it selects the most relevant spatial signatures and then on the basis of these signatures, it reduces the rank of the covariance matrix of the observed measurements, $\mathbf{\Sigma}_{\mathbf{x}}$. This statement can be explained by defining $\mathbf{\Psi}_l \triangleq \mathbf{P}_{\mathbf{H}_l}\hat{\mathbf{\Sigma}}_{\mathbf{x}}\mathbf{P}_{\mathbf{H}_l}$ and noticing that $\mathbf{\Psi}_l = \mathbf{P}_{\mathbf{H}_l}\hat{\mathbf{R}}_{\mathbf{x}}\mathbf{P}_{\mathbf{H}_l} - \mathbf{P}_{\mathbf{H}_l}\bar{\mathbf{x}}\bar{\mathbf{x}}^T\mathbf{P}_{\mathbf{H}_l}$, which can be equivalently expressed by using the properties of projection matrices as $\mathbf{\Psi}_l \triangleq \frac{1}{N}\sum_{n=0}^{N}\left[\mathbf{P}_{\mathbf{H}_l}\mathbf{x}(n)\right]\left[\mathbf{P}_{\mathbf{H}_l}\mathbf{x}(n)\right]^T - \left[\mathbf{P}_{\mathbf{H}_l}\bar{\mathbf{x}}\right]\left[\mathbf{P}_{\mathbf{H}_l}\bar{\mathbf{x}}\right]^T$. The expression of $\mathbf{\Psi}_l$ clearly shows that it is indeed the sample covariance matrix of a vector achieved by projecting the received observations $\mathbf{x}(n)$ onto the specific subspace being spanned by the signatures of active sensors. This will indeed result in an SNR gain due to the projection of the observation vector onto a reduced dimensionality subspace.

5.3.2.2 Implementation of the Structured-GLRT with MDL

The next step is to substitute the expression in (5.23) into (5.14), and perform the joint PU signal detection and model order selection from $l = 1, \ldots, K$. To do so, we summarize the implementation of the resulting detector in the pseudo-code description indicated in Algorithm 5.1.

Algorithm 5.1 MDL-based structured GLRT

1. With the received observation vectors $\mathbf{x}(n)$, $n = 1, 2, \cdots, N$, we calculate:

$$\hat{\mathbf{R}}_x = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}(n)\,\mathbf{x}^T(n) \text{ and } \bar{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}(n)$$

2. Find $\hat{\boldsymbol{\mu}}_{\mathbf{a}} = \mathbf{H}^{-1}\bar{\mathbf{x}}$.

3. Take the absolute values $|\hat{\boldsymbol{\mu}}_{\mathbf{a}}|$ and sort $|\hat{\boldsymbol{\mu}}_{\mathbf{a}}|$ in descending order, which results in $\tilde{\boldsymbol{\mu}}_{\mathbf{a}} = \text{sort}\left(|\hat{\boldsymbol{\mu}}_{\mathbf{a}}|\right)$.

4. Reorder the signature vectors in $\mathbf{H}$ according to the sorted $\tilde{\boldsymbol{\mu}}_{\mathbf{a}}$, to get $\widetilde{\mathbf{H}}$.

5. Implement the detector as:

   - Initialize $\mathbf{t} = \langle\,\rangle$ and $l = 1$.
   - while $l \leq K$ do:
     - Set $\mu_{\mathbf{a}_l}$=$\widetilde{\mu}_{\mathbf{a}}(1:l)$.
     - Set $\mathbf{H}_l$=$\widetilde{\mathbf{H}}(1:l)$.
     - Calculate $\mathbf{P}_{\mathbf{H}_l} = \mathbf{H}_l\left(\mathbf{H}_l^T\mathbf{H}_l\right)^{-1}\mathbf{H}_l^T$.
     - Calculate $\hat{\boldsymbol{\Sigma}}_x$ in (5.16).
     - Calculate $\Lambda_{\text{SG},l}(\mathbf{X})$ in (5.23).
     - Push the result of $l\log K - 2\log\Lambda_{\text{SG},l}(\mathbf{X})$ onto the vector $\mathbf{t}$.
     - $l = l + 1$.
   - end while
   - $\Lambda_{\text{SG}}(\mathbf{X}) = \min\{\mathbf{t}\} \gtrless_{\mathcal{H}_1}^{\mathcal{H}_0} \gamma$ as in (5.14).

5.4 Improved estimation of the covariance matrix

Both the unstructured and the structured-GLRT detectors presented in this paper are found to be based on the determinant of covariance matrices, which are typically estimated through the sample covariance, as in (5.10) and (5.17), respectively. Therefore, and although it is often taken for granted, a critical requirement for the GLRT detectors under study is that the sample covariance matrices must be non-singular and positive definite. To this end, we have to make sure that the number of available observations at the BS, given by N, is much larger than the number of sensors K (i.e. $N \gg K$). However, in many sensor network deployments we typically have a very large K, and thus, using a number of samples greater than K is a requirement difficult to fulfill in practice. In these circumstances it is therefore needed to estimate the covariance matrix with fewer samples while keeping a reasonable detection performance. Stein in [70], introduced the concept of shrinkage applied to high-dimensional estimators, and he derived the striking result that the performance of MLE can always be improved upon by shrinking with a given factor α (*shrinkage intensity*). This improved covariance estimator is well-conditioned and always positive definite, even for small sample sizes [61]. The basic principle of shrinking estimators is to shrink the variation

of the eigenvalues in the sample covariance matrix, proceeding as follows:

$$\breve{\mathbf{\Sigma}}_{\mathbf{x}} = \alpha \mathbf{F}_0 + (1-\alpha)\,\hat{\mathbf{\Sigma}}_{\mathbf{x}} \quad , \quad 0 \leq \alpha \leq 1 \tag{5.24}$$

where $\mathbf{F}_0$ is the target matrix, which is chosen to be positive definite (and therefore non singular), well-conditioned and we assume it herein as $\mathbf{F}_0 = \frac{1}{K}\mathrm{Tr}(\hat{\mathbf{\Sigma}}_{\mathbf{x}})\mathbf{I}_K$. The interested reader can find further details about the target matrix in [61]. Now for the expression in (5.24), we need to choose an appropriate α, the shrinkage intensity parameter. In [71], the authors discuss shrinkage methods that calculate the intensity parameter on the basis of received observations. They present what they call an oracle approximating shrinkage (OAS) estimator, which is an iterative method presented in Algorithm 5.2.

Algorithm 5.2 The Oracle approximating shrinkage estimator (OAS)

1. `Initialize` α `and` δ_{target}`.`
2. `Implement the shrinkage estimation as:`
   - `while covariance matrix estimation error` $\delta > \delta_{\text{target}}$ `do:`
     - `Calculate` $\breve{\mathbf{\Sigma}}_{\mathbf{x}}^{(k)} = \alpha^{(k)}\mathbf{F}_0 + \left(1-\alpha^{(k)}\right)\hat{\mathbf{\Sigma}}_{\mathbf{x}}$`.`
     - `Calculate` $\alpha^{(k+1)} = \frac{\left(1-\frac{2}{K}\right)\mathrm{Tr}\left(\breve{\mathbf{\Sigma}}_{\mathbf{x}}^{(k)}\hat{\mathbf{\Sigma}}_{\mathbf{x}}\right)+\mathrm{Tr}^2\left(\breve{\mathbf{\Sigma}}_{\mathbf{x}}^{(k)}\right)}{\left(N+1-\frac{2}{K}\right)\mathrm{Tr}\left(\breve{\mathbf{\Sigma}}_{\mathbf{x}}^{(k)}\hat{\mathbf{\Sigma}}_{\mathbf{x}}\right)+\left(1-\frac{N}{K}\right)\mathrm{Tr}^2\left(\breve{\mathbf{\Sigma}}_{\mathbf{x}}^{(k)}\right)}$
   - `end while`

In Algorithm 5.2 δ_{Target}, represents a specified threshold for the covariance matrix estimation error. The algorithm stops once the estimation error turns out to be less than this threshold, reaches to the following stable value of the shrinkage parameter,

$$\alpha_{\text{approx}} \approx \min\left\{ \frac{\left(1-\frac{2}{K}\right)\mathrm{Tr}(\hat{\mathbf{\Sigma}}_{\mathbf{x}}^2) + \mathrm{Tr}^2(\hat{\mathbf{\Sigma}}_{\mathbf{x}})}{\left(N+1-\frac{2}{K}\right)\left[\mathrm{Tr}(\hat{\mathbf{\Sigma}}_{\mathbf{x}}^2) + \mathrm{Tr}^2(\hat{\mathbf{\Sigma}}_{\mathbf{x}})\right]},\ 1 \right\}. \tag{5.25}$$

In our detection schemes, we use α_{approx} in (5.25), the approximate value of the shrinkage parameter in the covariance matrix estimation process indicated in (5.24).

5.5 Simulation results

The motivation of this section is to assess the performance of the proposed structured-GLRT detector in (5.14) and (5.23), which takes advantage of the novel concept of spatial signatures introduced in Section 5.2, and whose implementation is described in the pseudo-code description of Algorithm 5.1. For the analysis to be conducted herein, we consider a wireless sensor network with a total of $K = 30$ sensors deployed in a squared field. The sensors are randomly placed within the field following a uniform distribution, and we assume that the PU appears at an unknown position. We have tested the detectors considered in this

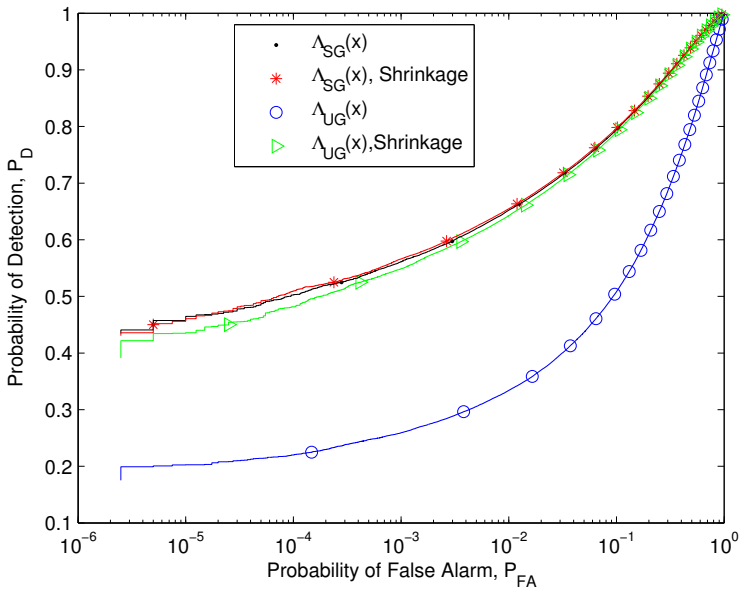


Figure 5.1: ROC curves ($\sigma_{\Delta_\sigma} = 0$): $\sigma_X = 3$ dB, $N = 100$, $M = 250$, $P_0 = -7$ dB and $\sigma_f = 0$ dB

paper for many different uniformly distributed topologies of K sensors, and we have found that the results have similar characteristics for different random topologies.

For the signal generation, we are assuming a quasi-static block-fading channel in which both the PU received power and the noise power at each sensor, do remain constant within the observation interval of N measurements. For a given observation interval, the PU received power at sensor i is given by $P_{\epsilon,i} = P_0 d_i^{-\beta} 10^{X_\sigma/10}$, where P_0 is the power at a reference distance from the PU, β is the signal decay exponent with typical values from $2-5$, d_i is the Euclidean distance between the PU and sensor i, and X_σ is the value of the log-normal shadowing. From one observation interval to the following, we allow the shadowing to vary according to $X_\sigma \sim \mathcal{N}(0, \sigma_X^2)$, with σ_X the dB-spread [63]. Regarding the noise power at sensor i, we are assuming $\sigma_\varepsilon^2 = 10^{\Delta_\sigma/10}$ for all sensors, with Δ_σ modeling the log-normal noise uncertainty as $\Delta_\sigma \sim \mathcal{N}(0, \sigma_{\Delta_\sigma}^2)$, from one observation interval to the following [72]. For the variability of power measurements at the BS, we are assuming $\sigma_w^2 = \frac{2\sigma_\varepsilon^4}{M} + \sigma_f^2$, as in (5.3), plus an additional disturbance σ_f^2 due to the noisy reporting links.

Regarding the assessment of the detectors being considered in this chapter, we will analyze their performance with and without the shrinkage estimation through the use of receiver operating characteristic (ROC) curves. Although the ROC curves fully characterize the performance, it is also desirable to have a single and quantitative figure of merit in order to compare different detectors. This metric is typically the area under the ROC curve (AUC), which varies between 0.5 (poor performance) and 1 (good performance).

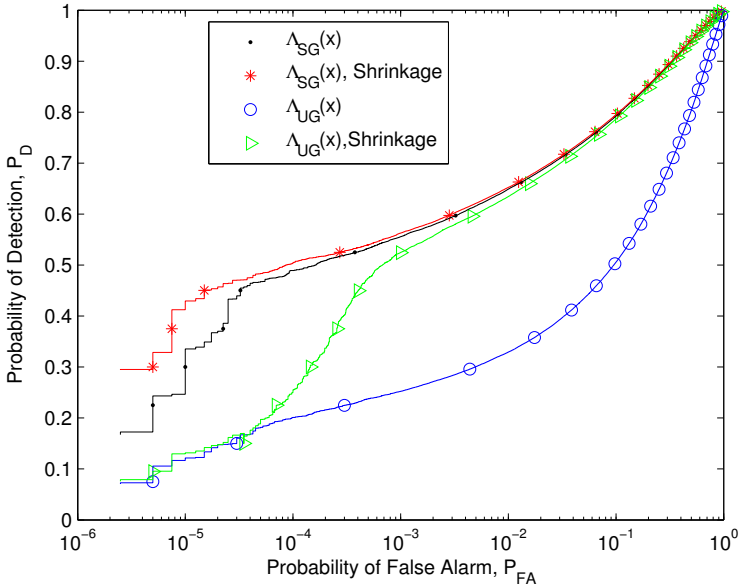


Figure 5.2: ROC curves ($\sigma_{\Delta_\sigma} = 3$): $\sigma_X = 3$ dB, $N = 100$, $M = 250$, $P_0 = -7$ dB and $\sigma_f = 0$ dB

Experiment 1: ROC curves for the detection schemes

In Figure 5.1 and 5.2, we evaluate the ROC curves for the proposed detection schemes by setting the PU transmit power $P_0 = -7$ dB, so that the mean received power in the sensor field turns out to be -37 dB. In Figure 5.1 , we present the result in the absence of noise power uncertainty (i.e. $\sigma_{\Delta_\sigma} = 0$ dB) and in Figure 5.2 we plot the ROC curve for the case of $\sigma_{\Delta_\sigma} = 3$ dB. From these results, it can be observed that at a lower probability of false alarm, the structured-GLRT $\Lambda_{\mathrm{SG}}(\mathbf{X})$ outperforms the unstructured one, $\Lambda_{\mathrm{UG}}(\mathbf{X})$, specially in the case of severe noise power uncertainty in Figure 5.2. Similarly, results also show that shrinkage estimation indeed improves the performance, which is more significant in the case of $\Lambda_{\mathrm{UG}}(\mathbf{X})$. In the case of the structured detector $\Lambda_{\mathrm{SG}}(\mathbf{X})$, there are no significant changes in performance due to the inclusion of shrinkage estimation. The reason for this is that with spatial signatures, we are indeed using a subspace of reduced dimension as we select the important samples in the received observations. Due to this selection, it indeed automatically shrinks the variance of the eigenvalues of the sample covariance matrix $\hat{\mathbf{\Sigma}}_{\mathbf{x}}$. That is why, further shrinkage of the covariance matrix does not bring significant changes in the detection result. In spite of that, it is clearly evident from the plot that the performance of the detector with the spatial signatures is better than the detector that is not using any spatial structure.

Experiment 2: Effects of noise power uncertainty

In Figure 5.3 and , we show the AUC plots to analyze the effects of noise power uncertainty (i.e. σ_{Δ_σ}). The AUC plots clearly show that between $\sigma_{\Delta_\sigma} = 0$ and $\sigma_{\Delta_\sigma} = 3$ dB, both detectors

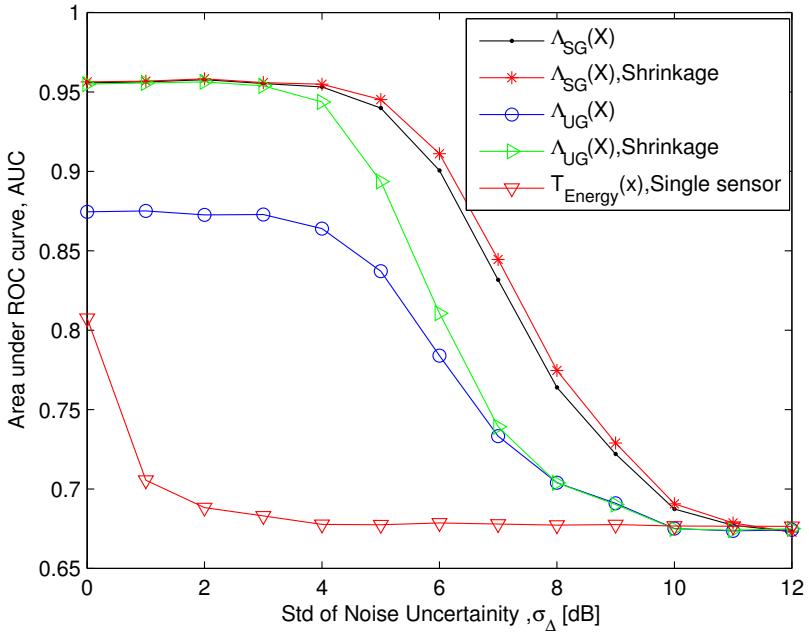


Figure 5.3: AUC curves to assess the effects of noise power uncertainty: $\sigma_X = 0$ dB, $N = 100$, $M = 250$, $P_0 = -7$ dB and $\sigma_f = 0$ dB

are robust against the noise power uncertainty, and $\Lambda_{\mathrm{SG}}(\mathbf{X})$ even shows robustness up to $\sigma_\Delta = 5$ dB. Indeed, the impact of noise power uncertainty is more severe in the case of low SNR [72] and $\Lambda_{\mathrm{SG}}(\mathbf{X})$ is able to counteract this situation by increasing the system's SNR by selecting the samples of active sensors. It can also be inferred from the results that the performance of the detection schemes improves by using shrinkage estimation, though the improvement is very very small in the case of $\Lambda_{\mathrm{SG}}(\mathbf{X})$. In this experiment, we have also analyzed the impact of noise power uncertainty on a energy detector at single node. We remark here that we selected a sensor that is located close to the PU and it receives signal with high SNR. In spite of that, we can see that the performance of the energy detector at a single node is severely effected by noise uncertainty, thus confirming the advantages of the proposed approach of collaborating sensing with spatial information.

Experiment 3: Effect of shadowing present in the channel between PU and the SUs

In this experiment, we analyze the effect of shadowing present in the channel between PU and the SUs. We analyze the two detection schemes for different values of shadowing, quantified by the parameter σ_X. For this analysis, we consider two different SNR regimes (low- and high-SNR), and noise power uncertainties calibrated by σ_{Δ_σ}. In Figure 5.4 and 5.5 we plot the AUC curves for low SNR regime. Figure 5.4 is for $\sigma_{\Delta_\sigma} = 0$ dB, and Figure 5.5 for $\sigma_{\Delta_\sigma} = 3$ dB. The results show that by using spatial signatures and shrinkage estimation, the detection performance improves. Interestingly, in the low-SNR regime both detection schemes perform better as the shadow fading becomes more variable (i.e. higher σ_X). This

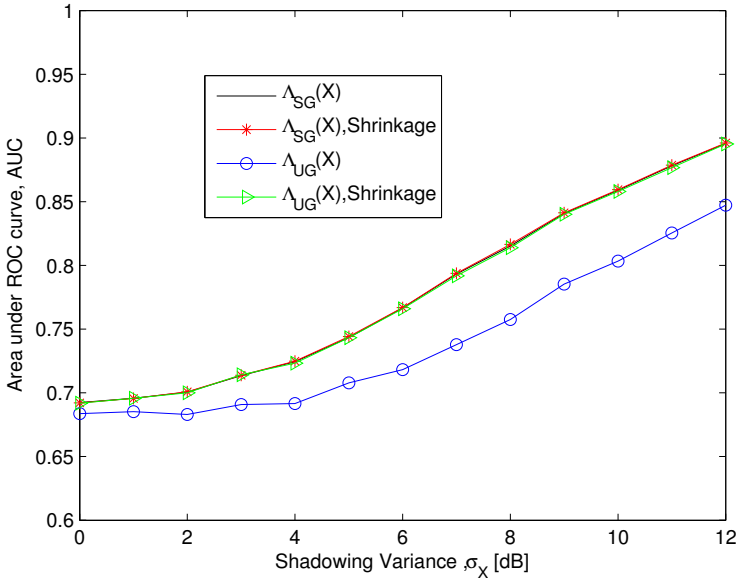


Figure 5.4: AUC curves to assess the effects of shadowing with $\sigma_{\Delta_\sigma} = 0$ dB: $N = 100$ and $M = 250$, $P_0 = -16$ dB and $\sigma_f = 0$ dB.

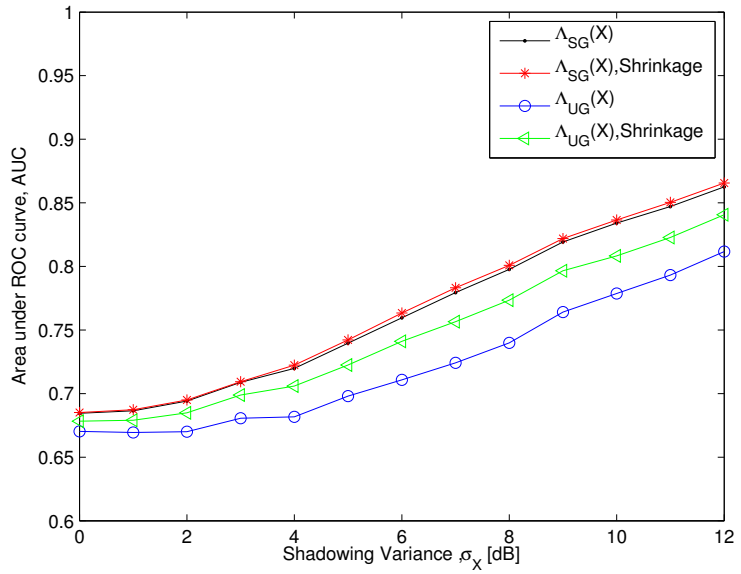


Figure 5.5: AUC curves to assess the effects of shadowing with $\sigma_{\Delta_\sigma} = 3$ dB: $N = 100$ and $M = 250$, $P_0 = -16$ dB and $\sigma_f = 0$ dB.

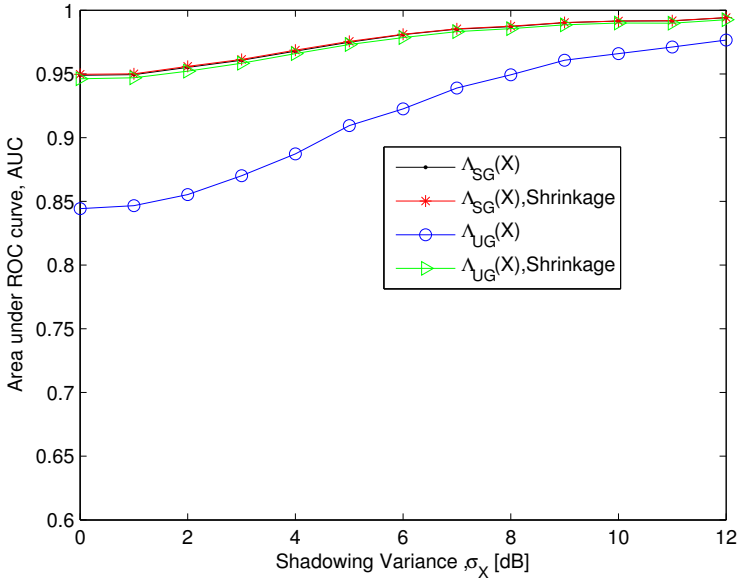


Figure 5.6: AUC curves to assess the effects of shadowing with $\sigma_{\Delta_\sigma} = 0$ dB:: $N = 100$ and $M = 250$, $P_0 = -16$ dB and $\sigma_f = 0$ dB.

is because of the heavy-tailed distribution of the primary signal strength, due to the log-normally-distributed shadow fading [73]. Therefore, in the case of low-SNR and large dB-spread, σ_X, the detection performance is found to improve. Similarly, in Figure 5.6 and Figure 5.7 we plot the AUC curves for the high-SNR regime. Figure 5.6 is for $\sigma_{\Delta_\sigma} = 0$ dB, and Figure 5.6 for $\sigma_{\Delta_\sigma} = 3$ dB. Here again, we can see that by using spatial signatures and shrinkage estimation, the detection performance improves, too. The behavior of the unstructured detector $\Lambda_{\mathrm{UG}}(\mathbf{X})$ is the same as it is in the case of low-SNR. In this experiment, it is also interesting to note that the effect of shrinkage estimation over $\Lambda_{\mathrm{SG}}(\mathbf{X})$ is negligible, since the plots with and without shrinkage are indistinguishable. On the other hand, the performance of $\Lambda_{\mathrm{UG}}(\mathbf{X})$ significantly increases due to introduction of shrinkage approach.

Experiment 4: Effect of sample support available to estimate covariance matrix

In this final experiment, we analyze the two detectors, for different values of the number of samples N, used to estimate the covariance matrices. The results are shown in Figure 5.8 and Figure 5.9, for noise power uncertainty $\sigma_{\Delta_\sigma} = 0$ dB and $\sigma_{\Delta_\sigma} = 3$ dB, respectively. We can see that the performance of $\Lambda_{\mathrm{SG}}(\mathbf{X})$ is better than $\Lambda_{\mathrm{UG}}(\mathbf{X})$. The interesting observation is that in the case of $\Lambda_{\mathrm{SG}}(\mathbf{X})$, for $\sigma_{\Delta_\sigma} = 0$, there are no significant changes in the performance by introducing the shrinkage estimation. In the case of traditional GLRT, the shrinkage estimation indeed improves the detection process, while considering small number of samples, N.

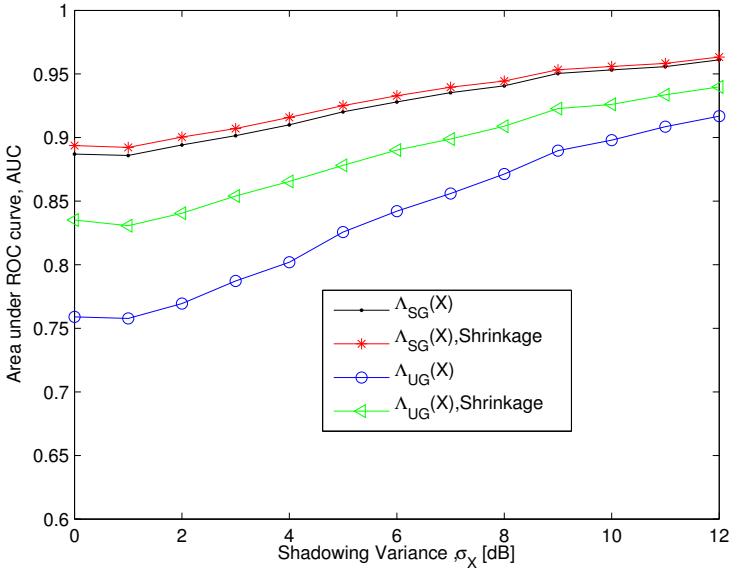


Figure 5.7: AUC curves to assess the effects of shadowing with $\sigma_{\Delta_\sigma} = 3$ dB:: $N = 100$ and $M = 250$, $P_0 = -16$ dB and $\sigma_f = 0$ dB.

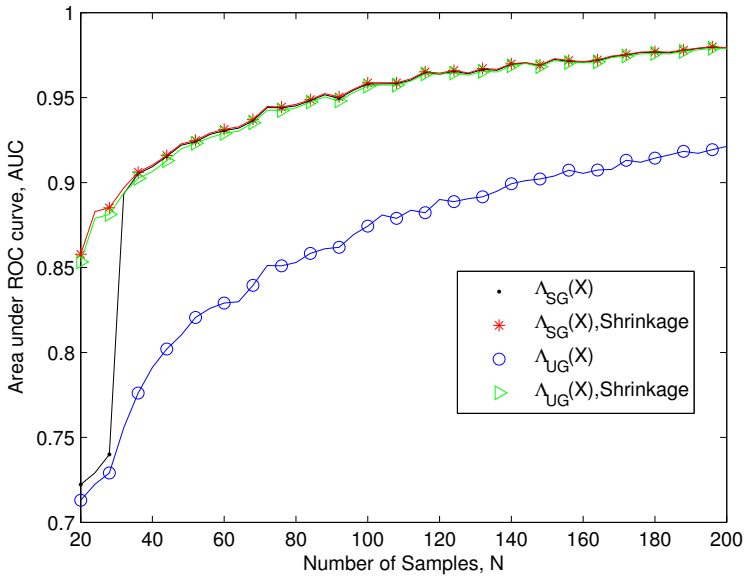


Figure 5.8: AUC curves to analyze the effects of sample size N : $\sigma_{\Delta_\sigma} = 0$ dB, $\sigma_X = 3$ dB, $M = 250$, $P_0 = -7$ dB and $\sigma_f = 0$ dB.

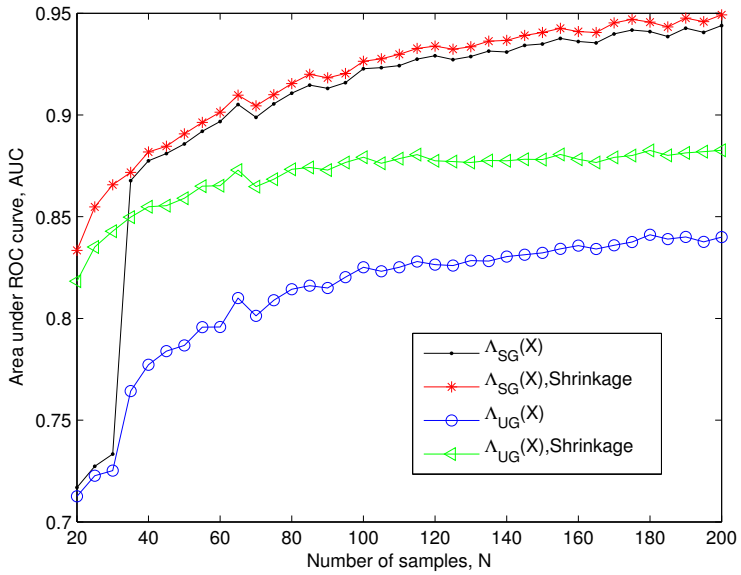


Figure 5.9: AUC curves to analyze the effects of sample size N : $\sigma_{\Delta_\sigma} = 3$ dB, $\sigma_{\mathrm{X}} = 3$ dB, $M = 250$, $P_0 = -7$ dB and $\sigma_f = 0$ dB.

CHAPTER 6

Multi-antenna and Multi-sensor Detection

In the literature it has been widely recognized that multiple antennas can offer space diversity and improve the spectrum sensing performance [74, 75]. Hence, in this chapter, we present novel spectrum sensing techniques for cognitive radio (CR) users that are equipped with multiple antennas. We present two novel schemes, the first one is for the case of a single user equipped with multiple antennas and the second one is for a cooperative approach (i.e. multi-user) with multiple sensors, where each sensor is equipped with multiple antennas.

For the single user case, we exploit the fact that the presence of any primary signal should result in spatial correlation in the observations received at different antennas of the user terminal [10, 76]. In addition to being spatially correlated, the received signal samples are usually correlated in time due to presence of temporal correlated channel, oversampling of the received signal or time correlation of the transmitted signal [19, 20]. This spatio-temporal correlation is a feature that can be used for detection purposes, since the remaining (i.e. undesired) noise processes at different antennas can be safely assumed statistically independent both in time and space.

Spectrum sensing methods that only exploit the spatial structure of the received signal covariance matrix have been of great interest in recent time [74]. The majority of these schemes are based on the multivariate statistical inference theory [36, 35] and interested readers can find a comprehensive details in [35, Ch. 9-10], which discusses in detail multivariate detectors for testing the independence of random observations with the help of the Generalized Likelihood Ratio Test (GLRT) on the basis of covariance matrices. These GLRT-based detectors typically end up with a simple quotient between the determinant of the sample covariance matrix and the determinant of its diagonal version, and these tests have been widely applied to the detection of signals especially in the context of cognitive radios [10, 76].

Through careful study on various existing spectrum sensing techniques, one can conclude that the signal's temporal correlation is not fully exploited in most of these techniques. In fact in most of the the existing spectrum sensing techniques, the temporal correlation is ignored or considered as a deleterious effect. One of the reasons is that temporal correlation often makes it difficult to achieve tractable solutions. However, the exploitation of temporal correlation jointly with spatial correlation can provide us extra side information to enhance the detection performance. Hence, it will be interesting to find ways to devise detection

mechanisms that exploit temporal correlation with tractable solutions. Considering these facts into consideration, the main focus of this work is to devise a multi-antenna detector that robustly exploits spatio-temporal correlation. In order to do so we propose a signal model that leads us to tractable detection schemes while exploiting the temporal correlation jointly with spatial correlation. In the proposed signal model, we consider the concept of correlated block-fading channel where the channel is temporally correlated inside sub-blocks of samples and independent between sub-blocks [19].

Based on the proposed signal model, we adopt the GLRT formulation that makes a decision based on whether the spatio-temporal sample covariance matrix is diagonal or not [34]. To make the discussion and notation simple, we call this GLRT scheme as spatio-temporal GLRT (ST-GLRT). Compared to the traditional spatial covariance based GLRT, the ST-GLRT provides some improved performance. It is because the ST-GLRT scheme exploits temporal correlation as an additional feature on top of spatial correlation and energy. However, since the GLRT involves the estimation of unknown parameters (i.e. covariance matrix), it depends on the sample size and the dimensionality. In practice, the GLRT is used based on the assumption that the sample size is large while the sample dimension is small. When this is not the case, the GLRT degenerates due to the singular and ill-conditioned sample covariance matrix [77, 78]. In the case of the ST-GLRT, we have to deal with both the spatial and temporal dimensions, and hence, the over all data dimension is even larger. Hence, the ST-GLRT has some limitations when the detection process requires quick decision and the sample support is small, as it is in the case of detection of primary signals in cognitive radio. Thus, although for the large sample support, the ST-GLRT can certainly achieve an improved detection performance, for small sample support it has some limitations that deserve a detailed study.

In order to bring robustness to the ST-GLRT, one may assume the existence of some underlying structure based on the spatial and temporal components of the covariance matrix. By using this prior information, we can reduce the demand for large sample support. Having said this, convenient structures that can be assumed for the spatio-temporal modeling of multi-antenna measurements are those given by persymmetric and Kronecker product structures [79, 80, 81]. In [34], the authors exploited the Toeplitz structure of the covariance matrix by assuming wide-sense stationarity. Doing so they proposed an approximated GLRT in the frequency domain that leads to robustness against the small sample support. Contrary to that work, in the present work we rather focus on exploiting the covariance structures without assuming approximation in the frequency domain. In particular, we will take advantage of the result in [80] which states that by exploiting the persymmetric structure, the number of independent vector measurements required for the covariance matrix estimator can be decreased by up to a factor of two. This will certainly bring down the demand for high sample support required for the ST-GLRT not to degenerate. Hence, we take advantage of the use of the persymmetric structure to modify the ST-GLRT and provide robustness against the repercussions of small sample support and large dimensional data.

Recently, the concept of exploiting the Kronecker structure of the covariance matrix has received a lot of interest in statistics [82, 83]. The formulations of the maximum likelihood

(ML) method for estimating the covariance matrix based on the Kronecker product has been previously discussed in [82, 83]. Similarly, in the cases where the correlation structure is not separable by using the Kronecker product, [84] discusses details about the nearest Kronecker product approximations. In this paper we use the Kronecker structure to modify the ST-GLRT by taking advantage of the inherent spatio-temporal structure of the received observations. In order to do so, we adopt and extend our earlier work [85] by using the Kronecker product to efficiently exploit the space-time correlation in a multi-antenna spectrum sensing scheme. In addition to the Kronecker product based factorization, we also assess the effects of exploring the persymmetric property of the covariance matrix, thus exploiting the fact that factored matrices have additional persymmetric structure [81]. Therefore, by exploiting the Kronecker product structure jointly with the persymmetric structure, the performance of detection schemes can further be improved in terms of efficiency of the number of sample to estimate the covariance matrix. Apart from the single user can with multiple antenna, we also exploit spatio-temporal correlation in the case of a WSN that monitors environmental effects in large region.

Towards the end of chapter we propose a spectrum sensing scheme with multiple distributed users where each user has multiple antennas. For both of these cases we start by adopting the traditional GLRT. Then we propose the proposed detection schemes for both of these two cases. In the the second part of this chapter, we consider a novel spectrum sensing for a cognitive radio network with multiple distributed radios where each radio has multiple antennas. The proposed spectrum sensing scheme exploits the fact that when any primary signal is present, measurements are spatially correlated due to inter-antennas and inter-users spatial correlation. The use of multiple antennas for detection at the multiple distributed radios has been previously discussed in [86, 87]. However, the detection schemes in these papers ignore the inter-radios spatial correlation. In contrast we consider a more general detection problem based on joint exploitation of inter-antennas and inter-users spatial structure. In order to exploit this correlation we propose two novel detectors that are robust against the high dimensionality of data and small sample support. The proposed detectors exploit the embedded spatial structure and decompose the large covariance matrix into inter-user and inter-antennas matrices by using single-pair and multi-pairs Kronecker products.

Finally, for each case to compare the proposed methods with the traditional techniques, numerical results are drawn. These results illustrate that the proposed detection schemes indeed out perform the traditional approaches especially in the case of small sample support.

6.1 Single-user multi-antenna Spatio-temporal GLRT

There are two kinds of techniques for spectrum sensing that take advantage of spatial diversity. One is cooperative spectrum sensing which involves single antenna multiple users (i.e. sensor nodes, radios) to detect the spectrum together, as we have seen in the previous

chapters. Each user conducts its own operation on the signal and transmits the result to a fusion center (base station) through signaling channel, and then the fusion center will combine the results from different users. The other approach is multi-antenna spectrum sensing [20, 10], which is performed by one user, and thus no signaling overhead is required. In this section we present novel detectors for spectrum sensing for a single user equipped with multi-antenna.

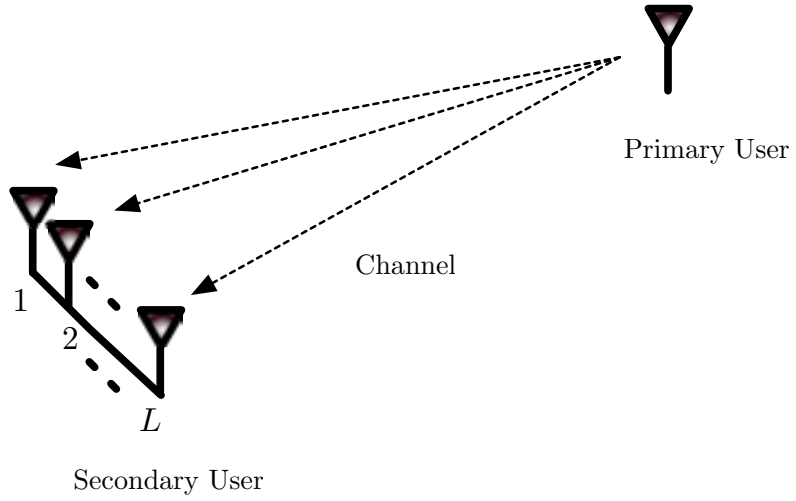


Figure 6.1: System model for a single user equipped with multiple antennas.

6.1.1 Problem statement

Herein, we address the problem of detecting the presence of a primary user by a single cognitive radio that is equipped with L antennas as shown in Fig 6.1. We assume no prior knowledge about the primary transmission, or the noise processes except that the noise is spatio-temporally independent. We focus on a practical scenario where due to the presence of a primary user (PU) signal, the received signals at the L antennas are correlated in space as well as time. Moreover, we assume uniform linear array (ULA), hence, the spatial correlation structure is represented by a $L \times L$ symmetric Toeplitz matrix $\mathbf{C}_s$. Similarly, we take into consideration that the received signal samples are correlated in time due to presence of temporal correlated channel, oversampling of the received signal or time correlation of the transmitted signal [19]. Hence, the consecutive received samples of the $L \times 1$ vector $\mathbf{x}(n),\ n = 1, \cdots, N_T$ at L antennas of a user are temporally correlated, where $\mathbf{x}(n) \triangleq [x_1(n), x_2(n), \cdots, x_L(n)]^T$ and N_T is total number received samples. The exploitation of temporal correlation jointly with spatial correlation can provide us extra side information to enhance the detection performance. In order to devise detection mechanisms that exploit temporal correlation with tractable solutions, the proposed technique can be described as follow:

1. We consider to split the received block of N_T vectors $\mathbf{x}(n)$ into M sub-blocks where each block contains N samples of vector $\mathbf{x}(n)$, as shown in the Fig. 6.2.

2. We assume that the consecutive vectored samples within each sub-block are considered to be temporally correlated with $N \times N$ temporal correlation matrix $\mathbf{C}_t$.

3. Independence is assumed between consecutive sub-blocks.

In the proposed mechanism, we consider to take into account the concept of correlated block-fading single input multiple output (SIMO) channel. The main feature of the correlated block-fading model is that the the channel is correlated in each block that contains N sample. However, the fading channel is independent between consecutive blocks [19, 88].

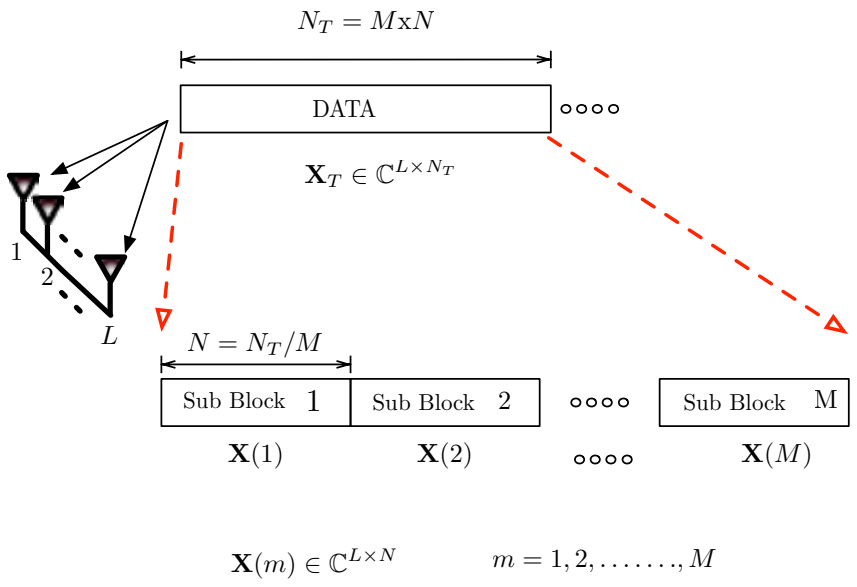


Figure 6.2: Schematic representation of the proposed methodology for slicing the observation block into M sub-blocks

6.1.2 Problem formulation and proposed signal model

In order to proceed, we need the distribution of $\{\mathbf{x}(n)\}_{n=1}^{N_T}$. We take it to be zero-mean complex Gaussian. In addition to resulting in tractable models and useful detectors, this assumption is reasonable if the primary network employs orthogonal frequency division multiplexing (OFDM) as modulation format[1] [34]. Based on the above facts, under the alternate hypotheses the vector sample $\mathbf{x}(n)$ can be represented as:

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{x}(n) = \mathbf{w}(n), \qquad n = 1, \cdots, N_T, \\ \mathcal{H}_1 &: \mathbf{x}(n) = \mathbf{s}(n) + \mathbf{w}(n), \quad n = 1, \cdots, N_T, \end{aligned} \tag{6.1}$$

where $\mathbf{s}(n)$ is the vector with the samples of the primary signal at the L antennas at time n and $\mathbf{w}(n)$ is the the additive noise vector. To process the data for the decision making, the SU collects N_T consecutive samples of vector $\mathbf{x}(n)$. Based on the proposed three steps

[1]We begin with the complex base-band signal sampled at the specific Nyquist rate.

approach discussed above, the received N_T vector samples are divided into M blocks and each block consists of N vector samples, such that $MN = N_T$. Let us define the m-th block as:

$$\mathbf{X}(m) = \begin{bmatrix} x_1(m) & x_1(2m) & \cdots & x_1(Nm) \\ x_2(m) & x_2(2m) & \cdots & x_2(Nm) \\ \vdots & \vdots & \ddots & \vdots \\ x_L(m) & x_L(2m) & \cdots & x_L(Nm) \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^T(m) \\ \mathbf{x}_2^T(m) \\ \vdots \\ \mathbf{x}_L^T(m) \end{bmatrix}, \tag{6.2}$$

where the i-th row, $\mathbf{x}_i(m) = [x_i(1m), x_i(2m), \cdots, x_i(Nm)]^T$ for $m = 1, ..., M$, contains N-samples at the i-th antenna. Let us define a vector $\mathbf{z}(m) \triangleq \text{vec}(\mathbf{X}(m))$. The covariance matrix of the $LN \times 1$ vector $\mathbf{z}(m)$ under hypothesis $\mathcal{H}_1$ is

$$\mathbf{\Sigma}_1 = \text{E}\left[\mathbf{z}\mathbf{z}^H\right] = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} & \cdots & \mathbf{\Sigma}_{1N} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} & \cdots & \mathbf{\Sigma}_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{N1} & \mathbf{\Sigma}_{N2} & \cdots & \mathbf{\Sigma}_{NN} \end{bmatrix} \in \mathbb{C}^{LN \times LN}, \tag{6.3}$$

where the sub-block covariance matrices $\mathbf{\Sigma}_{ik} = \text{E}\left[\mathbf{x}_i \mathbf{x}_l^T\right] \in \mathbb{C}^{L \times L}$, $1 \leq i, l \leq N$ in $\mathbf{\Sigma}_1$ capture all space-time second order information about the random vectors $\{\mathbf{x}_i\}_{i=1}^{L}$. Thus, the hypothesis testing problem becomes

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{z} \sim C\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_0), \\ \mathcal{H}_1 &: \mathbf{z} \sim C\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_1), \end{aligned} \tag{6.4}$$

where $C\mathcal{N}(0, \mathbf{\Sigma}_h)$ $h = 0, 1$, denotes the complex Gaussian distribution with zero mean and covariance $\mathbf{\Sigma}_h$. Based on the assumption about the noise powers at the L antennas of the user, we can classify the detection problem (6.4) into two categories. If the noise powers at the antennas are considered to be the same, the multi-antenna cognitive radio is called *calibrated* cognitive radio. In this case, under $\mathcal{H}_0$, $\mathbf{\Sigma}_0 = \sigma_w^2 \mathbf{I} \in \mathbb{C}^{LN \times LN}$ with σ_w^2 as the common noise power at different antennas. However, here in this work we focus on the most generic case of *uncalibrated* multiple antennas user, hence, $\mathbf{\Sigma}_0 = \mathbf{\Sigma}_{\text{A},0} \otimes \mathbf{I}_{N \times N}$ or

$$\mathbf{\Sigma}_0 = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_L^2 \end{bmatrix} \otimes \mathbf{I}_{N \times N}, \tag{6.5}$$

where $\sigma_i^2, i = 1, 2, ..., L$ is noise power at i-th antenna of the array. The diagonal structure of $\mathbf{\Sigma}_0$assumes that when no PU signal is present, the observations are uncorrelated in both spatial and temporal dimensions.

6.1.3 GLRT based on Spatio-Temporal correlation

In this section, we derive a spectrum sensing scheme that exploits both the spatial and temporal correlation. Since, the parameters $\{\mathbf{\Sigma}_0, \mathbf{\Sigma}_1\}$ under both hypotheses hence need to adopt the GLRT approach, the test statistic of ST-GLRT can be formulated as:

$$\Lambda_{\text{ST}}\left(\mathbf{Z}\right) = \frac{\max\limits_{\mathbf{\Sigma}_0} f_{\mathbf{z}}\left(\mathbf{Z}; \mathbf{\Sigma}_0\right)}{\max\limits_{\mathbf{\Sigma}_1} f_{\mathbf{z}}\left(\mathbf{Z}; \mathbf{\Sigma}_1\right)} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \gamma, \tag{6.6}$$

where $f_{\mathbf{z}}\left(\mathbf{Z}, \mathbf{\Sigma}_0\right)$ and $f_{\mathbf{z}}\left(\mathbf{Z}; \mathbf{\Sigma}_1\right)$ are the likelihood functions under hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Similarly, we assume that we have M independent blocks of the data $\mathbf{X}$, or equivalently vector $\mathbf{z}$, $\mathbf{Z} = \begin{bmatrix} \mathbf{z}\left(1\right) & \mathbf{z}\left(2\right), & \cdots, & \mathbf{z}\left(M\right) \end{bmatrix}$ available. In order to solve GLRT (6.6) we have to derive the maximum likelihood estimates of the parameters for each of the hypotheses. Note that the maximum likelihood estimates is asymptotically an unbiased and efficient estimator [25]. The expression for the likelihood function $f_{\mathbf{z}}\left(\mathbf{Z}; \mathbf{\Sigma}_1\right)$ under hypothesis $\mathcal{H}_1$ can be written as:

$$f_{\mathbf{z}}\left(\mathbf{X}; \mathbf{\Sigma}_1\right) = \frac{1}{\left(\pi\right)^{MLN} \left|\mathbf{\Sigma}_1\right|^M} \exp\left\{-M \text{tr}\left(\mathbf{\Sigma}_1^{-1} \hat{\mathbf{\Sigma}}_1\right)\right\}, \tag{6.7}$$

where $\hat{\mathbf{\Sigma}}_1 = \frac{1}{N}\sum_{m=1}^{M} \mathbf{z}(m)\mathbf{z}^H(m)$. As we have assumed independence between the M sub-blocks, the M snapshots of the observation vector $\mathbf{z}\left(m\right)$ are statistically independent and the maximum likelihood estimator is equal to the sample covariance matrix $\hat{\mathbf{\Sigma}}_1$. Under the alternate hypothesis the expression for $f_{\mathbf{z}}\left(\mathbf{Z}, \mathbf{\Sigma}_0\right)$ is

$$\begin{aligned} f_{\mathbf{z}}\left(\mathbf{Z}, \mathbf{\Sigma}_0\right) &= f_{\mathbf{z}}\left(\mathbf{Z}; \mathbf{\Sigma}_{\text{A},0} \otimes \mathbf{I}_{N\times N}\right) \\ &= \left(\pi\right)^{-\frac{1}{2}MLN} \left|\mathbf{\Sigma}_{\text{A},0} \otimes \mathbf{I}_{N\times N}\right|^{-M} \exp\left[-\sum_{m=1}^{M} \mathbf{z}(m) \left(\mathbf{\Sigma}_{\text{A},0}^{-1} \otimes \mathbf{I}_{N\times N}\right) \mathbf{z}^H(m)\right], \end{aligned} \tag{6.8}$$

Now we can write $f_{\mathbf{z}}\left(\mathbf{Z}; \mathbf{\Sigma}_{\text{A},0} \otimes \mathbf{I}_{N\times N}\right) = f_{\mathbf{x}}\left(\mathbf{X}_M; \mathbf{\Sigma}_{\text{A},0}, \mathbf{I}_{N\times N}\right)$ [89] and the expression for $f_{\mathbf{x}}\left(\mathbf{X}_M; \mathbf{\Sigma}_{\text{A},0}, \mathbf{I}_{N\times N}\right)$ can be written as:

$$\begin{aligned} f_{\mathbf{z}}\left(\mathbf{Z}, \mathbf{\Sigma}_0\right) &= f_{\mathbf{x}}\left(\mathbf{X}_M; \mathbf{\Sigma}_{\text{A},0}, \mathbf{I}_{N\times N}\right) \\ &= \left(\pi\right)^{-\frac{1}{2}MLN} \left|\mathbf{\Sigma}_{\text{A},0}\right|^{-MN} \left|\mathbf{I}_{N\times N}\right|^{-ML} \\ &\quad \exp\left[-\sum_{m=1}^{M} \text{tr}\left\{\mathbf{\Sigma}_{\text{A},0}^{-1} \mathbf{X}\left(m\right) \mathbf{I}_{N\times N} \mathbf{X}^H\left(m\right)\right\}\right], \end{aligned} \tag{6.9}$$

where $\mathbf{X}_M = \begin{bmatrix} \mathbf{X}\left(1\right) & \mathbf{X}\left(2\right), & \cdots, & \mathbf{X}\left(M\right) \end{bmatrix}$. By using (6.9) we can get MLE of $\mathbf{\Sigma}_{\text{A},0}$ as: $\hat{\mathbf{\Sigma}}_{\text{A},0} = \text{diag}\left(\frac{1}{MN}\sum_{m=1}^{M} \mathbf{X}\left(m\right) \mathbf{X}^H\left(m\right)\right)$ [90] and thus $\hat{\mathbf{\Sigma}}_0 = \hat{\mathbf{\Sigma}}_{\text{A},0} \otimes \mathbf{I}_{N\times N}$. In case, user is equipped with calibrated antennas, then the only unknown parameter under $\mathcal{H}_0$ is the noise power σ_w^2 , and therefore $\hat{\sigma}_w^2 = \frac{1}{KN}\text{tr}\left(\hat{\mathbf{\Sigma}}_1\right)$ [47]. To solve (6.6), in the case of uncalibrated

case we have to replace $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$ by $\hat{\mathbf{\Sigma}}_0$ and $\hat{\mathbf{\Sigma}}_1$, respectively. Doing so, and solving (6.6) we can get the final expression of the GLRT as:

$$\Lambda_{\mathrm{ST}}\left(\mathbf{Z}\right) = \frac{\left|\hat{\mathbf{\Sigma}}_1\right|}{\left|\hat{\mathbf{\Sigma}}_0\right|} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \gamma, \tag{6.10}$$

Note that the detection scheme in (6.10) assumes no structure for the covariance matrix, except that the covariance matrix is symmetric and non-singular. Furthermore, we remark that in the case of calibrated case the only difference is present in the denominator of (6.10) as instead of $\left|\hat{\mathbf{\Sigma}}_0\right|$, we will have $\frac{1}{KN}\mathrm{tr}\left(\hat{\mathbf{\Sigma}}_1\right)$, thus the detection scheme for calibrated case can be expressed as: $\Lambda_{\mathrm{T}}\left(\mathbf{Z}\right) = \left|\hat{\mathbf{\Sigma}}_1\right| / \frac{1}{KN}\mathrm{tr}\left(\hat{\mathbf{\Sigma}}_1\right)$.

Before concluding the discussion in this section, as a reference we present the traditional GLRT, that only exploits the spatial correlation and ignores the temporal correlation. By assuming the fact that the vector samples $\mathbf{x}\left(n\right),\ \ n = 1, \cdots, N_T$ are temporally uncorrelated, the GLRT scheme for the detection problem (6.1) can be formulated as:

$$\Lambda_{\mathrm{T}}\left(\mathbf{X}_{N_T}\right) = \frac{\underset{\mathbf{\Sigma}_{A,0}}{\max} f_{\mathbf{x}}\left(\mathbf{X}_{N_T}; \mathbf{\Sigma}_{\mathrm{A}}^{\mathcal{H}_0}\right)}{\underset{\mathbf{\Sigma}_{A,1}}{\max} f_{\mathbf{x}}\left(\mathbf{X}_{N_T}; \mathbf{\Sigma}_{\mathrm{A}}^{\mathcal{H}_1}\right)} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \gamma, \tag{6.11}$$

where $f_{\mathbf{x}}\left(\mathbf{X}_{N_T}; \mathbf{\Sigma}_{\mathrm{A}}^{\mathcal{H}_0}\right)$ and $f_{\mathbf{x}}\left(\mathbf{X}_{N_T}; \mathbf{\Sigma}_{\mathrm{A}}^{\mathcal{H}_1}\right)$ are the likelihood functions, $\mathbf{\Sigma}_{\mathrm{A}}^{\mathcal{H}_0}$ and $\mathbf{\Sigma}_{\mathrm{A}}^{\mathcal{H}_1}$ represent the covariance matrices under hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Similarly, matrix $\mathbf{X}_{N_T} = \begin{bmatrix} \mathbf{x}\left(1\right) & \mathbf{x}\left(2\right), & \cdots, & \mathbf{x}\left(N_T\right) \end{bmatrix} \in \mathbb{C}^{L \times N_T}$ contains all of the available samples of vector $\mathbf{x}$. Solving (6.11), the final expression of the detector based on the traditional GLRT formulation can be written as:

$$\Lambda_{\mathrm{T}}\left(\mathbf{X}_{N_T}\right) = \frac{\left|\hat{\mathbf{\Sigma}}_{\mathrm{A}}^{\mathcal{H}_1}\right|}{\left|\hat{\mathbf{\Sigma}}_{\mathrm{A}}^{\mathcal{H}_0}\right|} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \gamma, \tag{6.12}$$

where $\hat{\mathbf{\Sigma}}_{\mathrm{A}}^{\mathcal{H}_1} = \frac{1}{N_T}\sum_{n=1}^{N_T} \mathbf{x}(n)\mathbf{x}^H(n)$ is a sample covariance matrix. As under the hypothesis $\mathcal{H}_0$, when we assume uncalibrated case and only noise is present, thus the estimate of the diagonal matrix can be expressed as: $\hat{\mathbf{\Sigma}}_{\mathrm{A}}^{\mathcal{H}_0} = \mathrm{diag}\left(\hat{\mathbf{\Sigma}}_{\mathrm{A}}^{\mathcal{H}_1}\right)$ [90]. The detector (6.12) only exploits the energy and spatial-correlation across the L antennas of the receiver. This scheme completely ignore the information provided by the temporal correlation. Compared to (6.12), the ST-GLRT (6.10) promises improved detection performance since it uses temporal correlation as an additional detection metric. With the help of computer simulations, we compare the two detection schemes for cognitive enabled receiver with $L = 4$ antennas with the help of AUC plots as shown in the Fig.6.3.

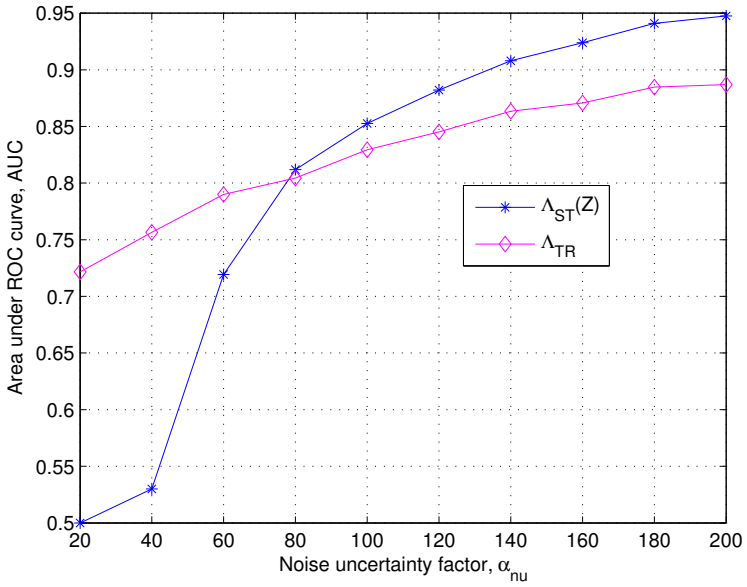


Figure 6.3: AUC curves: Comparison of the ST-GLRT $\Lambda_{\text{ST}}(\mathbf{Z})$ with the traditional GLRT $\Lambda_{\text{TR}}(\mathbf{X}_{N_T})$ for different sizes of the sample support (M), number of vector samples per sub-block $N = 15$ and number of antennas $L = 4$.

In Fig. 6.3 we can see that for the case where the sample support (M) is sufficient enough, the ST-GLRT clearly outperforms the traditional GLRT scheme in presence of similar conditions. However, in the case when $M < NL$, we can see that the ST-GLRT completely collapses. In order to circumvent this limitation in the following sections we propose some modifications in the proposed ST-GLRT by exploiting the presence of some inherent structures in the space-time correlation.

6.1.4 GLRT exploiting persymmetric structure

In order to solve the detection problem (6.4) with unknown covariance matrices, a critical requirement for the detectors based on the GLRT is that the sample covariance matrices must be non-singular and positive definite [83]. To this end, we have to make sure that the number of available observations given by M, is much larger than LN (i.e. $M \gg LN$). However, in quick spectrum sensing, a number of samples greater than LN is a requirement difficult to fulfill in practice [91]. Hence, the motivation of the remaining discussion is to bring robustness against this small sample support. Note that in (6.10) we assume no prior knowledge about the spatio-temporal structure of the covariance matrix except that it is symmetric. Hence, it is completely blind detector. One way to achieve the robustness against the small sample support is to look for possible a-priori known patterns/structures in the large spatio-temporal covariance matrix. Using this prior information in particular the persymmetric structure, more efficient detection schemes can be devised.

6.1.4.1 Persymmetric-block-Toeplitz structure

In this section we establish the fact that the spatio-temporal covariance matrix $\mathbf{\Sigma}_1$ has a persymmetric-block-Toeplitz structure. Toeplitz structured matrices belong to a sub-class of persymmetric matrices and a matrix is said to be Toeplitz matrix if its entries are constant on each diagonal [92]. Similarly, as matrix is a block-Toeplitz if its blocks are constant on each block diagonal. Similarly, a spatial process with equidistant observations (as in the case of ULA) and with a stationary covariance function yields a symmetric Toeplitz covariance matrix [84]. Moreover, if there is a regular temporal component (random spatial vector) with stationary covariance in time, the resulting space-time covariance matrix has an additional block Toeplitz structure. In our problem, we assume that the received $(L \times N)$ signal block at the structured antenna array as a multichannel auto-retrogressive process with $(N \times N)$ temporal correlation matrix $\mathbf{C}_t$, which is Toeplitz [93]. Similarly, the antenna spatial structure is represented by the Toeplitz correlation matrix $\mathbf{C}_s$ [93][94]. In particular, if the L elements of a linear array are symmetrically spaced with respect to the phase center (so that the l^{th} element and the $(L+1\text{–}l)^{\text{th}}$ element are at the same distance from the array phase center) the covariance matrix is persymmetric. Keeping the persymmetric nature of the matrices $\mathbf{C}_t$ and $\mathbf{C}_s$ in mind, the spatio-temporal covariance matrix $\mathbf{\Sigma}_1$ can be assumed as a persymmetric-block-Toeplitz [95]. Therefore, we have the following persymmetry condition satisfied [96]

$$\mathbf{\Sigma}_1 = \mathbf{J}_1 \mathbf{\Sigma}_1 \mathbf{J}_1, \tag{6.13}$$

where $(LN \times LN)$ matrix

$$\mathbf{J}_1 = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ \vdots & \cdots & 1 & 0 \\ 0 & & & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix} \tag{6.14}$$

denotes the reversal matrix. Based on these considerations, in Section 6.1.4.2 we present the modified GLRT that exploits the persymmetric property of the block Toeplitz covariance matrix.

6.1.4.2 Persymmetric GLRT(P-GLRT)

The difference in the formulation of the GLRT with the persymmetric comes due to the constraint (6.13). Therefore, the formulation of the GLRT based on the persymmetric covariance matrix can be represented as:

$$\begin{aligned} \Lambda_{\text{PS}}(\mathbf{Z}) &= \frac{\max\limits_{\mathbf{\Sigma}_0} f_{\mathbf{z}}(\mathbf{Z}; \mathbf{\Sigma}_0)}{\max\limits_{\mathbf{\Sigma}_1} f_{\mathbf{z}}(\mathbf{Z}; \mathbf{\Sigma}_1)} \gtrless_{\mathcal{H}_1}^{\mathcal{H}_0} \gamma \\ &\text{s.t } \mathbf{\Sigma}_1 = \mathbf{J}_1 \mathbf{\Sigma}_1 \mathbf{J}_1 \end{aligned}. \tag{6.15}$$

Comparing (6.15) to (6.6), we can see that the difference lies only in the denominator. In order to exploit the persymmetry of $\mathbf{\Sigma}_1$ we need to use the forward-backward (FB) log-likelihood [97] which is combination of the forward looking and the backward looking log-likelihood functions. The forward looking log-likelihood $\log f_{\mathbf{z}}\left(\mathbf{Z};\mathbf{\Sigma}_1\right)$ can be written as:

$$\log f_{\mathbf{z}}^{(F)}\left(\mathbf{Z};\mathbf{\Sigma}_1\right) = -MLN\log\pi - M\log\left|\mathbf{\Sigma}_1\right| - M\left(\sum_{m=1}^{M}\mathbf{z}^H(m)\mathbf{\Sigma}_1^{-1}\mathbf{z}(m)\right), \tag{6.16}$$

Now by dropping constant terms and using $\mathbf{z}^H(m)\mathbf{\Sigma}_1^{-1}\mathbf{z}(m) = \mathrm{tr}\left\{\mathbf{\Sigma}_1^{-1}\mathbf{z}(m)\mathbf{z}^H(m)\right\}$, (6.16) can be written as:

$$\log f_{\mathbf{z}}\left(\mathbf{Z};\mathbf{\Sigma}_1\right) = -\log\left|\mathbf{\Sigma}_1\right| - \mathrm{tr}\left\{\mathbf{\Sigma}_1^{-1}\hat{\mathbf{\Sigma}}_1\right\}, \tag{6.17}$$

where $\hat{\mathbf{\Sigma}}_1 = \frac{1}{N}\sum_{n=1}^{N}\mathbf{z}(m)\mathbf{z}^H(m)$. Similarly, with constraint $\mathbf{\Sigma}_1 = \mathbf{J}_1\mathbf{\Sigma}_1\mathbf{J}_1$ the backward looking log-likelihood $\log f_{\mathbf{z}}^{(B)}\left(\mathbf{Z};\mathbf{\Sigma}_1\right)$ can be written as [97]:

$$\log f_{\mathbf{z}}^{(B)}\left(\mathbf{Z};\mathbf{\Sigma}_1\right) = -\log\left|\mathbf{\Sigma}_1\right| - \mathrm{tr}\left\{\mathbf{\Sigma}_1^{-1}\mathbf{J}_1\hat{\mathbf{\Sigma}}_1\mathbf{J}_1\right\}. \tag{6.18}$$

Adding (6.17) and (6.18), gives us the forward-backward log-likelihood as:

$$\frac{1}{2}\log f_{\mathbf{z}}^{(FB)}\left(\mathbf{Z};\mathbf{\Sigma}_1\right) = -\log\left|\mathbf{\Sigma}_1\right| - \mathrm{tr}\left\{\mathbf{\Sigma}_1^{-1}\hat{\mathbf{\Sigma}}_{\mathrm{PS},1}\right\}, \tag{6.19}$$

where

$$\hat{\mathbf{\Sigma}}_{\mathrm{PS},1} = \frac{1}{2}\left(\hat{\mathbf{\Sigma}}_1 + \mathbf{J}_1\hat{\mathbf{\Sigma}}_1\mathbf{J}_1\right). \tag{6.20}$$

The covariance matrix estimator (6.20) is called forward-backward sample covariance matrix [97]. An exact theory indicating the performance of the estimator as a function of number of independent vectors $\mathbf{z}(m)$, $m = 1, 2, \cdots, M$ to estimate the covariance matrix $\mathbf{\Sigma}_1$ is not available. However, a qualitative discussion shows that the sample covariance matrix (6.20) allows the demand for required samples decreased by approximately a factor of two [80, 92]. Similarly, it is reported in [98] that the $\hat{\mathbf{\Sigma}}_{\mathrm{PS},1}$ has consistently lower variance than the variance of $\hat{\mathbf{\Sigma}}_1$. Solving (6.15) and using (6.20), the final expression for the modified GLRT becomes

$$\Lambda_{\mathrm{PS}}\left(\mathbf{Z}\right) = \frac{\left|\hat{\mathbf{\Sigma}}_1 + \mathbf{J}_1\hat{\mathbf{\Sigma}}_1\mathbf{J}_1\right|}{\left|\hat{\mathbf{\Sigma}}_0\right|} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \gamma. \tag{6.21}$$

where $\hat{\mathbf{\Sigma}}_0 = \hat{\mathbf{\Sigma}}_{\mathrm{A},0} \otimes \mathbf{I}_{N\times N}$. Please note that in the case of calibrated antennas the only difference lies in the denominator of (6.21) as instead of $\det\left|\hat{\mathbf{\Sigma}}_0\right|$, we will have $\frac{1}{KN}\mathrm{tr}\left(\hat{\mathbf{\Sigma}}_1\right)$.

Compared to the detection scheme in (6.10), the new one in (6.21) has an improved performance at small sample support the number of independent vector measurements required for the covariance matrix estimator can be decreased by up to a factor of two [80]. In Fig. 6.4, we compare the detector $\Lambda_{\mathrm{PS}}\left(\mathbf{Z}\right)$ in (6.21) with the frequency domain detector proposed in [34]. We can see that the frequency domain scheme performs slightly better than the time domain persymmetric detection scheme for sample size $M < 60$. In conclusion, we can say

that the exploitation of inherent structure of covariance matrix both in frequency and time domain, leads us to robustness against the small sample support compared to the ST-GLRT in (6.10).

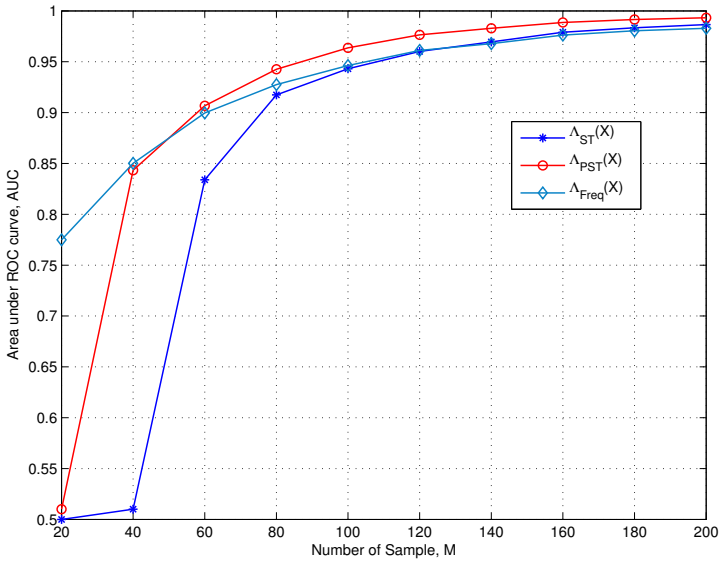


Figure 6.4: AUC curves: Comparison of the P-GLRT $\Lambda_{\text{PS}}(\mathbf{Z})$ with the ST-GLRT $\Lambda_{\text{ST}}(\mathbf{Z})$ for different sizes of the sample support (M), number of vector samples per sub-block $N = 15$, number of antennas $L = 4$ and average SNR$\bar{\kappa} = -8$dB

6.1.5 GLRT based on the Kronecker product

In Section 6.1.3, we have presented the ST-GLRT approach for the detection problem in (6.4) and argued that it degenerates due to the singularity issue that arises in detection problems with small samples sizes M and large LN. In order to counteract this issue, then we modified the ST-GLRT by exploiting the persymmetric properties of the covariance matrix. Similarly, we have also discussed that the spatio-temporal covariance matrix $\mathbf{\Sigma}_1$ has block-Toeplitz structure. Moreover, in [99, 100], it is reported that the block-Toeplitz structure can be factored by Kronecker product of two matrices. Taking this into effect, we factorize the covariance matrix $\mathbf{\Sigma}_1$ into a purely spatial and a purely temporal components based on the Kronecker product as [83]:

$$\mathbf{\Sigma}_1 = \mathbf{\Sigma}_{\text{A}} \otimes \mathbf{\Sigma}_{\text{T}}. \tag{6.22}$$

In (6.22), the sub-matrix $\mathbf{\Sigma}_{\text{A}}$ captures the spatial correlation between the observations received at different antennas and sub-matrix $\mathbf{\Sigma}_{\text{T}}$ captures the time correlation between N columns vector in block $\mathbf{X}$ where we have the relation vec $(\mathbf{X}) = \mathbf{z}$. Herein, we remark that the covariance structure (6.22) makes an implicit assumption that the temporal correlation structure remains the same at all spatial locations (elements of antenna). Similarly, the spatial correlation structure remains the same during time $1 : N$.

6.1.5.1 KR-GLRT

In this section, we apply the Kronecker product based factorization to derive the modified GLRT that can be expressed as:

$$\Lambda_{\mathrm{KR}}(\mathbf{Z}) = \frac{\max\limits_{\boldsymbol{\Sigma}_0} f_{\mathbf{z}}\left(\mathbf{Z}; \boldsymbol{\Sigma}_0\right)}{\max\limits_{\boldsymbol{\Sigma}_{\mathrm{T}}, \boldsymbol{\Sigma}_{\mathrm{A}}} f_{\mathbf{z}}\left(\mathbf{Z}; \boldsymbol{\Sigma}_{\mathrm{A}} \otimes \boldsymbol{\Sigma}_{\mathrm{T}}\right)} \gtrless_{\mathcal{H}_1}^{\mathcal{H}_0} \gamma. \tag{6.23}$$

In order to solve (6.23), under hypothesis $\mathcal{H}_1$, we need to estimate the unknown covariance matrices $\boldsymbol{\Sigma}_{\mathrm{A}}$ and $\boldsymbol{\Sigma}_{\mathrm{T}}$ by using maximum likelihood estimation (MLE) paradigm. Under $\mathcal{H}_1$, the likelihood function $f_{\mathbf{z}}\left(\mathbf{Z}; \boldsymbol{\Sigma}_{\mathrm{A}} \otimes \boldsymbol{\Sigma}_{\mathrm{T}}\right)$ can be written as:

$$\begin{aligned} f_{\mathbf{z}}\left(\mathbf{Z}; \boldsymbol{\Sigma}_{\mathrm{A}} \otimes \boldsymbol{\Sigma}_{\mathrm{T}}\right) = (\pi)^{-\frac{1}{2}MLN} \left|\boldsymbol{\Sigma}_{\mathrm{A}} \otimes \boldsymbol{\Sigma}_{\mathrm{T}}\right|^{-M} \\ \exp\left[-\sum_{m=1}^{M} \mathbf{z}(m) \left(\boldsymbol{\Sigma}_{\mathrm{A}}^{-1} \otimes \boldsymbol{\Sigma}_{\mathrm{T}}^{-1}\right) \mathbf{z}^{H}(m)\right], \end{aligned} \tag{6.24}$$

Now we can write $f_{\mathbf{z}}\left(\mathbf{Z}; \boldsymbol{\Sigma}_{\mathrm{A}} \otimes \boldsymbol{\Sigma}_{\mathrm{T}}\right) = f_{\mathbf{x}}\left(\mathbf{X}_M; \boldsymbol{\Sigma}_{\mathrm{T}}, \boldsymbol{\Sigma}_{\mathrm{A}}\right)$ [89] and we have the expression

$$\begin{aligned} f_{\mathbf{x}}\left(\mathbf{X}_M; \boldsymbol{\Sigma}_{\mathrm{A}}, \boldsymbol{\Sigma}_{\mathrm{T}}\right) = (\pi)^{-\frac{1}{2}MLN} \left|\boldsymbol{\Sigma}_{\mathrm{A}}\right|^{-MN} \left|\boldsymbol{\Sigma}_{\mathrm{T}}\right|^{-ML} \\ \exp\left[-\sum_{m=1}^{M} \mathrm{tr}\left\{\boldsymbol{\Sigma}_{\mathrm{T}}^{-1} \mathbf{X}^{H}(m) \boldsymbol{\Sigma}_{\mathrm{A}}^{-1} \mathbf{X}(m)\right\}\right], \end{aligned} \tag{6.25}$$

To find the matrices $\boldsymbol{\Sigma}_{\mathrm{A}}$ and $\boldsymbol{\Sigma}_{\mathrm{T}}$ that satisfy $\max\limits_{\boldsymbol{\Sigma}_{\mathrm{T}}, \boldsymbol{\Sigma}_{\mathrm{A}}} f_{\mathbf{X}}\left(\mathbf{X}_M; \boldsymbol{\Sigma}_{\mathrm{A}}, \boldsymbol{\Sigma}_{\mathrm{T}}\right)$, a common approach is to take derivative of $\log f_{\mathbf{X}}\left(\mathbf{X}_M; \boldsymbol{\Sigma}_A, \boldsymbol{\Sigma}_{\mathrm{T}}\right)$

$$\begin{aligned} \log f_{\mathbf{X}}\left(\mathbf{X}_M; \boldsymbol{\Sigma}_A, \boldsymbol{\Sigma}_{\mathrm{T}}\right) = -N \log\left|\boldsymbol{\Sigma}_{\mathrm{A}}\right| - L \log\left|\boldsymbol{\Sigma}_{\mathrm{T}}\right| \\ - \frac{1}{M} \sum_{m=1}^{M} \mathrm{tr}\left\{\boldsymbol{\Sigma}_{\mathrm{T}}^{-1} \mathbf{X}^{H}(m) \boldsymbol{\Sigma}_{\mathrm{A}}^{-1} \mathbf{X}(m)\right\} \end{aligned} \tag{6.26}$$

with respect to $\boldsymbol{\Sigma}_{\mathrm{A}}\left(\boldsymbol{\Sigma}_{\mathrm{T}}\right)$ keeping $\boldsymbol{\Sigma}_{\mathrm{T}}\left(\boldsymbol{\Sigma}_{\mathrm{A}}\right)$ fixed. Equating the result of the derivative to zero and after simple mathematical operations, the estimators under the hypothesis $\mathcal{H}_1$ can be written as [83]:

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{T}} = \frac{1}{LM} \sum_{m=1}^{M} \mathbf{X}^{T}(m) \hat{\boldsymbol{\Sigma}}_{\mathrm{A}}^{-1} \mathbf{X}(m), \tag{6.27}$$

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{A}} = \frac{1}{NM} \sum_{m=1}^{M} \mathbf{X}(m) \hat{\boldsymbol{\Sigma}}_{\mathrm{T}}^{-1} \mathbf{X}^{T}(m). \tag{6.28}$$

Expression (6.27) and (6.28) suggest that $\hat{\boldsymbol{\Sigma}}_{\mathrm{T}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{A}}$ can be achieved using an iterative method such as the Flip-Flop algorithm. The Flip-Flop algorithm is obtained by alternately maximizing $\log f_{\mathbf{x}}\left(\mathbf{X}_M; \boldsymbol{\Sigma}_{\mathrm{A}}, \boldsymbol{\Sigma}_{\mathrm{T}}\right)$ w.r.t. $\boldsymbol{\Sigma}_{\mathrm{A}}$ keeping the last available estimate of $\boldsymbol{\Sigma}_{\mathrm{T}}$ fixed and viceversa. In [83], numerical experiments have been reported which indicate that the Flip-Flop algorithm performs very well and is much faster than a more general purpose

Algorithm 6.1 ML based Non-Iterative Flip-Flop

- Choose a starting value for $\hat{\Sigma}_{\mathrm{A}}^{0}$ as $\mathbf{I}_{L\times L}$
- Estimate $\hat{\mathbf{\Sigma}}_{\mathrm{T}}^{1}$ from (6.27) with $\hat{\mathbf{\Sigma}}_{\mathrm{A}} = \hat{\mathbf{\Sigma}}_{\mathrm{A}}^{0}$.
- Find the following
  1. Estimate $\hat{\mathbf{\Sigma}}_{\mathrm{A}}$ from (6.28) with $\hat{\mathbf{\Sigma}}_{\mathrm{T}}^{1}$.
  2. Estimate $\hat{\mathbf{\Sigma}}_{\mathrm{T}}$ from (6.27) with $\hat{\mathbf{\Sigma}}_{\mathrm{A}}$ from previous step.

optimization algorithm such as Newton–Raphson [83]. In [82], it has been discussed that for the case of large enough M, there is no need to iterate the algorithm. Taking into account this fact, we adopt non-iterative Flip-Flop approach and only perform the the steps given in Algorithm 6.2. To begin with, we use an initial value of $\hat{\mathbf{\Sigma}}_{\mathrm{A}}^{0} = \mathbf{I}_{L\times L}$. On the other hand under $\mathcal{H}_0$, we have the estimate of $\mathbf{\Sigma}_0$ as $\hat{\mathbf{\Sigma}}_0 = \hat{\mathbf{\Sigma}}_{\mathrm{A},0} \otimes \mathbf{I}_{N\times N}$. Having all of the maximum likelihood based estimates, and solving (6.23), we can get the expression:

$$\Lambda_{\mathrm{KR}}(\mathbf{Z}) = \frac{\left|\hat{\mathbf{\Sigma}}_{\mathrm{T}}\right|^{L}\left|\hat{\mathbf{\Sigma}}_{\mathrm{A}}\right|^{N}}{\left|\hat{\mathbf{\Sigma}}_{0}\right|} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \gamma. \tag{6.29}$$

Note that in the case of calibrated antennas the only difference is present in the denominator of (6.29) as instead of $\left|\hat{\mathbf{\Sigma}}_0\right|$, we will have $\frac{1}{LN}\mathrm{tr}\left(\hat{\mathbf{\Sigma}}_1\right)$.

The main advantage of the proposed GLRT (6.29) over the traditional is that under $\mathcal{H}_1$ instead of $\frac{1}{2}LN(LN+1)$ parameters, it has only $\frac{1}{2}L\left(L+1\right)+\frac{1}{2}N\left(N+1\right)$ parameters to estimate. Furthermore, the dimensions of these two covariance matrices $\mathbf{\Sigma}_{\mathrm{T}}$ and $\mathbf{\Sigma}_{\mathrm{A}}$ are much smaller than the dimension of full covariance matrix $\mathbf{\Sigma}_1$, that is why the computations are much less demanding. Another important point to note that the detector (6.29) takes into account the Kronecker structure as an extra feature jointly with spatio-temporal structure and the energy in the received observations. Keeping these facts in mind, the KR-GLRT (6.29) indeed promises a performance improvement compared to the unstructured ST-GLRT (6.10).

6.1.5.2 Flip-Flop Algorithm with Least Square Estimation

In every iteration of algorithm 6.2, to calculate (6.28) and (6.27), the inverse of the covariance matrices is required which involves a significant complexity. Keeping this in mind and using results of [101, Theorem 4.1], we propose the Flip-Flop algorithm based on Least Square Estimation (LSE) to find $\hat{\mathbf{\Sigma}}_{\mathrm{T}}$ and $\hat{\mathbf{\Sigma}}_{\mathrm{A}}$. Since, we know that the GLRT is meant to find the unknown covariance matrix with the help of the MLE, it would be therefore an abuse of language to call the detector (6.29) with the LSE as a GLRT. It can be rather considered as an

ad-hoc alternative. The expressions for $\hat{\mathbf{\Sigma}}_{\mathrm{T}}$ and $\hat{\mathbf{\Sigma}}_{\mathrm{A}}$ in LS paradigm can be written as:

$$\hat{\mathbf{\Sigma}}_{\mathrm{T}} = \frac{(M-1)^{-1}}{\mathrm{Tr}\left(\hat{\mathbf{\Sigma}}_{\mathrm{A}}^{2}\right)} \sum_{m=1}^{M} \mathbf{X}^{H}(m)\, \hat{\mathbf{\Sigma}}_{\mathrm{A}} \mathbf{X}(m), \tag{6.30}$$

$$\hat{\mathbf{\Sigma}}_{\mathrm{A}} = \frac{(M-1)^{-1}}{\mathrm{Tr}\left(\hat{\mathbf{\Sigma}}_{\mathrm{T}}^{2}\right)} \sum_{m=1}^{M} \mathbf{X}(m)\, \hat{\mathbf{\Sigma}}_{\mathrm{T}} \mathbf{X}^{H}(m), \tag{6.31}$$

where $\mathrm{Tr}(\mathbf{A})$ denotes the trace of matrix $\mathbf{A}$. In the LS case, the Kronecker product of $\mathbf{\Sigma}_{\mathrm{A}}$ and $\mathbf{\Sigma}_{\mathrm{T}}$ is actually fitted to the sample covariance matrix by minimizing the difference in Frobenius norm with respect to $\mathbf{\Sigma}_{\mathrm{A}}$ and $\mathbf{\Sigma}_{\mathrm{T}}$ as:

$$\min_{\mathbf{\Sigma}_{\mathrm{A}}, \mathbf{\Sigma}_{\mathrm{T}}} \left\| \hat{\mathbf{\Sigma}}_{1} - \mathbf{\Sigma}_{\mathrm{A}} \otimes \mathbf{\Sigma}_{\mathrm{T}} \right\|_{F}^{2}. \tag{6.32}$$

The derivation process of (C.8) and (C.9) based on solving (6.32) is given in the Appendix C. Moreover, the steps in the algorithm based on the LS are the same as given in Algorithm 6.2. The only difference is that instead of (6.27) and (6.28), it uses (C.8) and (C.9). Furthermore, LS paradigm can be used without assuming any prior information or that the probability distribution of the received signal should be known.

In Fig. 6.5, we compare the performance of the KR-GLRT, while estimating the two matrices $\mathbf{\Sigma}_{\mathrm{A}}$ and $\mathbf{\Sigma}_{\mathrm{T}}$ with the help of MLE and LSE, respectively. In order to plot these ROC curves, we assume $L = 4$ antennas with SNRs $\{-5.8\text{dB}, -5.65\text{dB}, -5.5\text{dB}, -6.0\text{dB}\}$. The received mean power at the 4 antennas is 0dB. The rest of the parameters are given in the caption of the figure. From this experiment, we find that in the uncalibrated case where noise powers at different antennas are nonidentical, the MLE slightly outperforms the LSE methodology. However, the LSE based ad hoc methodology has slightly less computational complexity.

6.1.6 GLRT based on the Kronecker and persymmetric structure

In section 6.1.4, we have assumed that the covariance matrix $\mathbf{\Sigma}_{1}$ has a persymmetric-block-Toeplitz structure. Keeping this in mind, in this section we assess the possible improvement in the detection performance of (6.29) by exploiting the fact that the factored matrices $\mathbf{\Sigma}_{\mathrm{T}}$ and $\mathbf{\Sigma}_{\mathrm{A}}$ have individual persymmetric structure. We show that it is possible to account for the persymmetric structure by a simple modification of the flip-flop algorithm. Hence, for the estimation of the covariance matrix under $\mathcal{H}_1$, we explore both the Kronecker structure and the persymmetric structures. Once we know $\mathbf{\Sigma}_{\mathrm{T}}$ and $\mathbf{\Sigma}_{\mathrm{A}}$ are having the persymmetric structure, then the maximum likelihood based estimators for the persymmetric Kronecker based GLRT (6.35) are found by considering the following constraints under hypothesis $\mathcal{H}_1$,

$$\mathbf{\Sigma}_{\mathrm{A}} = \mathbf{J}_{\mathrm{L}} \mathbf{\Sigma}_{\mathrm{A}} \mathbf{J}_{\mathrm{L}}, \tag{6.33}$$

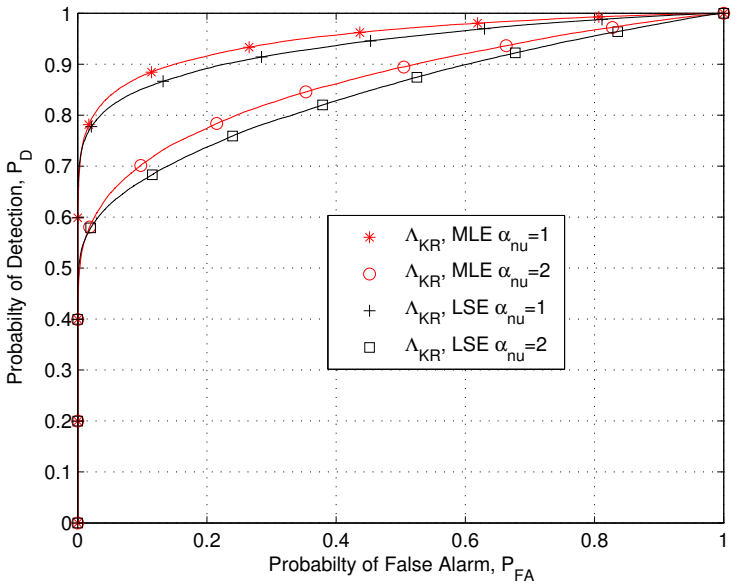


Figure 6.5: ROC Curves: Comparison of the KR-GLRT (MLE) with the KR-GLRT (LSE) for number of vector samples per sub-block $N = 10$, number of antennas $L = 4$, sample size$M = 40$, noise uncertainty $\alpha_{nu} = 1$ and shadowing effect $\sigma_{\text{dB-Spread}} = 4$.

$$\boldsymbol{\Sigma}_{\mathrm{T}} = \mathbf{J}_{\mathrm{N}}\boldsymbol{\Sigma}_{\mathrm{T}}\mathbf{J}_{\mathrm{N}}, \tag{6.34}$$

where $\mathbf{J}_{\mathrm{L}}$ and $\mathbf{J}_{\mathrm{N}}$ are the reversal matrices of dimensions $(L \times L)$ and $(N \times N)$, respectively. The modified version of KR-GLRT (6.23) that we denote as PK-GLRT here, can be written as:

$$\begin{aligned} \Lambda_{\mathrm{PK}}(\mathbf{Z}) &= \frac{\max\limits_{\boldsymbol{\Sigma}_0} f_{\mathbf{z}}\left(\mathbf{Z}; \boldsymbol{\Sigma}_0\right)}{\max\limits_{\boldsymbol{\Sigma}_{\mathrm{T}}, \boldsymbol{\Sigma}_{\mathrm{A}}} f_{\mathbf{z}}\left(\mathbf{Z}; \boldsymbol{\Sigma}_{\mathrm{T}} \otimes \boldsymbol{\Sigma}_{\mathrm{A}}\right)} \\ \text{s.t} \quad & \boldsymbol{\Sigma}_{\mathrm{A}} = \mathbf{J}_{\mathrm{L}}\boldsymbol{\Sigma}_{\mathrm{A}}\mathbf{J}_{\mathrm{L}} \\ & \boldsymbol{\Sigma}_{\mathrm{T}} = \mathbf{J}_{\mathrm{N}}\boldsymbol{\Sigma}_{\mathrm{T}}\mathbf{J}_{\mathrm{N}} \end{aligned} \tag{6.35}$$

We see that in the expression (6.35), under hypothesis $\mathcal{H}_0$ the solution is the same as in the cases of (6.15), (6.21) and (6.23). The difference lies in the case of hypothesis $\mathcal{H}_1$, where we need to solve

$$\begin{aligned} & \max_{\boldsymbol{\Sigma}_{\mathrm{A}}, \boldsymbol{\Sigma}_{\mathrm{T}}} \log f_{\mathbf{Z}}\left(\mathbf{Z}; \boldsymbol{\Sigma}_{\mathrm{T}} \otimes \boldsymbol{\Sigma}_{\mathrm{A}}\right), \\ & \text{s.t} \quad \begin{aligned} \boldsymbol{\Sigma}_{\mathrm{A}} &= \mathbf{J}_{\mathrm{L}}\boldsymbol{\Sigma}_{\mathrm{A}}\mathbf{J}_{\mathrm{L}} \\ \boldsymbol{\Sigma}_{\mathrm{T}} &= \mathbf{J}_{\mathrm{N}}\boldsymbol{\Sigma}_{\mathrm{T}}\mathbf{J}_{\mathrm{N}} \end{aligned} \end{aligned} \tag{6.36}$$

In (6.36), we used the fact that maximum of $\log f_{\mathbf{z}}\left(\mathbf{Z}; \boldsymbol{\Sigma}_{\mathrm{T}} \otimes \boldsymbol{\Sigma}_{\mathrm{A}}\right)$ and $f_{\mathbf{z}}\left(\mathbf{Z}; \boldsymbol{\Sigma}_{\mathrm{T}} \otimes \boldsymbol{\Sigma}_{\mathrm{A}}\right)$ is placed at the same value of the unknown parameters to be estimated [25]. To find the maximizer with the constraints $\boldsymbol{\Sigma}_{\mathrm{A}} = \mathbf{J}_{\mathrm{L}}\boldsymbol{\Sigma}_{\mathrm{A}}\mathbf{J}_{\mathrm{L}}$ and $\boldsymbol{\Sigma}_{\mathrm{T}} = \mathbf{J}_{\mathrm{N}}\boldsymbol{\Sigma}_{\mathrm{T}}\mathbf{J}_{\mathrm{N}}$ we need to develop the likelihood function in the form of forward-backward (FB) log-likelihood. From Section 6.1.4 we know that the FB is combination of the forward looking and the backward looking log-likelihood. In the case of using Kronecker product, the forward looking log-likelihood

function for estimating $\mathbf{\Sigma}_{\mathrm{T}}$ can be written as:

$$\begin{aligned} \log f_{\mathbf{z}}^{F}\left(\mathbf{Z};\mathbf{\Sigma}_{\mathrm{T}}\otimes\mathbf{\Sigma}_{\mathrm{A}}\right) &= -N\log\left|\mathbf{\Sigma}_{\mathrm{A}}\right| - L\log\left|\mathbf{\Sigma}_{\mathrm{T}}\right| \\ &\quad - \frac{1}{M}\sum_{m=1}^{M}\mathrm{tr}\left\{\mathbf{\Sigma}_{\mathrm{T}}^{-1}\mathbf{X}^{H}\left(m\right)\mathbf{\Sigma}_{\mathrm{A}}^{-1}\mathbf{X}\left(m\right)\right\}. \end{aligned} \tag{6.37}$$

or, the forward looking log-likelihood function for estimating $\mathbf{\Sigma}_{\mathrm{A}}$ can be written as:

$$\begin{aligned} \log f_{\mathbf{z}}^{F}\left(\mathbf{Z};\mathbf{\Sigma}_{\mathrm{T}}\otimes\mathbf{\Sigma}_{\mathrm{A}}\right) &= -N\log\left|\mathbf{\Sigma}_{\mathrm{A}}\right| - L\log\left|\mathbf{\Sigma}_{\mathrm{T}}\right| \\ &\quad - \frac{1}{M}\sum_{m=1}^{M}\mathrm{tr}\left\{\mathbf{\Sigma}_{\mathrm{A}}^{-1}\mathbf{X}\left(m\right)\mathbf{\Sigma}_{\mathrm{T}}^{-1}\mathbf{X}^{H}\left(m\right)\right\}. \end{aligned} \tag{6.38}$$

Similarly, using the discussion in 6.1.4, the backward looking log-likelihood can be written

$$\begin{aligned} \log f_{\mathbf{z}}^{B}\left(\mathbf{Z};\mathbf{\Sigma}_{\mathrm{T}}\otimes\mathbf{\Sigma}_{\mathrm{A}}\right) &= -N\log\left|\mathbf{\Sigma}_{\mathrm{A}}\right| - L\log\left|\mathbf{\Sigma}_{\mathrm{T}}\right| \\ &\quad - \frac{1}{M}\sum_{m=1}^{M}\mathrm{tr}\left\{\mathbf{\Sigma}_{\mathrm{T}}^{-1}\mathbf{J}_{\mathrm{N}}\mathbf{X}^{H}\left(m\right)\mathbf{\Sigma}_{\mathrm{A}}^{-1}\mathbf{X}\left(m\right)\mathbf{J}_{\mathrm{N}}\right\}. \end{aligned} \tag{6.39}$$

or

$$\begin{aligned} \log f_{\mathbf{z}}^{B}\left(\mathbf{Z};\mathbf{\Sigma}_{\mathrm{T}}\otimes\mathbf{\Sigma}_{\mathrm{A}}\right) &= -N\log\left|\mathbf{\Sigma}_{\mathrm{A}}\right| - L\log\left|\mathbf{\Sigma}_{\mathrm{T}}\right| \\ &\quad - \frac{1}{M}\sum_{m=1}^{M}\mathrm{tr}\left\{\mathbf{\Sigma}_{\mathrm{A}}^{-1}\mathbf{J}_{\mathrm{L}}\mathbf{X}\left(m\right)\mathbf{\Sigma}_{\mathrm{T}}^{-1}\mathbf{X}^{H}\left(m\right)\mathbf{J}_{\mathrm{L}}\right\}. \end{aligned} \tag{6.40}$$

Adding (6.37) and (6.39), or (6.38) and (6.40) give us $\log f_{\mathbf{z}}^{BF}\left(\mathbf{Z};\mathbf{\Sigma}_{\mathrm{T}}\otimes\mathbf{\Sigma}_{\mathrm{A}}\right)$, the Kronecker product based forward-backward log-likelihood. As mentioned before the approach is to use a cyclic maximization and alternate between maximizing with respect to $\mathbf{\Sigma}_{\mathrm{T}}$ and $\mathbf{\Sigma}_{\mathrm{A}}$, while keeping the other variable fixed. In the first step, we fix $\mathbf{\Sigma}_{\mathrm{A}}$ and find $\mathbf{\Sigma}_{\mathrm{T}}$ that maximizes $\log f_{\mathbf{z}}^{BF}\left(\mathbf{Z};\mathbf{\Sigma}_{\mathrm{T}}\otimes\mathbf{\Sigma}_{\mathrm{A}}\right)$ (obtained through adding (6.37) and (6.39)). Taking into account the results in [89] and the concept of forward-backward estimation in Section 6.1.4, the estimator of the persymmetric $\mathbf{\Sigma}_{\mathrm{T}}$ can be found as :

$$\hat{\mathbf{\Sigma}}_{\mathrm{PS,T}} = \frac{1}{2ML}\sum_{m=1}^{M}\left\{\mathbf{X}^{H}\left(m\right)\hat{\mathbf{\Sigma}}_{\mathrm{PS,A}}^{-1}\mathbf{X}\left(m\right) + \mathbf{J}_{\mathrm{N}}\left(\mathbf{X}^{H}\left(m\right)\hat{\mathbf{\Sigma}}_{\mathrm{PS,A}}^{-1}\mathbf{X}\left(m\right)\right)\mathbf{J}_{\mathrm{N}}\right\}. \tag{6.41}$$

Similarly, by following the same process with fixed $\mathbf{\Sigma}_{\mathrm{T}}$, the estimator of the persymmetric $\mathbf{\Sigma}_{\mathrm{A}}$ can be written as:

$$\hat{\mathbf{\Sigma}}_{\mathrm{PS,A}} = \frac{1}{2MN}\sum_{m=1}^{M}\left\{\mathbf{X}\left(m\right)\hat{\mathbf{\Sigma}}_{\mathrm{PS,T}}^{-1}\mathbf{X}^{H}\left(m\right) + \mathbf{J}_{\mathrm{L}}\left(\mathbf{X}\left(m\right)\hat{\mathbf{\Sigma}}_{\mathrm{PS,T}}^{-1}\mathbf{X}^{H}\left(m\right)\right)\mathbf{J}_{\mathrm{L}}\right\}. \tag{6.42}$$

As it was in the case of expressions (6.27) and (6.28), both of the expressions (6.41) and (6.42) suggest that $\hat{\mathbf{\Sigma}}_{\mathrm{PS,T}}$ and $\hat{\mathbf{\Sigma}}_{\mathrm{PS,A}}$ can be estimated using an iterative method such as the Flip-Flop algorithm, as shown in algorithm 6.2. Considering (6.41) and (6.42), the final expression for

Algorithm 6.2 Non-Iterative Flip-Flop (persymmetric)

- Choose a starting value for $\hat{\mathbf{\Sigma}}_{\text{A}}^{0}$ as $\mathbf{I}_{L\times L}$
- Estimate $\hat{\mathbf{\Sigma}}_{\text{T}}^{1}$ from (6.27) with $\hat{\mathbf{\Sigma}}_{\text{A}}^{1}=\hat{\mathbf{\Sigma}}_{\text{A}}^{0}$.
- Find the following
  1. Estimate $\hat{\mathbf{\Sigma}}_{\text{PS,A}}$ from (6.42) with $\hat{\mathbf{\Sigma}}_{\text{T}}^{1}$.
  2. Estimate $\hat{\mathbf{\Sigma}}_{\text{PS,T}}$ from (6.41) with $\hat{\mathbf{\Sigma}}_{\text{PS,A}}$ from step 1.

the GLRT function (6.35) can be written as:

$$\Lambda_{\text{PK}}(\mathbf{Z})=\frac{\left|\hat{\mathbf{\Sigma}}_{\text{PS,T}}\right|^{K}\left|\hat{\mathbf{\Sigma}}_{\text{PS,A}}\right|^{N}}{\left|\hat{\mathbf{\Sigma}}_{0}\right|}\gtrless_{\mathcal{H}_1}^{\mathcal{H}_0}\gamma. \tag{6.43}$$

The advantage of (6.43) over (6.29) is that under $\mathcal{H}_1$ instead of $\frac{1}{2}L(L+1)+\frac{1}{2}N(N+1)$ parameters, it has approximately only $(2L-1)+(2N-1)$ parameters to estimate. In conclusion, by exploiting the underlying structure of the covariance matrix $\mathbf{\Sigma}_1$ via the persymmetric ML estimates of covariance matrices $\mathbf{\Sigma}_{\text{T}}$ and $\mathbf{\Sigma}_{\text{A}}$, it further increases the robustness of (6.29) at small sample support. Here once again we remark that in the case of calibrated case the only difference is present in the denominator of (6.43) as instead of $\left|\hat{\mathbf{\Sigma}}_{0}\right|$, we will have $\frac{1}{KN}\text{tr}\left(\hat{\mathbf{\Sigma}}_{1}\right)$.

6.1.7 Numerical results

In this Section we present numerical results to analyze the proposed detection schemes that are discussed in the preceding sections. For the analysis to be conducted herein, we use the receiver operating characteristic ROC curve and the area under the ROC curve (AUC), which varies between 0.5 (poor performance) and 1 (good performance) [27].

For the assessment of the detectors being considered in this chapter, the SNRs at the different antennas are allocated randomly. We use the average SNR of all antennas as: $\bar{\kappa}=\frac{1}{K}\sum_{i=1}^{L}\kappa_i$ and the SNR of i^{th} antenna is $\kappa_i=\zeta^{i-1}\kappa_{\text{min}}$ where κ_{min} is the minimum SNR among those of the L antennas. For a given average SNR $\bar{\kappa}$ and SNR gap ζ, we can generate the SNRs of the antennas, though the match of the SNRs to the sensors can be random. Moreover, we assume that the receiver is equipped with $L=4$ antennas and the number of vector samples in a sub-block is $N=15$. The remaining parameters for every experiment are described in the captions of the corresponding diagrams, accordingly.

In in the first experiment, we simulate with help of the Matlab, the ROC curves in order to compare the performance of the proposed schemes to the traditional schemes. In this experiment we plot ROC curves for two different cases. In the first case we assume that the unknown noise powers at different antennas are perfectly estimated and there is no noise power uncertainty. The ROC plots are given in Fig.6.6. In the second part of this experiment shown in Fig. 6.7, we repeat the same experimental setup for the case when the noise power

uncertainty is present. Note that we model the noise power uncertainty by generating the noise power at the $\{l^{th}\}_{l=1}^{L}$ antenna as $\sigma_{w,l}^2 \sim \mathcal{U}\left(\frac{\sigma_{n,l}^2}{\alpha_{nu}}, \alpha_{nu}\sigma_{n,l}^2\right)$, where $\alpha_{nu} \geq 1$ and $\alpha_{nu} = 1$ means no noise uncertainty [15]. From the results it is clear that the proposed schemes clearly outperform traditional schemes. Moreover, from the experiment we can also conclude that the noise power uncertainty slightly deteriorates the performance of all detectors. To further analyze the effect of the sample support, noise power uncertainty and shadowing parameters, we need to have a single and quantitative figure of merit. Hence, next we use AUC curves to see these effects on the detection performance of the spectrum sensing schemes

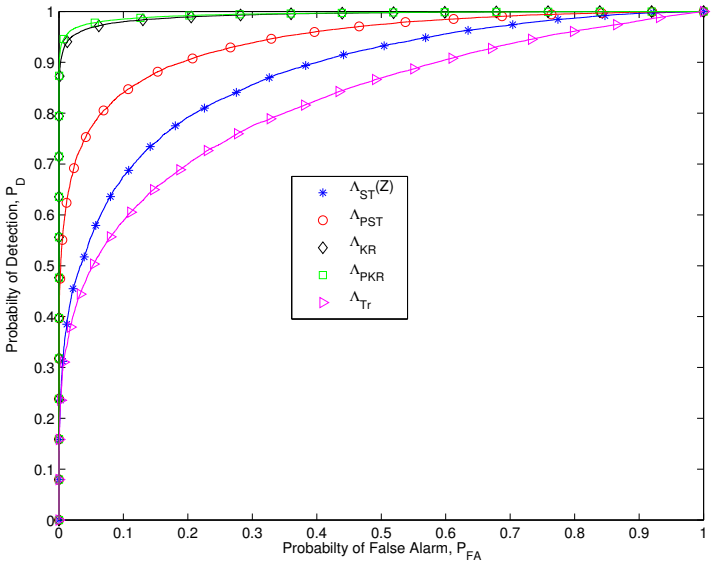


Figure 6.6: ROC curves: Comparison of detection schemes for sample support size $M = 100$, number of vector samples per sub-block $N = 15$, number of antennas $L = 4$, shadowing effect $\sigma_{\text{dB}-\text{Spread}} = 4$, noise uncertainty$\alpha_{nu} = 1$ and average SNR $\bar{\kappa} = -8$dB.

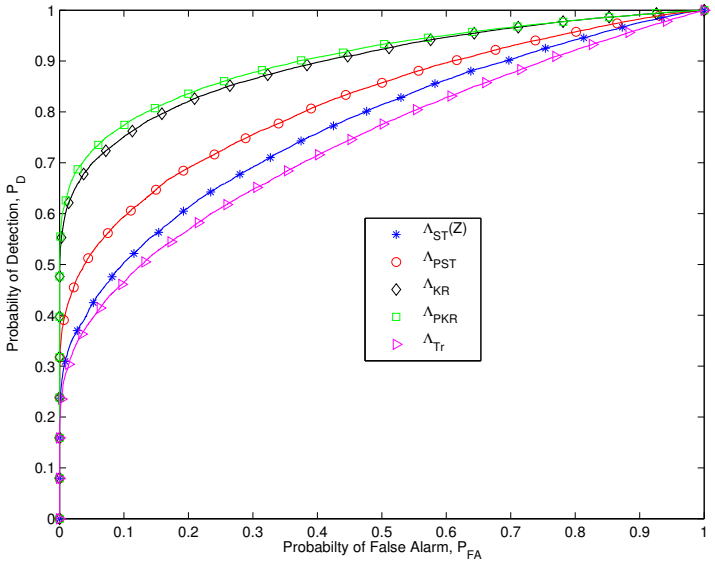


Figure 6.7: ROC curves: Comparison of detection schemes for sample support size $M = 100$, number of vector samples per sub-block $N = 15$, number of antennas $L = 4$, shadowing effect $\sigma_{\text{dB}-\text{Spread}} = 4$, noise uncertainty$\alpha_{nu} = 2$ and average SNR $\bar{\kappa} = -8$dB.

In Fig. 6.8, we plot the AUC plots to analyze the effects of sample size in the presence shadowing, with and without the effects of of noise power uncertainty. In Fig. 6.8, the curves with dashed lines represent the case where both shadowing and the effects of noise power uncertainty, $\alpha_{nu} = 1.5$. On the other hand the curves with solid lines show the case with no noise power uncertainty (i.e. $\alpha_{nu} = 1$). From these plots, it can be concluded that the proposed schemes $\Lambda_{\text{KR}}(\mathbf{Z})$ and $\Lambda_{\text{PK}}(\mathbf{Z})$ are robust against the small sample support both in the presence and absence of noise power uncertainty. Particularly, in the region $20 \leq M \leq 80$, the detectors $\Lambda_{\text{KR}}(\mathbf{Z})$ and $\Lambda_{\text{PK}}(\mathbf{Z})$ clearly outperform the other detectors. We can also see that in the small sample regime $\Lambda_{\text{PK}}(\mathbf{Z})$ performs better than $\Lambda_{\text{KR}}(\mathbf{Z})$. The obvious reason for this is that under $\mathcal{H}_1$ instead of $\frac{1}{2}L(L+1)+\frac{1}{2}N(N+1)$ parameters in the case of $\Lambda_{\text{KR}}(\mathbf{Z})$, the detection scheme $\Lambda_{\text{PK}}(\mathbf{Z})$ has approximately only $(2L-1)+(2N-1)$ parameters to estimate. For $N = 15$, $L = 4$, Table 6.1 lists the number of parameters to be estimated by different spectrum sensing scheme under hypothesis $\mathcal{H}_1$.

Table 6.1: For $N = 15$, $L = 4$, number of parameters to be estimated for the covariance matrix under hypothesis $\mathcal{H}_1$.

S.No	Detection scheme	Number of parameters to be estimated
1	$\Lambda_{\text{ST}}(\mathbf{Z})$	1836
3	$\Lambda_{\text{KR}}(\mathbf{Z})$	130
4	$\Lambda_{\text{PK}}(\mathbf{Z})$	36

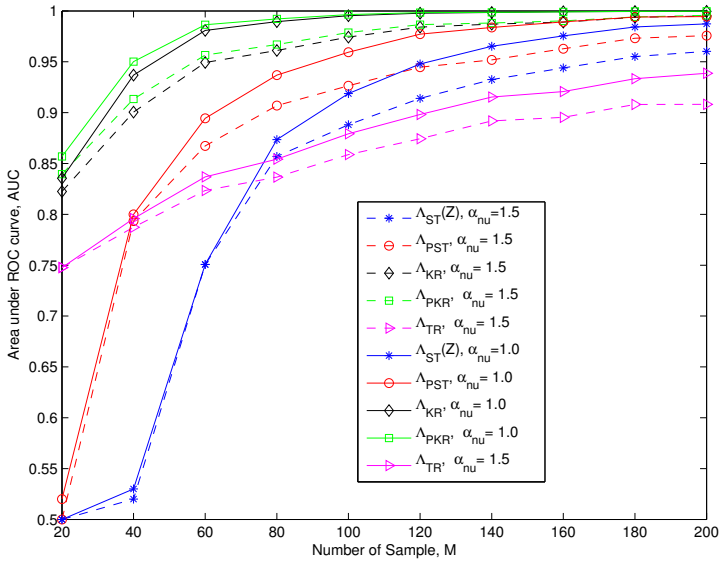


Figure 6.8: AUC curves: To asses the effects of number of samples M, with number of vector samples per sub-block $N = 15$, number of antennas $L = 4$, shadowing $\sigma_{\text{dB}-\text{Spread}} = 4$, and average SNR$\bar{\kappa} = -8$dB.

In order to confirm the results listed in the Table 6.1, we simulate the normalized minimum square error (MSE) of the estimator of the covariance matrix under hypothesis $\mathcal{H}_1$, expressed as :

$$\text{RMSE} = \sqrt{\frac{1}{J_{\text{avg}}} \sum_{j=2}^{J_{\text{avg}}} \frac{\left\| \hat{\boldsymbol{\Sigma}}_1 - \boldsymbol{\Sigma}_1 \right\|_F^2}{\|\boldsymbol{\Sigma}_1\|_F^2}}. \tag{6.44}$$

The plot of the normalized MSE is given in Fig 6.9.

In Fig.6.10, we show the AUC plots to analyze the effect of shadowing (i.e. $\sigma_{\text{dB}-\text{Spread}}$) both in the presence and absence of noise uncertainty, $\alpha_{nu} = 1.5$ and $\alpha_{nu} = 1$, respectively. It is evident from thee results that the effects of shadowing are very small on the performance of the detection schemes. However, we can see that incrementing $\sigma_{\text{dB}-\text{Spread}}$, a slight improvement occurs in the performance of the detection schemes $\Lambda_{\text{TR}}(\mathbf{Z})$, $\Lambda_{\text{ST}}(\mathbf{Z})$ and $\Lambda_{\text{PS}}(\mathbf{Z})$. The most obvious reason for this interesting outcome can be the heavy-tailed distribution of the primary signal strength due to the log-normally-distributed shadow fading that behave in such a way at lower SNR [73].

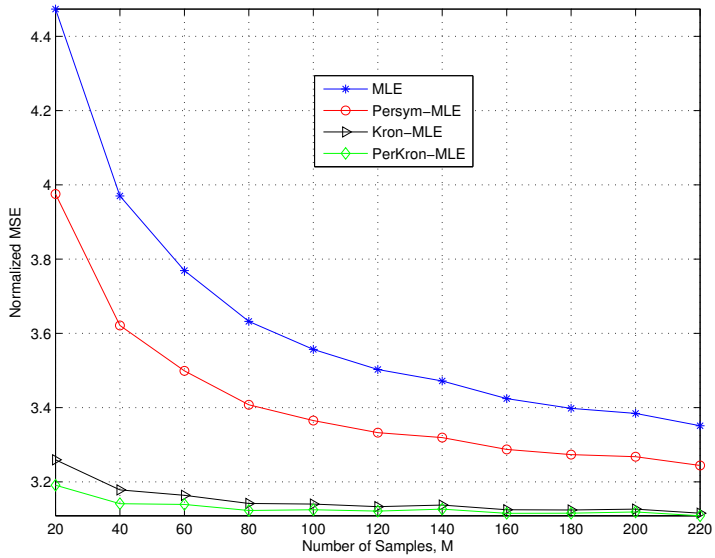


Figure 6.9: Normalized MSE of the estimator of the covariance matrix under the hypothesis $\mathcal{H}_1$.

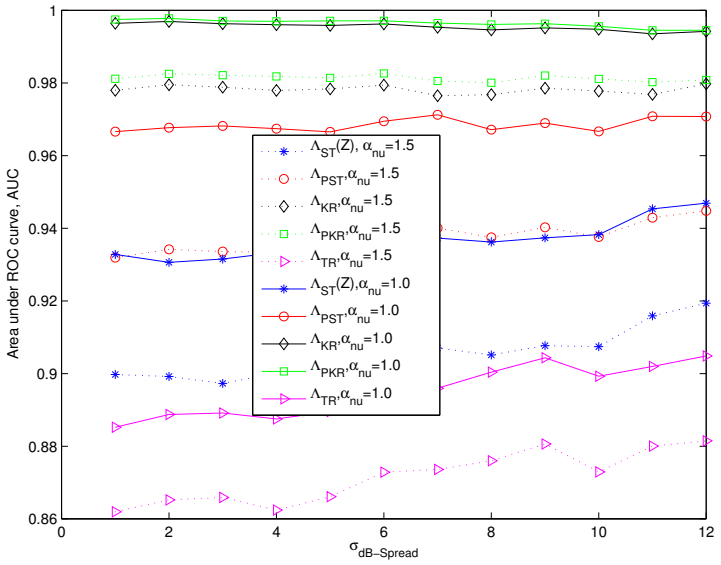


Figure 6.10: AUC curves: To asses the effect of shadowing $\sigma_{\mathrm{dB-Spread}}$ with sample size $M = 80$, number of vector samples per sub-block $N = 15$, number of antennas $L = 4$, noise uncertainty $\alpha_{nu} = 1.5$ and average SNR $\bar{\kappa} = -8$dB

In the final experiment that is depicted in Fig. 6.11, we show the AUC plots to analyze the effects of noise power uncertainty. The results show a robust behaviour for the detection schemes against the noise power uncertainty. Once again, we observe that the performance of the proposed schemes $\Lambda_{\mathrm{KR}}(\mathbf{Z})$ and $\Lambda_{\mathrm{PK}}(\mathbf{Z})$ is better than other schemes that do not exploit the underlying structure of the received signal.

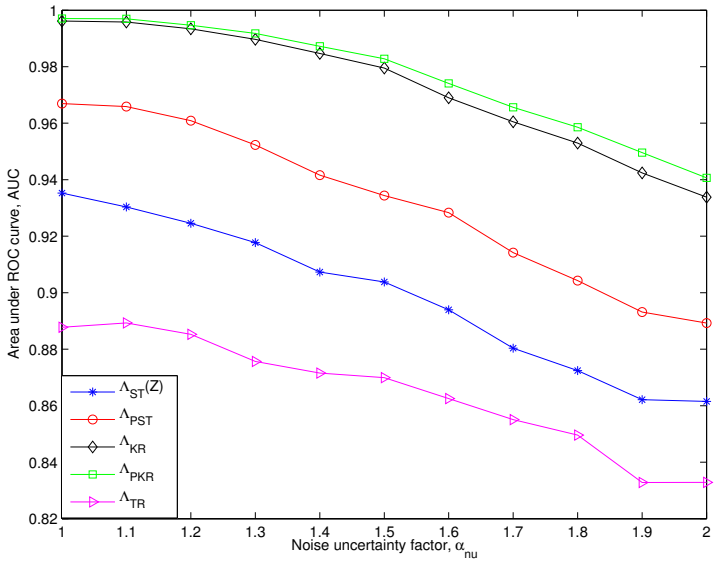


Figure 6.11: AUC curves: To asses the effects of noise uncertainty α_{nu} with sample size $M = 80$, number of vector samples per sub-block $N = 15$, number of antennas $L = 4$, effect of shadowing $\sigma_{\text{dB}-\text{Spread}} = 4$ and average SNR$\bar{\kappa} = -8$dB.

6.1.8 Application to single-user (virtual) multi-antenna detection

In Section 6.1.5 we have presented Kronecker based detection schemes to exploit spatio-temporal correlation for a single receiver (user) equipped with multi-antennas. In the present subsection we extend that scheme to the case of single-antenna sensors in a WSN, by considering the fact that a set of single-antenna sensors could be thought as the elements of a single-user multi-antenna device. Keeping this in mind, the problem can be formulated using the tools of multi-antenna signal processing explained in previous subsections.

6.1.8.1 Problem formulation

We consider an infrastructure based sensor network where a fusion center is supposed to receive observations from total K single-antenna sensor nodes and it makes the final decision about the presence of any event in the field. Furthermore, we assume that the signal emitted by the presence of the event can be modeled as an electromagnetic field that can be measured by a radio-frequency receiver. At the start of the detection process, the fusion center sends an inquiry message to all of the K sensors, asking them to send their observations. After receiving this signal from the fusion center, each sensor takes measurements (total M measurements each consist of N samples) and send them to the fusion center. At m-th measurement stage, samples collected from K sensors are stored in

the $K \times N$ matrix $\mathbf{X}(m) \triangleq \left[\begin{array}{cccc} \mathbf{x}(m) & \mathbf{x}(2m) & \cdots & \mathbf{x}(Nm) \end{array} \right]$ as:

$$\mathbf{X}(m) = \left[\begin{array}{cccc} x_1(m) & x_1(2m) & \cdots & x_1(Nm) \\ x_2(m) & x_2(2m) & \cdots & x_2(Nm) \\ \vdots & \vdots & \ddots & \vdots \\ x_K(m) & x_K(2m) & \cdots & x_K(Nm) \end{array} \right] = \left[\begin{array}{c} \mathbf{x}_1^T(m) \\ \mathbf{x}_2^T(m) \\ \vdots \\ \mathbf{x}_K^T(m) \end{array} \right] \tag{6.45}$$

where the k-th row, $\mathbf{x}_k(m) = [x_k(nm), x_k(nm), \cdots, x_k(nm)]^T$, contains N-samples received at the k-th sensor in the$m$-th measurement stage. By considering K sensors as if they were the multi-antennas of the previous sections, we can see that matrix $\mathbf{X}$ in this problem is similar to the one in Section 6.1.2 Hence, we can use the Kronecker based detection scheme in Section 6.1.5 by considering that $\mathbf{\Sigma}_{\mathrm{S}} = \mathbf{\Sigma}_{\mathrm{A}}$, where $\mathbf{\Sigma}_{\mathrm{S}}$ quantifies correlation between the rows of matrix $\mathbf{X}$ in (6.45). Apart from this, in this problem we also compare the result of Kronecker based detection scheme with some ad hoc tests. For instance, we recall CAV detector presented in Chapter 3, that is nothing but a ratio between the sum of elements of the spatio-temporal sample covariance matrix and the sum of diagonal elements of that matrix as:

$$\Lambda_2(\mathbf{Z}) = \frac{\sum_i^{NK} \sum_j^{NK} \left| \rho_1^{(i,j)} \right|}{\sum_i^{NK} \left| \rho_0^{(i,i)} \right|} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma \tag{6.46}$$

where $\rho_1^{(i,j)}$ is (i,j)-th element of the spatio-temporal sample covariance matrix $\hat{\mathbf{\Sigma}}_1$ and $\rho_0^{(i,i)}$ (i,i)-th diagonal element of $\hat{\mathbf{\Sigma}}_0 = \mathrm{diag}\hat{\mathbf{\Sigma}}_1$. Inspired by CAV detector (6.46), based on spatial and temporal covariance matrices achieved through algorithm 6.1, we propose

$$\Lambda_4(\mathbf{Z}) = \frac{\left(\sum_i^K \sum_j^K \left| \rho_{\mathrm{S}}^{(i,j)} \right| \right) \left(\sum_i^N \sum_j^N \left| \rho_{\mathrm{T}}^{(i,j)} \right| \right)}{\sum_i^{KN} \left| \rho_0^{(i,i)} \right|} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma \tag{6.47}$$

where $\rho_S^{(i,j)}$ and $\rho_{\mathrm{T}}^{(i,j)}$ are (i,j)-th elements of spatial sample covariance matrix $\hat{\mathbf{\Sigma}}_{\mathrm{S}}$ and temporal sample covariance matrix $\hat{\mathbf{\Sigma}}_{\mathrm{T}}$, respectively. Similarly, $\rho_0^{(i,i)}$ are (i,i)-th diagonal elements of $\hat{\mathbf{\Sigma}}_0$. Motivated by the John's U-statistic [77], here we also propose another ad hoc test statistic that uses the two separate spatial and temporal sample covariance matrices. The expression for this Ad hoc test is

$$\Lambda_5(\mathbf{Z}) = \frac{1}{KN} \mathrm{tr}\left(\hat{\Psi}_{\mathrm{S}} - \mathbf{I}_K \right)^2 \mathrm{tr}\left(\hat{\Psi}_{\mathrm{T}} - \mathbf{I}_N \right)^2 \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma \tag{6.48}$$

where $\hat{\Psi}_{\mathrm{S}} = \frac{\hat{\mathbf{\Sigma}}_S}{\frac{1}{K}\mathrm{tr}\left(\hat{\mathbf{\Sigma}}_S\right)}$ and $\hat{\Psi}_{\mathrm{T}} = \frac{\hat{\mathbf{\Sigma}}_T}{\frac{1}{N}\mathrm{tr}\left(\hat{\mathbf{\Sigma}}_T\right)}$. The detectors in (6.46), (6.47) and (6.48) do not involve determinants so they have robustness against the high dimensionality and they also do not assume any prior information about the signal and noise distribution.

6.1.8.2 Numerical example

For simulation we consider a WSN, where sensors are uniformly spread and we assume that the event appears at an unknown position. We analyze the performance of the detection approaches for different values of M, the number of samples by using AUC curves.

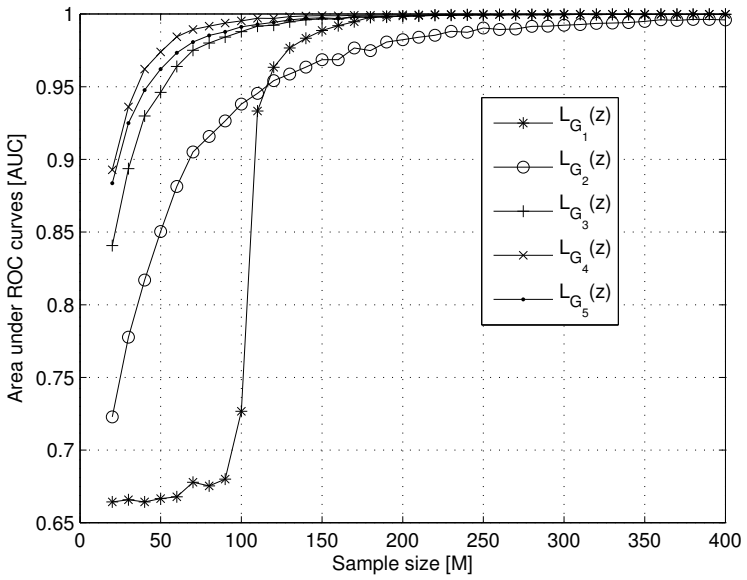


Figure 6.12: AUC curves: To asses the effects sample size M on the detection schemes for number of sensors K = 15, temporal dimension N = 5 and emitted power $P_0 = -20$dB.

In Fig.6.12 we analyze the the detection schemes for $K = 15$, $N = 5$. In the figure $L_{G_1}(\mathbf{Z})$ and $L_{G_3}(\mathbf{Z})$ represent detectors obtained by using $\Lambda_{\text{ST}}(\mathbf{Z})$ and $\Lambda_{\text{KR}}(\mathbf{Z})$, respectively. The results show that by increasing the value of M, of course the detection performances of all detectors increase, but the proposed schemes perform better than the traditional schemes. For example for a specific value of AUC, lets say AUC $= 0.9$, the proposed detectors $L_{G_3}(\mathbf{Z}) = \Lambda_3(\mathbf{Z})$, $L_{G_4}(\mathbf{Z}) = \Lambda_4(\mathbf{Z})$ and $L_{G_5}(\mathbf{Z}) = \Lambda_5(\mathbf{Z})$ need $M \leq 50$, while the traditional GLRT $L_{G_1}(\mathbf{Z}) = \Lambda_1(\mathbf{Z})$ and $L_{G_2}(\mathbf{Z}) = \Lambda_2(\mathbf{Z})$ need $M \geq 100$.

From these results we can conclude that the proposed detection schemes $\Lambda_3(\mathbf{Z})$, $\Lambda_4(\mathbf{Z})$ and $\Lambda_5(\mathbf{Z})$ are performing better than the traditional schemes $\Lambda_1(\mathbf{Z})$, $\Lambda_2(\mathbf{Z})$. The obvious reason for this enhancement at small sample support regime is the exploitation of underlying spatio-temporal structure, that results in reduction of the number of parameters (elements) to be estimated in the unknown covariance matrix. Amongst the proposed detectors, $\Lambda_4(\mathbf{Z})$ and $\Lambda_5(\mathbf{Z})$ show more robustness as the AUC curves reach to the reasonable high value at a very small M. It is because, they not only take advantage of the exploitation of covariance structure but also avoid matrix operations that are sensitive to the singularity of a matrix. Therefore, they show more robustness against the high dimensionality and small sample support compared to the other scheme presented in this subsections.

6.2 Multi-user multi-antenna spectrum sensing

In this section, we present a novel spectrum sensing for a cognitive radio network with multiple distributed radios where each radio has multiple antennas. We consider that the cognitive network has K secondary users, each user is equipped with L antennas. The proposed spectrum sensing scheme exploits the fact that when any primary signal is present, measurements are spatially correlated due to inter-antennas and inter-users spatial correlation.

6.2.1 Problem formulation

In the presence of the primary signal, the received signal at the output of L antennas of k-th user can be expressed as: $\mathbf{x}_k(n) = \mathbf{s}_k(n) + \mathbf{w}_k(n),\ n = 1, \cdots, N$, where $L \times 1$ vector $\mathbf{s}_k(n)$ contains the samples of the primary signal received and vector $\mathbf{w}_k(n)$ consists of additive noise at time n. The k-th user sends $\mathbf{x}_k(n)$ to the fusion center (i.e. base station, central controller) via ideal links[2]. The fusion center stacks the received signal vectors $\{\mathbf{x}_k(n)\}_{k=1}^{K}$ in the $KL \times 1$ vector $\mathbf{x}(n)$ as shown in Figure 6.13. Hence, the hypothesis testing can be represented as:

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{x}(n) = \mathbf{w}(n), \qquad n = 1, \cdots, N, \\ \mathcal{H}_1 &: \mathbf{x}(n) = \mathbf{s}(n) + \mathbf{w}(n),\ n = 1, \cdots, N, \end{aligned} \tag{6.49}$$

where

$$\mathbf{x}(n) = \left[\mathbf{x}_1^T(n), \mathbf{x}_2^T(n), \cdots, \mathbf{x}_K^T(n)\right]^T \in \mathbb{C}^{KL \times 1},$$

$$\mathbf{s}(n) = \left[\mathbf{s}_1^T(n), \mathbf{s}_2^T(n), \cdots, \mathbf{s}_K^T(n)\right]^T \in \mathbb{C}^{KL \times 1},$$

and

$$\mathbf{w}(n) = \left[\mathbf{w}_1^T(n), \mathbf{w}_2^T(n), \cdots, \mathbf{w}_K^T(n)\right]^T \in \mathbb{C}^{KL \times 1}.$$

We have the covariance matrix $\mathbf{\Sigma}_1 = \mathrm{E}\left[\mathbf{x}\mathbf{x}^H\right]$ as:

$$\mathbf{\Sigma}_1 = \mathrm{E}\left[\mathbf{x}\mathbf{x}^H\right] = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} & \cdots & \mathbf{\Sigma}_{1K} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} & \cdots & \mathbf{\Sigma}_{1K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{K1} & \mathbf{\Sigma}_{K2} & \cdots & \mathbf{\Sigma}_{KK} \end{bmatrix} \in \mathbb{C}^{KL \times KL}, \tag{6.50}$$

where the covariance matrices $\mathbf{\Sigma}_{k,k} = E\left[\mathbf{x}_k \mathbf{x}_k^H\right] \in \mathbb{C}^{L \times L}$, $1 \leq k \leq K$ capture the spatial correlation present in the signal received at L antennas of the k-th user. Similarly, the non-zero off-diagonal blocks $\mathbf{\Sigma}_{k,j} = E\left[\mathbf{x}_k \mathbf{x}_j^H\right] \in \mathbb{C}^{L \times L}$, $1 \leq k, j \leq K$ capture the spatial correlation between proximal users. Consequently, in order to simplify the detection

[2]In the network of cognitive-enabled femto-cells, already existing wired network can be considered as ideal links between the fusion center and the users [3]

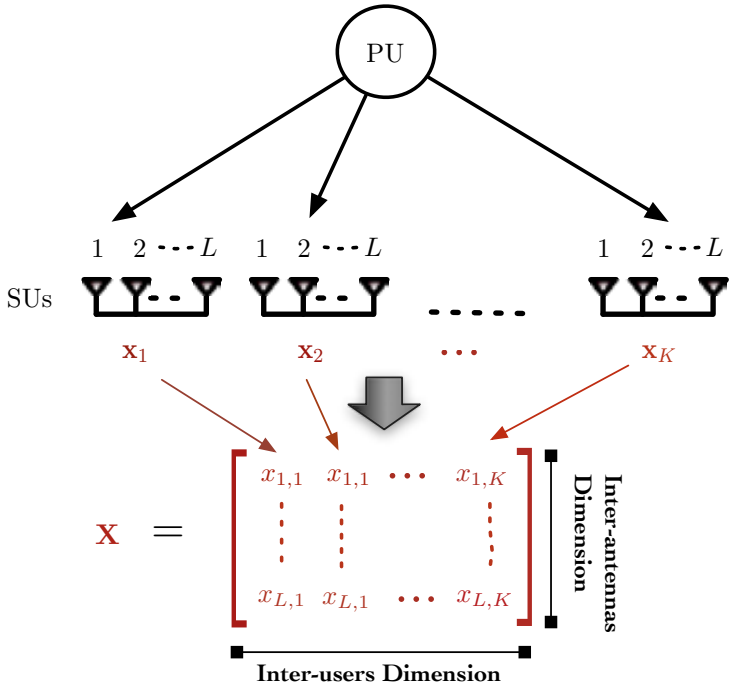


Figure 6.13: Schematic representation of the proposed multi-sensor, multi-antenna methodology.

problem, we need the distribution of $\mathbf{x}(n)$. We take it to be zero-mean Gaussian as in addition to resulting in tractable models and useful detectors, this assumption is reasonable for instance if the primary user employs orthogonal frequency division multiplexing (OFDM) as a modulation format [3] [34]. Hence, the hypothesis testing problem becomes:

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{x} \sim \mathcal{CN}(0, \mathbf{\Sigma}_0) \\ \mathcal{H}_1 &: \mathbf{x} \sim \mathcal{CN}(0, \mathbf{\Sigma}_1) \end{aligned} \tag{6.51}$$

where $\mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_h)$ $h = \{0, 1\}$, denotes the complex Gaussian distribution with zero mean and covariance $\mathbf{\Sigma}_h$. Under $\mathcal{H}_0$, $\mathbf{\Sigma}_0$ is an unknown diagonal matrix, since in the absence of the primary signal the observations are assumed to be white. For this problem, we will next present the traditional detector, and then derive the proposed improved detector by introducing the pair Kronecker product.

6.2.2 Conventional GLRT

In this section, we adopt the GLRT for the detection problem introduced in (6.51) since it is asymptotically an optimal detector. Later on it will be as a bench mark to be compared with the other proposed techniques. For the detection problem (6.51), the GLRT statistic can be

[3] We begin with the complex base-band signal sampled at the specific Nyquist rate.

formulated as:

$$\Lambda_{\mathrm{T}}\left(\mathbf{X}\right)=\frac{\underset{\boldsymbol{\Sigma}_0}{\max} f_{\mathbf{x}}\left(\mathbf{X};\boldsymbol{\Sigma}_0\right)}{\underset{\boldsymbol{\Sigma}_1}{\max} f_{\mathbf{x}}\left(\mathbf{X};\boldsymbol{\Sigma}_1\right)}\underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}}\gamma, \tag{6.52}$$

where $f_{\mathbf{x}}\left(\mathbf{X},\boldsymbol{\Sigma}_0\right)$ and $f_{\mathbf{x}}\left(\mathbf{X};\boldsymbol{\Sigma}_1\right)$ are likelihood functions under hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Similarly, $\mathbf{X}=\begin{bmatrix}\mathbf{x}\left(1\right) & \mathbf{x}\left(2\right), & \cdots, & \mathbf{x}\left(N\right)\end{bmatrix}$ is the matrix that contains all of the available samples of $KL\times 1$ vector $\mathbf{x}$. Solving (6.52) we can get the final expression of the GLRT as:

$$\Lambda_{\mathrm{T}}\left(\mathbf{X}\right)=\frac{\left|\hat{\boldsymbol{\Sigma}}_1\right|}{\left|\hat{\boldsymbol{\Sigma}}_0\right|}\underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}}\gamma, \tag{6.53}$$

where $\hat{\boldsymbol{\Sigma}}_1=\frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^H(n)$ is the maximum likelihood estimator (MLE) of $\boldsymbol{\Sigma}_1$. Similarly, under the alternate hypothesis, we assume that only noise is present and $\boldsymbol{\Sigma}_0$ is diagonal matrix, then $\hat{\boldsymbol{\Sigma}}_0=\mathrm{diag}\left(\frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^H(n)\right)$. In practice, the GLRT is used based on the assumption that the sample size N is large while the sample dimensions $\{K,L\}$ are small. However, when the sample support is limited (in particular, when $N\leq KL$) , the GLRT degenerates due to the ill-conditioning of the estimated covariance matrix [79]. A way to reduce these limitations will be presented next.

6.2.3 GLRT based on Kronecker product

The GLRT based detection scheme in (6.52) assumes no structure for the covariance matrix, except that the covariance matrix is symmetric. However, in order to circumvent the issue of ill-conditioning problems, one may assume an a-priori suitable structure on the covariance matrix [79]. Hence, we propose two detection techniques in Sections 6.2.3.1 and 6.2.3.2 that exploit the embedded correlation structure based on the single-pair Kronecker product (SPKP) and multi-pairs Kronecker product (MPKP), respectively.

6.2.3.1 SPKP-GLRT

In Section 6.1.2, we have discussed that the $(KL\times 1)$ vector $\mathbf{x}$(n) contains K sub-vectors of dimension $(L\times 1)$ corresponding to the arrays at the K different SUs. Consequently, we see that the covariance matrix $\boldsymbol{\Sigma}_1$ has mainly two types of spatial correlation structures. Inter-receiver correlation structure appears between the SUs due to their proximity, and inter-antenna correlation appears between samples from the antenna elements of the same SU. Moreover, by assuming the fact that the spatial correlation structure of arrays at different SUs is similar, the observation matrix $\mathbf{X}$ can be considered as a matrix normal with separable structure [82, 84].

Keeping the above facts and discussion in [79, 84] into considerations, the overall (inter-antenna plus inter-receiver) spatial covariance $\boldsymbol{\Sigma}_1$ can be represented with the help of the SPKP model as:

$$\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_{\mathrm{S}}\otimes\boldsymbol{\Sigma}_{\mathrm{A}} \tag{6.54}$$

where $\mathbf{\Sigma}_{\mathrm{S}} \in \mathbb{C}^{K\times K}$ and $\mathbf{\Sigma}_{\mathrm{A}} \in \mathbb{C}^{L\times L}$ capture the inter-users and inter-antennas spatial correlation, respectively. By using the results of (6.54) the SPKP-GLRT can be expressed as:

$$\Lambda_{\mathrm{SPKP}}\left(\mathbf{X}_N\right) = \frac{\max\limits_{\mathbf{\Sigma}_0} f_{\mathbf{x}}\left(\mathbf{X}_N; \mathbf{\Sigma}_0\right)}{\max\limits_{\mathbf{\Sigma}_{\mathrm{S}},\mathbf{\Sigma}_{\mathrm{A}}} f_{\mathbf{x}}\left(\mathbf{X}_N; \mathbf{\Sigma}_{\mathrm{A}} \otimes \mathbf{\Sigma}_{\mathrm{S}}\right)} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \gamma. \tag{6.55}$$

Solving (6.55), under the hypothesis $\mathcal{H}_1$, we need to estimate covariance matrices $\mathbf{\Sigma}_{\mathrm{S}}$ and $\mathbf{\Sigma}_{\mathrm{A}}$ by using the MLE paradigm. The MLEs under the hypothesis $\mathcal{H}_1$, can be written as [102]:

$$\hat{\mathbf{\Sigma}}_{\mathrm{A}} = \frac{1}{KN}\sum_{n=1}^{N} \mathbf{X}\left(n\right) \hat{\mathbf{\Sigma}}_{\mathrm{S}}^{-1} \mathbf{X}^{H}\left(n\right), \tag{6.56}$$

$$\hat{\mathbf{\Sigma}}_{\mathrm{S}} = \frac{1}{LN}\sum_{n=1}^{N} \mathbf{X}^{H}\left(n\right) \hat{\mathbf{\Sigma}}_{\mathrm{A}}^{-1} \mathbf{X}\left(n\right). \tag{6.57}$$

The non-iterative Flip-Flop approach to find $\hat{\mathbf{\Sigma}}_{\mathrm{A}}$ and $\hat{\mathbf{\Sigma}}_{\mathrm{S}}$ based on expressions (6.56) and (6.57) is given in Algorithm 6.3 with an initial value of $\hat{\mathbf{\Sigma}}_{\mathrm{S}}^{0} = \mathbf{I}_{K\times K}$. Finally, solving (6.55), we can get the expression:

$$\Lambda_{\mathrm{SPKP}}(\mathbf{X}) = \frac{\left|\hat{\mathbf{\Sigma}}_{\mathrm{A}}\right|^{K} \left|\hat{\mathbf{\Sigma}}_{\mathrm{S}}\right|^{L}}{\left|\hat{\mathbf{\Sigma}}_{0}\right|^{KL}} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \gamma. \tag{6.58}$$

The main advantage of the SPKP-GLRT in (6.58) over the traditional GLRT in (6.10) is that under $\mathcal{H}_1$, instead of $\frac{1}{2}KL(KL+1)$ parameters, it has only $\frac{1}{2}K\left(K+1\right) + \frac{1}{2}L\left(L+1\right)$ parameters to estimate. Therefore, the dimensions of $\mathbf{\Sigma}_{\mathrm{A}}$ and $\mathbf{\Sigma}_{\mathrm{S}}$ are much smaller than the dimension of the full covariance matrix $\mathbf{\Sigma}_1$, thus allowing a relaxation on the sample size that is required to avoid ill-conditioning of the MLE estimate $\hat{\mathbf{\Sigma}}_1$. Hence, the SPKP model is a good approximation that captures important information about the correlations and $\hat{\mathbf{\Sigma}}_1$ found under (6.22) is generally positive definite for $N > \max(K, L)$ [89].

Algorithm 6.3 ML based Non-Iterative Flip-Flop

1. Choose a starting value for $\hat{\Sigma}_{\mathrm{S}}^{0}$ as $\mathbf{I}_{K\times K}$
2. Estimate $\hat{\mathbf{\Sigma}}_{\mathrm{A}}^{1}$ from (6.56) with $\hat{\mathbf{\Sigma}}_{\mathrm{S}}^{0}$.
3. Find the following
   - Estimate $\hat{\mathbf{\Sigma}}_{\mathrm{S}}$ from (6.57)with $\hat{\mathbf{\Sigma}}_{\mathrm{A}}^{1}$.
   - Estimate $\hat{\mathbf{\Sigma}}_{\mathrm{A}}$ from (6.56) with $\hat{\mathbf{\Sigma}}_{\mathrm{S}}$.

6.2.3.2 MPKP-GLRT

In Section 6.2.3.1, we have approximated $\mathbf{\Sigma}_1$ in the form of two smaller covariance matrices through the SPKP model (6.22). The aforementioned SPKP model can also be written as: $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_{\mathrm{A}} \otimes \sum_{k=1}^{K} \lambda_k \mathbf{S}_k$, where $\mathcal{S}_k \triangleq \mathbf{v}_k \mathbf{v}_k^H$ is an orthonormal basis component represented as a singular matrix [103]. Furthermore, by using the identity $\mathbf{B} \otimes (\mathbf{O} + \mathbf{D}) = \mathbf{B} \otimes \mathbf{O} + \mathbf{B} \otimes \mathbf{D}$, we can write $\mathbf{\Sigma}_1 = \sum_{k=1}^{K} \mathbf{\Sigma}_{\mathrm{A}} \otimes \lambda_k \mathcal{S}_k = \sum_{k=1}^{K} \lambda_k \mathbf{\Sigma}_{\mathrm{A}} \otimes \mathcal{S}_k$. The expression makes it obvious that each spatial component $\mathcal{S}_k$ has the same inter-antenna covariance matrix. Hence, the SPKP covariance model is based on a very rigid assumption since it assumes similar inter-antenna spatial covariance structures at different SUs. However, in a more realistic case, different SUs are likely to have different inter-antenna spatial covariance. Therefore, to relax this rigid assumption and better explain the separation of inter-receiver and inter-antenna correlation, we consider a more general *multi-pair* Kronecker product (MPKP) model as:

$$\mathbf{\Sigma}_1 = \sum_{k=1}^{K} \mathbf{A}_k \otimes \mathcal{S}_k. \tag{6.59}$$

Note that by setting $\mathbf{A}_k = \lambda_k \mathbf{\Sigma}_{\mathrm{A}}$ we can see how (6.59) subsumes the SPKP model. Moreover, it is to be noted that the inter-receiver correlation components $\{\mathcal{S}_k\}_{k=1}^{K}$ are $(K \times K)$ rank-1 matrices and their corresponding inter-antenna correlation matrices $\mathbf{A}_k$ are full rank $(L \times L)$ matrices. The estimate of $\mathcal{S}_k$ or the spatial orthogonal components the $\{\mathbf{v}_k\}_{k=1}^{K}$ can be found by singular value decomposition (SVD) of the $(LN \times K)$ matrix $\mathbf{X}_V = \left[\mathbf{X}^T(1), \mathbf{X}^T(2), \cdots, \mathbf{X}^T(N)\right]^T$ as: $\mathbf{X}_V = \mathbf{U}\mathbf{\Psi}\mathbf{V}^H$[103]. Where $\mathbf{U}$ is a $(LN \times K)$ orthogonal matrix, $\mathbf{\Psi}$ is a $(K \times K)$ diagonal matrix with singular values $\{\eta_k\}_{k=1}^{K}$ of $\mathbf{X}_V$ as diagonal elements and $\mathbf{V}$ is a $(K \times K)$ orthogonal matrix of the spatial components. Each row of $\mathbf{V}^H$ is $\mathbf{v}_k$ that forms $\mathcal{S}_k$ as: $\mathcal{S}_k = \mathbf{v}_k \mathbf{v}_k^H$ for the model (6.59). Similarly, the estimate of $\mathbf{A}_k$ can be found based on the MLE paradigm as [103]:

$$\hat{\mathbf{A}}_k = \frac{1}{N} \sum_{n=1}^{N} \mathbf{X}(n) \hat{\mathcal{S}}_k \mathbf{X}^H(n), \tag{6.60}$$

and $\hat{\mathcal{S}}_k = \mathbf{v}_k \mathbf{v}_k^H$. Hence, we can write $\hat{\mathbf{\Sigma}}_1 = \sum_{k=1}^{K} \hat{\mathbf{A}}_k \otimes \hat{\mathcal{S}}_k$ and the MPKP-GLRT becomes:

$$\Lambda_{\mathrm{MPKP}}(\mathbf{X}_N) = \frac{\left| \sum_{k=1}^{K} \hat{\mathbf{A}}_k \otimes \hat{\mathcal{S}}_k \right|}{\left| \hat{\mathbf{\Sigma}}_0 \right|} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \gamma. \tag{6.61}$$

Compared to the SPKP-GLRT in (6.58), the MPKP-GLRT in (6.61) can better account for the different inter-antenna structures, since it is not constrained to the assumption of identical inter-antenna correlation at different SUs. However, the computational cost of the detection scheme (6.61) is slightly higher than the (6.58). This is because the number of free parameters in the covariance matrix to be estimated under hypothesis $\mathcal{H}_1$ has increased to $L^2 + LK(K-1)/2$ [103].

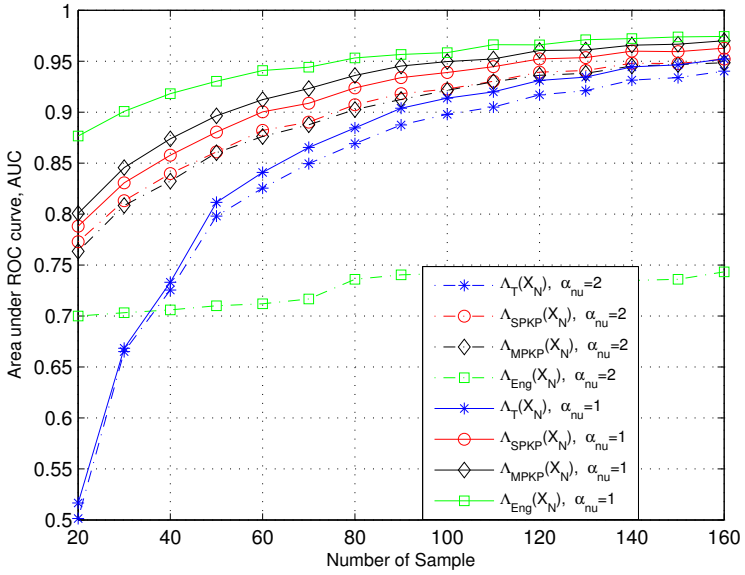


Figure 6.14: AUC curves (Solid lines for $\alpha_{nu} = 1$ and dashed lines for $\alpha_{nu} = 2$): To assess the effects of the sample size (N), with shadowing $\sigma_{\text{dB}-\text{Spread}} = 6$ and average SNR$\bar{\kappa} = -15$dB.

6.2.4 Numerical results

For the analysis to be conducted herein, we consider a wireless network with a total of K = 10 SUs (equipped with $L = 4$ antennas) randomly deployed to detect a PU that appears at an unknown position. The spatial correlation between the antennas of a SU is modeled herein as $c_{i,j} = \rho^{|i-j|}$, with $\{i, j\} = 1, \cdots, L$. Moreover, $0 < \rho < 1$ with $\rho = e^{-23\kappa^2(d\lambda_c)^2}$, which is called correlation coefficient between two adjacent antennas, and it relies on the angular spread κ, the wavelength λ_c and the distance d between two adjacent antennas [104]. Finally, the spatial correlation between SUs due to the correlated shadowing effects is modeled based on the model given in [50].

In order to analyze the performance of the proposed detectors, we use the area under the ROC curve (AUC), which varies between 0.5 (poor performance) and 1 (good performance). Moreover, we define the average SNR of all SUs as: $\bar{\kappa} = \frac{1}{K}\sum_{i=1}^{K}\kappa_i$ and the SNR of i-th SU is $\kappa_i = \zeta^{i-1}\kappa_{\text{min}}$ where κ_{min} is the minimum SNR among those of the SUs. For a given average SNR $\bar{\kappa}$ and SNR gap ζ, we can generate the random SNRs of the SUs.

In Fig. 6.14, we show the AUC plots to analyze the effects of the sample size N in the presence of noise power uncertainty and shadowing. We model the noise power uncertainty by generating the noise power at k-th SU as $\sigma^2_{w,k} \sim \mathcal{U}\left(\frac{\sigma^2_{n,k}}{\alpha_{nu}}, \alpha_{nu}\sigma^2_{n,k}\right)$, where $\alpha_{nu} \geq 1$, and $\alpha_{nu} = 1$ means no noise uncertainty. The dashed lines represent AUC curves for $\alpha_{nu} = 2$ and solid lines represent the case when $\alpha_{nu} = 1$. With these considerations, the results clearly show that Λ_{MPKP} and Λ_{SPKP} outperform the traditional Λ_{T} with a smaller sample support. Moreover, we also compare the results of these three detectors with energy detector

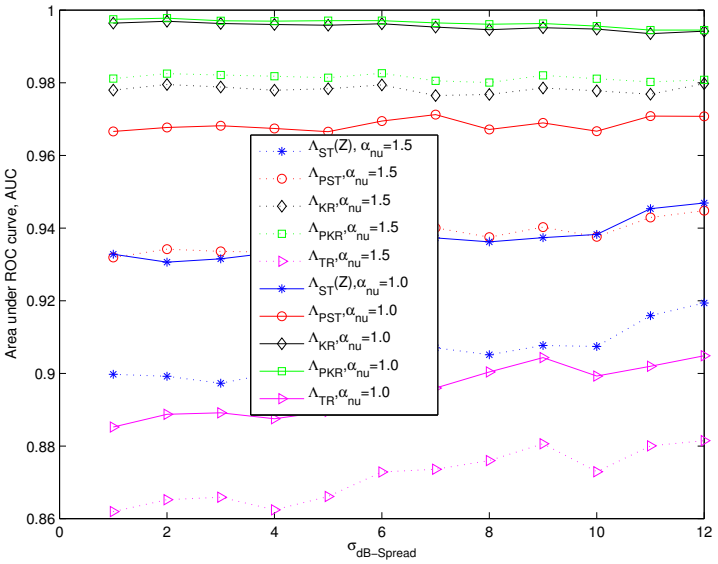


Figure 6.15: AUC curves (Solid lines for $\alpha_{nu} = 1$ and dashed lines for $\alpha_{nu} = 2$): To assess the effect of shadowing, $\sigma_{\text{dB}-\text{Spread}}$, with sample size $N = 70$ and average SNR $\bar{\kappa} = -15$dB.

given as:

$$\Lambda_{\text{Eng}}\left(\mathbf{X}_N\right) = \sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{l=1}^{L}\left|x_{k,l}\left(n\right)\right|^2. \tag{6.62}$$

We can see that the energy detector performs better in the absence of the noise power uncertainty, however, it has poor performance in the presence of noise power uncertainty. Interestingly, we can also observe that in the presence of noise power uncertainty, the increase of sample size has very small impact on the performance of the energy based detection scheme Λ_{ENG}.

In Fig.6.15, we show the AUC plots to analyze the effect of shadowing (i.e. $\sigma_{\text{dB}-\text{Spread}}$). From the result we can see that the effect of the shadowing is very small over the detection performance of the detection schemes as the presented spectrum sensing schemes are cooperative. However, in the case of Λ_{ENG}, the detection performance slightly increases with the increase in the $\sigma_{\text{dB}-\text{Spread}}$. The most obvious reason for this can be the heavy-tailed distribution of the primary signal strength due to the log-normally-distributed shadow fading that behave in such a way at lower SNR [73].

In Fig. 6.16, we show the AUC plots to analyze the effects of noise power uncertainty. Once again, we can see that Λ_{MPKP} and Λ_{SPKP} have superior performance than Λ_{T}. Moreover, we can observe the proposed schemes are robust against the noise power uncertainty compare to Λ_{ENG}.

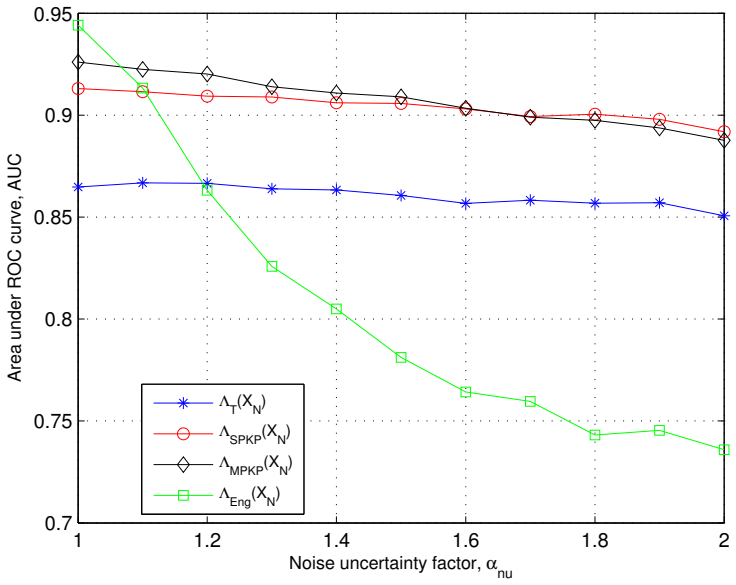


Figure 6.16: AUC curves: To asses the effects of noise uncertainty α_{nu}, with shadowing $\sigma_{\text{dB}-\text{Spread}} = 6$, sample size $N = 70$ and average SNR $\bar{\kappa} = -15$dB.

6.2.5 Application to multi-user (virtual) multi-antenna detection

The focus of this subsection is to use the tools already shown for the multi-user multi-antenna case, to a different scenario of multiple single-antenna devices in a WSN as shown in Fig. 6.17. In order to do so, the clusters in Fig. 6.17 can be understood as a multi-antenna device.

6.2.5.1 Proposed Methodology

In the proposed method we consider to slice the whole field of K sensors equally, into L_c clusters and the number of clusters L_c should be sub-multiple of K. Consequently, the total received vector $\mathbf{x}(n) \in \mathbb{R}^K$ is sliced into L_c sub-vectors as: $\mathbf{x}_l(n) \in \mathbb{R}^{K_c}$: $l = 1, 2, \cdots, L_c$ in a way that the l-th sub-vectors correspond to l-th cluster as elaborated in Fig. 6.17.

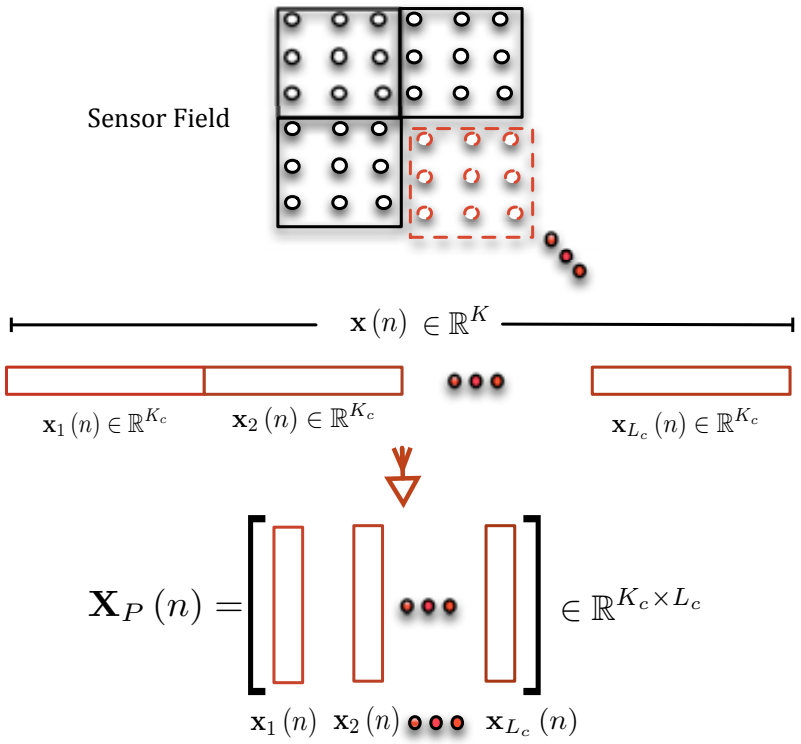


Figure 6.17: Schematic representation of the slicing process of a sensor field and the observation vector.

Once we have these sub-vectors of the received observation $\mathbf{x}(n)$ then all of the sub-vectors are stacked into a $K_c \times L_c$ matrix $\mathbf{X}_p(n) = \{\mathbf{x}_1(n), \mathbf{x}_2(n), \cdots, \mathbf{x}_{L_c}(n)\} \in \mathbb{R}^{K_c \times L_c}$. The elements present in each column of the matrix $\mathbf{X}_p(n)$ are spatially correlated and this correlation is quantified by inter-cluster covariance matrix $\mathbf{\Sigma}_{K_c}$. Similarly, the cross-correlation between the columns of the matrix $\mathbf{X}_p(n)$ is quantified by intra-cluster covariance matrix $\mathbf{\Sigma}_{L_c}$. Now considering the fact that the sensors are placed in a uniform grid, we can infer that the covariance structure of the columns of matrix $\mathbf{X}_p(n)$ remains the same as the topology of the K_c sensors in all of the L_c clusters is the same as if the clusters are uniform antenna arrays as in Section 6.2.2. This leads us to say that $\mathbf{X}_p(n)$ is indeed comply with the matrix $\mathbf{X}$ in the case of multi-user multi-antenna scheme in Section 6.2.2. Hence, this motivates to use $\Lambda_{\text{SPKP}}(\mathbf{X})$ in (6.55) for the detection problem in the present section.

6.2.5.2 Numerical results

For the purpose of simulation we consider a sensor network, where $K = 36$ sensors are placed in the uniform grid. For the simulations we consider $L_c = 4$ clusters, that are basically located in the form of four quadrants of a cartesian co-ordinate system and each cluster consists of $K_c = 9$ sensors. Furthermore, we assume that the event appears at the center $(x = 0, y = 0)$.

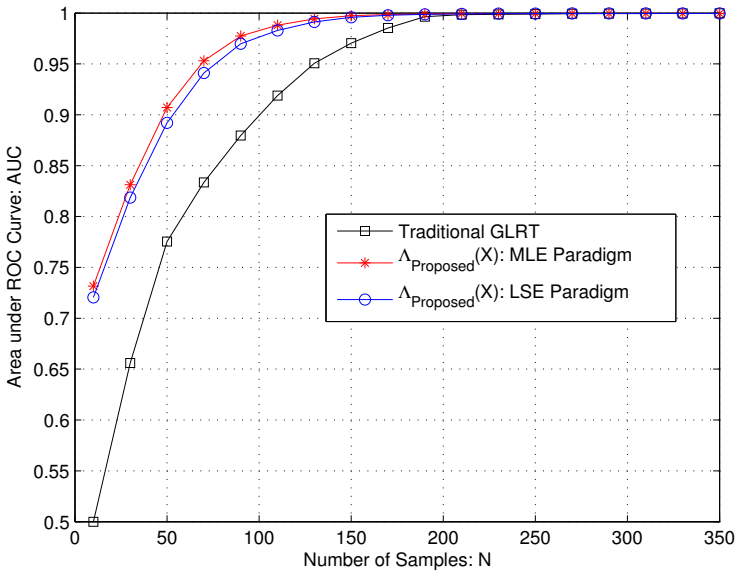


Figure 6.18: AUC curves: To assess the effects of the sample size for total number of sensors, $K = 36$.

In Fig 6.18 we plot AUC curves to analyze the detectors, for different values of the number of samples N used to estimate the covariance matrices. It is to noted that $\Lambda_{\text{Proposed}}(\mathbf{X})$ represents detection scheme (6.55) for the problem under observation. As it was expected, the results confirm that the performance of the proposed scheme is better even in the case of very small N compared to the traditional GLRT. From the experiment it can be concluded that when $N < K$, the traditional GLRT completely collapse and the proposed schemes have reasonable good performance even under $N < K$ regime. Another interesting conclusion can be drawn from the overall results is that the performance of proposed scheme with MLE paradigm is slightly better than the LSE paradigm.

CHAPTER 7

Conclusions

The aim of this contribution has been to achieve robust centralized multi-sensor detection schemes based on the exploitation of available prior information. In this context, we have exploited the fact that in dense WSN, the signal is typically a local phenomenon that only affects a small subset of sensors. These affected sensors will normally be located close to the event as well as close to each other in the form of a spatial cluster. Based on this intuition, novel detection schemes have been presented with a two-fold motivation: first, the exploitation of the relevant set of sensors, which helps in rejecting the noise; second, to take advantage of the signal correlation occurring by using a-priori information about the positions of sensors through a signature matrix, which captures the correlation among different sensors. From the in-depth analysis of such scheme, we have found that it improves the detection performance with respect to conventional multi-sensor detection schemes. We have also applied the proposed scheme to collaborative spectrum sensing and the aim has been to achieve an improvement in the sensing performance. The performance of the proposed spectrum sensing scheme has been tested under harsh conditions, such as the presence of severe noise, shadowing effects and noise power uncertainties due to the presence of interferences. The simulation results have confirmed the convenience of this novel approach, showing a superior performance, robustness against noise power uncertainty and shadowing effects, compared to traditional detectors that ignore sensor selection and spatial information.

Similarly, we have also focused on exploiting the expected structure and patterns in the received covariance matrix due to the underlying topology. In this regard, we have proposed novel detection schemes that exploit the spatio-temporal correlation present in the received observations at the multi-antenna receiver. We have also extended the concept of exploiting such correlation to the case of event detection in WSN. Moreover, we have proposed several collaborative spectrum sensing schemes that exploit the spatial correlation present in the received observations at sensors (secondary users), each of them equipped with multiple antennas. For all these proposed schemes, first we have adopted the traditional GLRT based on the observed covariance matrix that is considered to be asymptotically an optimal detector. However, we have observed that the GLRT performs poorly when the sample support is small. To cope with this, we have proposed several extensions that bring robustness against the sample support. The proposed schemes exploit the inherent structure of the received covariance matrix such as persymmetry and the Kronecker product of matrices. The perfor-

mance of the proposed schemes has been evaluated in presence of uncertain additive noise and shadowing with the help of numerical simulations. The results have shown that the proposed detection schemes consistently have better detection performance in the case when the sample size is much smaller than the dimensionality of data.

Future work

As most of the work in this thesis assumes that the links between the sensors and the fusion center are ideal, hence, a possible extension could be to analyze these schemes by incorporating the actual channel effects that exist between the sensors and the fusion center.

In the case of detection schemes based on the spatial signatures, we have assumed that the sensor to sensor distances are perfectly estimated. However, in practice the estimation of these distances could be erroneous, therefore, it will be interesting to analyze the performance of these schemes by including the effects of these errors. Another possible extension could be to extend the concept of spatial signatures to the distributed detection system, where instead of fusion center, each local sensor has the knowledge of signatures of the neigbouring sensors. Similarly, the focus of these schemes has been to exploit the spatial proximity while dealing with soft information, therefore, we would also be interested to exploit this known spatial proximity while considering exchange of one bit hard information between the sensors and the fusion center.

In the case of the multi-antenna, multi-sensor detection schemes that exploit the inherent structures and patterns in the observed covariance matrices, a possible next step could be to compare the detection performance of these schemes for practically obtained signals i.e., OFDM signals. Moreover, the results obtained for these schemes can also be applied to the spoofing detection in a GNSS receiver equipped with multiple antennas.

References

[1] J.-F. Chamberland and V. Veeravalli, "Wireless sensors in distributed detection applications," *IEEE Signal Processing Mag.*, vol. 24, pp. 16–25, 2007.

[2] D. Culler, D. Estrin, and M. Srivastava, "Overview of sensor networks," *IEEE Computer Society*, vol. 37, p. 8, 2004.

[3] G. Gür, S. Bayhan, and F. Alagoz, "Cognitive femtocell networks: an overlay architecture for localized dynamic spectrum access [dynamic spectrum management]," *IEEE Wireless Communications*, vol. 17, no. 4, pp. 62–70, 2010.

[4] R. R. Tenney and N. R. Sandell., "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-17, pp. 501–510, 1980.

[5] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: Part I–Fundamentals," *Proceedings of IEEE*, vol. 85, pp. 54–63, 1997.

[6] T. Berg and H. Durrant-Whyte, "Distributed and decentralized estimation," in *Proc. of the Singapore International Conference on Intelligent Control and Instrumentation (SICICI)*, 1992.

[7] S. Barbarossa and G. Scutari, "Bio-inspired sensor network design: Distributed decisions through self-synchronization," *IEEE Signal Processing Mag.*, vol. 24, pp. 26–35, 2007.

[8] P. Varshney, *Distributed Detection and Data Fusion.* Springer, 1996.

[9] J. Ma, G. Zhao, and Y. Li, "Soft combination and detection for cooperative spectrum sensing in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, pp. 4502–4507, 2008.

[10] Y. Zeng and Y.-C. Liang, "Eigenvalue-based spectrum sensing algorithms for cognitive radio," *IEEE Transactions on Communications*, vol. 57, no. 6, pp. 1784 – 1793, June 2009.

[11] ——, "Spectrum-sensing algorithms for cognitive radio based on statistical covariances," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 1804 – 1815, May 2009.

[12] E. Axell and E. Larsson, "A unified framework for GLRT-based spectrum sensing of signals with covariance matrices with known eigenvalue multiplicities," in *Proc. ICASSP*, May 2011, pp. 2956 –2959.

[13] E. Axell, G. Leus, E. Larsson, and H. Poor, "Spectrum sensing for cognitive radio : State-of-the-art and recent advances," *IEEE Signal Processing Magazine*, vol. 29, no. 3, pp. 101 –116, May 2012.

[14] E. Visotsky, S. Kuffner, and R. Peterson, "On collaborative detection of TV transmissions in support of dynamic spectrum sharing," in *Proc. 1st IEEE DySPAN*, 2005, pp. 338–345.

[15] R. Tandra and A. Sahai, "SNR walls for signal detection," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 24–17, Feb 2008.

[16] S. J. Shellhammer, S. Shankar, R. Tandra, and J.Tomcik, "Performance of power detector sensors of DTV signals in IEEE 802.22 WRANs," in *Proc. 1st International Workshop on Technology and policy for accessing spectrum (TAPAS)*, 2006, pp. 4–13.

[17] I. F. Akyildiz, M. C. Vuran, and O. B. Akan, "On exploiting spatial and temporal correlation in wireless sensor networks," in *Proc. of WiOpt*, 2004, pp. 71–80.

[18] A. Jindal and K. Psounis, "Modeling spatially-correlated sensor network data," in *Proc. First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks*, 2004, pp. 162–171.

[19] W. Yang, G. Durisiand, V. I. Morgenshtern, and E. Riegler, "Capacity pre-log of SIMO correlated block-fading channels," in *Proc. 8th International Symposium Wireless Communication Systems (ISWCS)*, nov. 2011, pp. 869 –873.

[20] Y. Zeng, Y.-C. Liang, A. T. Hoang, and R. Zhang, "A review on spectrum sensing for cognitive radio: challenges and solutions," *EURASIP J. Adv. Signal Process*, pp. 2:2–2:2, January 2010.

[21] B. Kasiri and C. Jun, "Effects of correlated shadowing on soft decision fusion in cooperative spectrum sensing," in *Proc. of IEEE Conference on Computer Communications*, 2010, pp. 1–6.

[22] A. Ghasemi and E. S. Souca, "Asymptotic performance of collaborative spectrum sensing under correlated log-normal shadowing," *IEEE Communications Letters*, vol. 11, no. 1, pp. 34–36, Jan 2007.

[23] H. L. V. Trees, *Detection, Estimation, and Modulation Theory, Part I.* John Wiley & Sons, 2001.

[24] M. A. Richards, *Fundamentals of Radar Signal Processing*, 1st ed. McGraw-Hill, 2005.

[25] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*. Prentice Hall, 1998.

[26] J. Theiler, "Confusion and clairvoyance: some remarks on the composite hypothesis testing problem," *SPIE Proceedings*, vol. 8390, p. 10, 2012.

[27] T. Fawcett, "ROC graphs: Notes and practical considerations for researchers," HP Laboratories, Tech. Rep. HPL-2003-4, 2003.

[28] V. Madisetti, *Digital Signal Processing Fundamentals*, 2nd ed. Boca Raton, FL, USA: CRC Press, Inc., Nov. 2009.

[29] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Processing Mag.*, vol. 21, pp. 36–47, 2004.

[30] C. Xu and S. M. Kay, "On centralized composite detection with distributed sensors," in *Proc. of the IEEE Radar Conference*, 2008, pp. 1–6.

[31] F. Vincent, O. Besson, and C. Richard, "Matched subspace detection with hypothesis dependent noise power," *IEEE Transactions on Signal Processing*, vol. 56, no. 11, pp. 5713–5718, 2008.

[32] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 1: Estimation Theory*. Prentice Hall, 1993.

[33] R. Niu and P. K. Varshney, "Distributed detection and fusion in a large wireless sensor network of random size," *EURASIP Journal on Wireless Communications and Networking*, vol. 4, pp. 462–472, 2005.

[34] D. Ramirez, J. Via, I. Santamaria, and L. L. Scharf, "Detection of spatially correlated gaussian time series," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5006 –5015, Oct. 2010.

[35] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley-Interscience, Ed. Wiley-Interscience, 2003.

[36] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. Academic Press, 1979.

[37] P. Belanovic, S. Valcarcel-Macua, and S. Zazo, "Location-aided Distributed Primary User Identification in a Cognitive Radio Scenario," in *In the proc of IEEE ICASSP*, 2012.

[38] F. Ding, G. Song, K. Yin, J. Li, and A. Song, "A GPS-enabled wireless sensor network for monitoring radioactive materials," *Sensors and Actuators A: Physical*, vol. 155, no. 1, pp. 210–215, 2009.

[39] N. Vankayalapati and S. M. Kay, "Asymptotically optimal detection/ localization of LPI signals of emitters using distributed sensors," *Proc. of the SPIE*, vol. 7706, pp. 77 060U–77 060U–9, 2010.

[40] R. Shibata, "Selection of the order of an autoregressive model by Akaike's information criterion," *Biometrika*, vol. 63, pp. 117–126, 1976.

[41] S. Zarrin and T. J. Lim, "Cooperative quickest spectrum sensing in cognitive radios with unknown parameters," in *IEEE GLOBECOM*, 2009, pp. 1–6.

[42] V. V. Veeravalli, T. Basar, and H. V. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 433 –442, March 1993.

[43] O. Hadjiliadis, H. Zhang, and H. V. Poor, "One shot schemes for decentralized quickest change detection," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3346–3359, July 2009.

[44] H. V. Poor and O. Hadjiliadis, *Quickest Detection*. Cambridge University Press, 2008.

[45] L. Lai, Y. Fan, and H. V. Poor, "Quickest detection in cognitive radio: A sequential change detection framework," in *Proc. IEEE GLOBECOM*, Dec 2008, pp. 1–5.

[46] H. Li, C. Li, and H. Dai, "Quickest spectrum sensing in cognitive radio," in *Proc. 42nd Annual Conference on Information Sciences and Systems*, 2008.

[47] S. Ali, G. Seco-Granados, and J. A. Lopez-Salcedo, "Spectrum sensing with spatial signatures in the presence of noise uncertainty and shadowing," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, June 2013.

[48] R. Niu and P. Varshney, "Decision fusion in a wireless sensor network with a random number of sensors," in *Proc. IEEE ICASSP*, vol. 4, March 2005, pp. 861–864.

[49] G. Lorden, "Procedures for reacting to a change in distribution," *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1897–1908, December 1971.

[50] J. Unnikrishnan and V. Veeravalli, "Cooperative sensing for primary detection in cognitive radio," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 18 –27, Feb. 2008.

[51] W. Headley, V. Chavali, and C. da Silva, "Exploiting radio correlation and reliability information in collaborative spectrum sensing," *IEEE Communications Letters*, vol. 15, no. 8, pp. 825 –827, august 2011.

[52] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *IEEE Electronics Letters*, vol. 27, pp. 2145 – 2146, 1991.

[53] S. Kullback, *Information Theory and Statistics*, 2nd ed. Dover Publications, 1968.

[54] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Select. Areas Commun.*, vol. 23, no. 2, pp. 201–220, February 2005.

[55] K. Shin, H. Kim, A. Min, and A. Kumar, "Cognitive radios for dynamic spectrum access: from concept to reality," *IEEE Wireless Communications*, vol. 17, no. 6, pp. 64–74, December 2010.

[56] Federal Communications Commission, "Facilitating opportunities for flexible, efficient and reliable spectrum use employing cognitive radio technologies," *FCC 03-322*, 2003.

[57] Y. Zeng, Y. C. Liang, Z. Lei, S. W. Oh, F. Chin, and S. Sun, "Worldwide regulatory and standardization activities on cognitive radio," in *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum*, April 2010, pp. 1–9.

[58] A. F. Molisch, L. J. Greenstein, and M. Shafi, "Propagation issues for cognitive radio," *Proc. IEEE*, vol. 97, no. 5, pp. 787–804, May 2009.

[59] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," vol. 11, no. 1, pp. 116–130, March 2009.

[60] Z. Quan, S.Cui, H. V. Poor, and A. H. Sayed, "Collaborative wideband sensing for cognitive radios," *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 60 –73, Nov. 2008.

[61] J. Schafer and K. Strimmer, "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," *Statist. Appl. Genet. Mol. Biol.*, vol. 4, p. 32, 2005.

[62] Y.-R. Tsai, "Sensing coverage for randomly distributed wireless sensor networks in shadowed environments," *IEEE Trans. Veh. Technol.*, vol. 57, no. 1, pp. 556–564, January 2008.

[63] A. W. Min, K. G. Shin, and X. Hu, "Secure cooperative sensing in IEEE 802.22 WRANs using shadow fading correlation," *IEEE Transactions on Mobile Computing*, vol. 10, no. 10, pp. 1434–1447, October 2011.

[64] P. K. Varshney and C. S. Burrus, *Distributed detection and data fusion.* New-York: Springer-Verlag, 1997.

[65] S. Shellhammer and G. Chouinard, "Spectrum sensing requirements summary," *IEEE P802.22-06/0089r1, Tech. Rep.*, June 2006.

[66] M. Di Renzo, L. Imbriglio, F. Graziosi, and F. Santucci, "Cooperative spectrum sensing over correlated log-normal sensing and reporting channels," in *Proc. IEEE GLOBECOM*, 2009, pp. 1–8.

[67] G. Vazquez-Vilar, D. Ramírez, R. López-Valcarce, J. Vía, and I. Santamaría, "Spatial rank estimation in cognitive radio networks with uncalibrated multiple antennas," in *Proc. 4th International Conference on Cognitive Radio and Advanced Spectrum Management.* New York, USA: ACM, 2011, pp. 35:1–35:5.

[68] D. Ramirez, G. Vazquez-Vilar, R. Lopez-Valcarce, J. Via, and I. Santamaria, "Detection of rank-P Signals in Cognitive Radio Networks With Uncalibrated Multiple Antennas," *IEEE Transactions on Signal Processing*, vol. 59, pp. 3764–3774, August 2011.

[69] H. L. V. Trees, *Optimum Array Processing (Detection, Estimation, and Modulation Theory, Part IV).* Wiley-Interscience, 2002.

[70] C. Stein, "Inadmissibility of the usual estimator of the mean of a multivariate normal distribution." *Proc. of the 3rd Berkeley Symp. on Math. Statist. and Prob*, vol. 1, pp. 197–206, 1956.

[71] Y. Chen, A. Wiesel, and A. O. Hero, "Shrinkage estimation of high dimensional covariance matrices," in *Proc. of IEEE ICASSP*, 2009, pp. 2937–2940.

[72] W. Lin and Q. Zhang, "A design of energy detector in cognitive radio under noise uncertainty," in *Proc. IEEE Singapore International Conference on Communication Systems*, Nov. 2008, pp. 213 –217.

[73] T. Muetze, P. Stuedi, F. Kuhn, and G. Alonso, "Understanding radio irregularity in wireless networks," in *Proc. of IEEE SECON*, June 2008, pp. 82 –90.

[74] Y. Zeng, Y.-C. Liang, A. T. Hoang, and R. Zhang, "A review on spectrum sensing for cognitive radio: challenges and solutions," *EURASIP J. Adv. Signal Process*, vol. 2010, p. 15, January 2010.

[75] L. Shen, H. Wang, W. Zhang, and Z. Zhao, "Multiple Antennas Assisted Blind Spectrum Sensing in Cognitive Radio Channels," *IEEE Communications Letters*, vol. 16, no. 1, pp. 92–94, Jan. 2012.

[76] A. W. Min, X. Zhang, and K. G. Shin, "Detection of small-scale primary users in cognitive radio networks," *IEEE J. Select. Areas Commun.*, vol. 29, pp. 349 – 361, 2011.

[77] O. Ledoit and M. Wolf, "Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size," *The Annals of Statistics*, vol. 30, no. 4, pp. 1081–1102, 2002.

[78] T. Ayoub and A. Haimovich, "Modified GLRT signal detection algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 3, pp. 810–818, Jul 2000.

[79] D. Akdemir, "Slicing: Nonsingular estimation of high dimensional covariance matrices using multiway Kronecker delta covariance structures," *Submitted arXiv:1104.1767v2 [math.ST]*, 04 2011. [Online]. Available: http://arxiv.org/abs/1104.1767

[80] R. Nitzberg, "Application of maximum likelihood estimation of persymmetric covariance matrices to adaptive processing," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 16, no. 1, pp. 124–127, Jan. 1980.

[81] P. Wirfalt and M. Jansson, "On Toeplitz and Kronecker structured covariance matrix estimation," in *Proc. Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Oct. 2010, pp. 185–188.

[82] K. Werner, M. Jansson, and P. Stoica, "On estimation of covariance matrices with Kronecker product structure," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 478 –491, Feb. 2008.

[83] N. Lu and D. L. Zimmerman, "The likelihood ratio test for a separable covariance matrix," *Statistics & Probability Letters*, vol. 73, no. 5, pp. 449–457, July 2005.

[84] M. G. Genton, "Separable approximations of space-time covariance matrices," *Environmetrics*, vol. 18, no. 7, pp. 681–695, 2007.

[85] S.Ali, J. A. López-Salcedo, and G. Seco-Granados, "Exploiting structure of spatio-temporal correlation for detection in wireless sensor networks," in *Proc. 20th European Signal Processing Conference*, Aug. 2012, pp. 774–778.

[86] Y. Zeng, Y.-C. Liang, and E. C. Y. Peh, "Optimal cooperative sensing for sensors equipped with multiple antennas," in *Proc. ICC*, 2012, pp. 1571–1575.

[87] D. Romero and R. Lopez-Valcarce, "Distributed spectrum sensing with multiantenna sensors under calibration errors," in *Proc. SPAWC*, 2011, pp. 441–445.

[88] Y. Liang and V. Veeravalli, "Capacity of noncoherent time-selective rayleigh-fading channels," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3095–3110, 2004.

[89] P. Dutilleul, "The MLE algorithm for the matrix normal distribution." *Journal of Statistical Computation and Simulation*, vol. 64, pp. 105–123, 1999.

[90] K. V. Mardia, *Statistics of Directional Data*. Academic Press, 1972.

[91] L. Lu, H.-C. Wu, and S. Iyengar, "A novel robust detection algorithm for spectrum sensing," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 2, pp. 305 –315, February 2011.

[92] M. Wax and T. Kailath, "Efficient inversion of Toeplitz-block Toeplitz matrix," *IEEE Transactions on Acoustics, Speech and Signal Processing*,, vol. 31, no. 5, pp. 1218–1221, Oct 1983.

[93] L. Cai and H. Wang, "A persymmetric multiband GLR algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 3, pp. 806–816, 1992.

[94] S. Kim, J. Lee, H. Wang, and D. Hong, "Sensing Performance of Energy Detector With Correlated Multiple Antennas," *IEEE Signal Processing Letters*, vol. 16, no. 8, pp. 671–674, Aug. 2009.

[95] P. Wang, Z. Sahinoglu, M.-O. Pun, and H. Li, "Persymmetric parametric adaptive matched filter for multichannel adaptive signal detection," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 3322–3328, June 2012.

[96] T. K. Huckle, K. Waldherr, and T. Schulte-Herbrüggen, "Exploiting matrix symmetries and physical symmetries in matrix product states and Tensor Trains," *Linear and Multilinear Algebra*, vol. 61, no. 1, pp. 91–122, 2013.

[97] M. Jansson and P. Stoica, "Forward-only and forward-backward sample covariances - a comparative study," *Signal Processing*, vol. 77, no. 3, pp. 235–245, Sept 1999.

[98] B. Armour, "Structured covariance autoregressive parameter estimation," Ph.D. dissertation, Department of Electrical Engineering, McGill Canada, 1989.

[99] J. Kamm and J. Nagy, "Optimal Kronecker Product Approximation of Block Toeplitz Matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 22, no. 1, pp. 155–172, 2000.

[100] M. W. Mitchell, M. G. Genton, and M. L. Gumpertz, "A likelihood ratio test for separability of covariances," *J. Multivar. Anal.*, vol. 97, no. 5, pp. 1025–1043, May 2006.

[101] C. V. Loan and N. Pitsianis, "Approximation with Kronecker products," *Linear Algebra for Large Scale and Real Time Applications*, pp. 293–314, 1993.

[102] S. Ali, M. Jansson, G. Seco-Granados, and J. A. López-Salcedo, "Novel collaborative spectrum sensing based on spatial covariance structure," in *Proc. 21th European Signal Processing Conference*, 2013.

[103] S. Plis, D. Schmidt, S. Jun, and D. Ranken, "A generalized spatiotemporal covariance model for stationary background in analysis of MEG data," in *Proc. IEEE EMBS, New York*, 2006, pp. 3680–3683.

[104] S. Kim, J. Lee, H. Wang, and D. Hong, "Sensing performance of energy detector with correlated multiple antennas," *IEEE Signal Processing Letters*, vol. 16, no. 8, pp. 671–674, Aug. 2009.

[105] G. H. Golub and C. F. Van Loan, *Matrix computations (3rd ed.)*. Baltimore, MD, USA: Johns Hopkins University Press, 1996.

[106] F. Bijma, J. C. de Munck, and R. M. Heethaar, "The spatiotemporal MEG covariance matrix modeled as a sum of Kronecker products," *Neuro-Image*, vol. 27, pp. 402–415, 2005.

[107] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley-Interscience, 1999.

APPENDIX A

Matrix Differentiation

Let us define the derivative of $f(\mathbf{X})$ with respect to $N \times P$ matrix $\mathbf{X}$ as:

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \left(\frac{\partial f(\mathbf{X})}{\partial x_{i,j}}\right) \tag{A.1}$$

then we have the following results

1. $\frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$
2. $\frac{\partial \mathbf{x}^T \mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{x}$
3. $\frac{\partial \mathbf{x}^T \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$
4. $\frac{\partial \mathbf{x}^T \mathbf{A}\mathbf{y}}{\partial \mathbf{x}} = \mathbf{A}\mathbf{y}$
5. $\frac{\partial |\mathbf{X}|}{\partial x_{i,j}} = X_{i,j}$ if all elements of $N \times P$ matrix $\mathbf{X}$ are distinct, where $X_{i,j}$ is the (i,j)-th cofactor of $\mathbf{X}$.
6. $\frac{\partial |\mathbf{X}|}{\partial x_{i,j}} = \begin{cases} X_{i,i} & i = j \\ 2X_{i,j} & i \neq j \end{cases}$ if $\mathbf{X}$ is symmetric.
7. $\frac{\partial \mathrm{Tr}(\mathbf{X}\mathbf{Y})}{\partial \mathbf{X}} = \begin{cases} \mathbf{Y}^T & \text{elements of } N \times P\text{matrix } \mathbf{X} \text{ are distinct.} \\ \mathbf{Y} + \mathbf{Y}^T - \mathrm{diag}(\mathbf{Y}) & \text{if } \mathbf{X} \text{ is symmetric.} \end{cases}$
8. $\frac{\partial \mathbf{X}^{-1}}{\partial x_{i,j}} = \begin{cases} -\mathbf{X}^{-1}\mathbf{J}_{i,i}\mathbf{X}^{-1} & i = j \\ -\mathbf{X}^{-1}\left(\mathbf{J}_{i,j} + \mathbf{J}_{j,i}\right)\mathbf{X}^{-1} & i \neq j \end{cases}$ if $\mathbf{X}$ is symmetric.

   where $\mathbf{J}_{i,j}$ is matrix with 1 at i,j-th location and zeros elsewhere.

APPENDIX B

Kronecker Product and Persymmetric Matrices

List of some important properties of the Kronecker product and the persymmetric matrices.

Persymmetric matrix

An arbitrary matrix $\mathbf{A}$ is called symmetric, if $\mathbf{A} = \mathbf{A}^T$ (i.e. $a_{i,j} = a_{j,i}$) and skew-symmetric, if $\mathbf{A}^T = -\mathbf{A}$. Similarly, persymmetric matrix is a square matrix which is symmetric in the northeast-to-southwest diagonal. In other words, if a symmetric matrix is rotated by 90°, it becomes a persymmetric matrix. Moreover, symmetric persymmetric matrices are sometimes called bisymmetric matrices [105].

Definition B.1. Let $\mathbf{A} = (a_{i,j})$ be an $N_A \times N_A$ matrix. Then matrix $\mathbf{A}$ to be a persymmetric matrix requires that

$$a_{ij} = a_{n-j+1,n-i+1} \text{for all } i, j. \tag{B.1}$$

Let $\mathbf{J} \in \mathbb{R}^{N_A \times N_A}$ be the counter-identity matrix which is also called exchange matrix. Then persymmetry can also be expressed by

$$\mathbf{JAJ} = \mathbf{A}^T. \tag{B.2}$$

Based on this persymmetry property, a matrix is a symmetric persymmetric, if it is symmetric about both diagonals, i.e.,

$$\mathbf{JAJ} = \mathbf{A}^T = \mathbf{A} \tag{B.3}$$

or component-wise

$$a_{i,j} = a_{j,i} = a_{n+1-i,n+1-j}. \tag{B.4}$$

Note that a matrix with the property $\mathbf{JAJ} = \mathbf{A}$ is called centrosymmetric. Therefore, symmetric persymmetric or symmetric centrosymmetric are the same. The set of all symmetric persymmetric $N_A \times N_A$ matrices is closed under addition and under scalar multiplication. Similarly, a matrix $\mathbf{A}$ is called symmetric skew-persymmetric if $\mathbf{JAJ} = -\mathbf{A}^T = -\mathbf{A}$. Furthermore, a persymmetric determinant is the determinant of a persymmetric matrix. A matrix for which the values on each line parallel to the main diagonal are constant, is called a

Toeplitz matrix. It should be noted that the Toeplitz structured matrices belong to a subclass of the persymmetric matrices [81, 80, 92].

Kronecker Product

In mathematics, the Kronecker product, denoted by $\otimes$ is an operation on two matrices of arbitrary size resulting in a block matrix. It is a generalization of the outer product (which is denoted by the same symbol) from vectors to matrices, and gives the matrix of the tensor product with respect to a standard choice of basis. The Kronecker product should not be confused with the usual matrix multiplication, which is an entirely different operation [101].

Definition B.2. If $\mathbf{A} = (a_{i,j})$ is an $N_A \times N_A$ matrix and $\mathbf{B} = (a_{i,j})$ is a $M_B \times M_B$ matrix, then the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is the $M_B N_A \times N_A M_B$ block matrix:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{1,1}\mathbf{B} & a_{1,2}\mathbf{B} & \cdots & a_{1,N_A}\mathbf{B} \\ a_{2,1} & a_{2,2} & & \vdots \\ \vdots & & \ddots & \\ a_{N_A,1}\mathbf{B} & \cdots & & a_{N_A,N_A}\mathbf{B} \end{bmatrix} \tag{B.5}$$

Obviously, the same definition holds if $\mathbf{A}$ and $\mathbf{B}$ are either complex-valued matrices or real-valued matrices. The Kronecker product has a rich and very pleasing algebra that supports a wide range of fast, elegant, and practical algorithms. Several trends in scientific computing suggest that this important matrix operation will have an increasingly greater role to play in the future[81]. First, the application areas where Kronecker products abound are all thriving. These include signal processing, image processing, semidefinite programming, and quantum computing[101]. Here, in this work we take advantage in reducing the computational cost of GLRT that exploit correlation present in the received observations.

Important properties

Here we list three important properties about the Kronecker Product of two matrices.

- If $\mathbf{A}$ and $\mathbf{B}$ are invertible matrices, then

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} \tag{B.6}$$

- If $\mathbf{A}$ and $\mathbf{B}$ are square matrices, then

$$(\mathbf{A} \otimes \mathbf{B})^{T} = \mathbf{A}^{T} \otimes \mathbf{B}^{T} \tag{B.7}$$

- If $\mathbf{A}$ is an $N_A \times N_A$ matrix and $\mathbf{B}$ is a $M_B \times M_B$ matrix

$$|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^{M_B} |\mathbf{B}|^{N_A} \tag{B.8}$$

- If $\mathbf{A}$ is an $N_A \times N_A$ matrix and $\mathbf{B}$ is a $M_B \times M_B$ matrix

$$(a\mathbf{A}) \otimes \mathbf{B} = a\,(\mathbf{A} \otimes \mathbf{B}) = \mathbf{A} \otimes (a\mathbf{B}) \tag{B.9}$$

Lemma B.3. *The Kronecker product of two symmetric persymmetric matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *is again symmetric persymmetric.*

Proof. Let $\mathbf{J}_A$ and $\mathbf{J}_B$ denote the exchange matrices which correspond to the size of $\mathbf{A}$ and $\mathbf{B}$, respectively. Then the exchange matrix $\mathbf{J}_{A\otimes B}$ of $\mathbf{A} \otimes \mathbf{B}$ is given by $\mathbf{J}_{A\otimes B} = \mathbf{J}_A \otimes \mathbf{J}_B$. Therefore,

$$(\mathbf{J}_A \otimes \mathbf{J}_B)\,(\mathbf{A} \otimes \mathbf{B})\,(\mathbf{J}_A \otimes \mathbf{J}_B) = (\mathbf{J}_A\mathbf{A}\mathbf{J}_A) \otimes (\mathbf{J}_B\mathbf{B}\mathbf{J}_B) = \mathbf{A}^T \otimes \mathbf{B}^T = \mathbf{A} \otimes \mathbf{B} \tag{B.10}$$

□

APPENDIX C

Derivation of the Least Square Estimate of the Kronecker Product of Covariance Matrices

Let $\mathbf{X} \in \mathbb{R}^{K\times N}$ is matrix normal. We assume that we have M realizations of $\mathbf{X}$. Let us define the $\left\{m^{th}\right\}_{m=1}^{M}$ realization as:

$$\mathbf{X}\left(m\right) \triangleq \begin{bmatrix} x_1\left(m\right) & x_1\left(2m\right) & \cdots & x_1\left(Nm\right) \\ x_2\left(m\right) & x_2\left(2m\right) & \cdots & x_2\left(Nm\right) \\ \vdots & \vdots & \ddots & \vdots \\ x_K\left(m\right) & x_K\left(2m\right) & \cdots & x_K\left(Nm\right) \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_K^T \end{bmatrix} \tag{C.1}$$

Similarly, define a vector $\mathbf{z}\left(m\right) \triangleq \text{vec}\left\{\mathbf{X}\left(m\right)\right\}$. The covariance matrix of the $KN\times 1$ vector $\mathbf{z}$ under hypothesis $\mathcal{H}_1$ is

$$\mathbf{\Sigma} = \text{E}\left[\mathbf{z}\mathbf{z}^T\right] = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} & \cdots & \mathbf{\Sigma}_{1K} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} & \cdots & \mathbf{\Sigma}_{1K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{K1} & \mathbf{\Sigma}_{K2} & \cdots & \mathbf{\Sigma}_{KK} \end{bmatrix} \in \mathbb{R}^{KN\times KN} \tag{C.2}$$

The estimate of $\mathbf{\Sigma}$ can be found as:

$$\hat{\mathbf{\Sigma}} = \frac{1}{M}\sum_{m=1}^{M} \mathbf{z}(m)\mathbf{z}^T(m). \tag{C.3}$$

In case the $\mathbf{X} \in \mathbb{R}^{K\times N}$ is matrix normal with separable structure then we can write [106]

$$\mathbf{\Sigma} = \mathbf{\Sigma}_{\text{A}} \otimes \mathbf{\Sigma}_{\text{B}}, \tag{C.4}$$

where the sub-matrix $\mathbf{\Sigma}_{\text{A}}$ captures the correlation between the elements of each column of $\mathbf{X}$ and sub-matrix $\mathbf{\Sigma}_{\text{B}}$ captures the correlation between N columns vector in block $\mathbf{X}$. The expression for the Least Square estimators of $\hat{\mathbf{\Sigma}}_{\text{A}}$ and $\hat{\mathbf{\Sigma}}_{\text{B}}$ are derived based on minimization of the the LS-cost function,

$$\left\|\hat{\boldsymbol{\Sigma}} - \hat{\boldsymbol{\Sigma}}_{\mathrm{A}} \otimes \hat{\boldsymbol{\Sigma}}_{\mathrm{B}}\right\|^2 = \mathrm{Tr}\left(\boldsymbol{\Sigma}_{\mathrm{A}}\boldsymbol{\Sigma}_{\mathrm{B}}^T\right) + \mathrm{Tr}\left(\boldsymbol{\Sigma}_{\mathrm{B}}^2\right)\mathrm{Tr}\left(\boldsymbol{\Sigma}_{\mathrm{A}}^2\right) - \frac{2}{N-1}\sum_{m=1}^{M}\mathrm{Tr}\left\{\mathbf{X}(m)\boldsymbol{\Sigma}_{\mathrm{B}}\left(\mathbf{X}\left(m\right)\right)^T\boldsymbol{\Sigma}_{\mathrm{A}}\right\}, \tag{C.5}$$

and the solutions for $\hat{\boldsymbol{\Sigma}}_{\mathrm{B}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathrm{A}}$ are found by differentiating (C.5) using results in [107] and subsequently equating the first derivative to zero. For the case of $\hat{\boldsymbol{\Sigma}}_{\mathrm{A}}$, this yields

$$\mathrm{d}_{\boldsymbol{\Sigma}_{\mathrm{A}}}\left\{\left\|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{\mathrm{A}} \otimes \boldsymbol{\Sigma}_{\mathrm{B}}\right\|^2\right\} = 2\mathrm{Tr}\left(\boldsymbol{\Sigma}_{\mathrm{B}}^2\right)\mathrm{Tr}\left(\boldsymbol{\Sigma}_{\mathrm{A}}\mathrm{d}\boldsymbol{\Sigma}_{\mathrm{A}}\right) - \frac{2}{N-1}\sum_{m=1}^{M}\mathrm{Tr}\left\{\mathbf{X}(m)\boldsymbol{\Sigma}_{\mathrm{B}}\mathbf{X}^T\left(m\right)\mathrm{d}\boldsymbol{\Sigma}_{\mathrm{A}}\right\}. \tag{C.6}$$

Equating to zero and after some algebraic manipulation, we get

$$\left(\mathrm{Tr}\boldsymbol{\Sigma}_{\mathrm{B}}\right)^2\boldsymbol{\Sigma}_{\mathrm{A}} - \frac{1}{N-1}\sum_{m=1}^{M}\mathbf{X}(m)\boldsymbol{\Sigma}_{\mathrm{B}}\mathbf{X}^T\left(m\right) = 0. \tag{C.7}$$

Arranging the above equation yields

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{A}} = \frac{(M-1)^{-1}}{\mathrm{Tr}\left(\hat{\boldsymbol{\Sigma}}_{\mathrm{B}}^2\right)}\sum_{m=1}^{M}\mathbf{X}^T\left(m\right)\hat{\boldsymbol{\Sigma}}_{\mathrm{B}}\mathbf{X}\left(m\right), \tag{C.8}$$

and similar derivations can be performed to find:

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{B}} = \frac{(M-1)^{-1}}{\mathrm{Tr}\left(\hat{\boldsymbol{\Sigma}}_{\mathrm{A}}^2\right)}\sum_{m=1}^{M}\mathbf{X}\left(m\right)\hat{\boldsymbol{\Sigma}}_{\mathrm{A}}\mathbf{X}^T\left(m\right), \tag{C.9}$$

Using (C.8) and (C.9), we get $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}_{\mathrm{A}} \otimes \hat{\boldsymbol{\Sigma}}_{\mathrm{B}}$.