BAYESIAN APPROACH FOR THE SPECTRUM SENSING MIMO-COGNITIVE RADIO NETWORK WITH PRESENCE OF THE UNCERTAINTY

A Dissertation presented to

the Faculty of the Graduate School

at the University of Missouri Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Muthana Al-Amidie

Dr. Naz Islam, Dissertation Supervisor

Dec. 2021

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

BAYESIAN APPROACH FOR THE SPECTRUM SENSING
MIMO-COGNITIVE RADIO NETWORK
WITH PRESENCE OF THE UNCERTAINTY

presented by Muthana Al-Amidie,
a candidate for the degree of Doctor of Philosophy and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Islam Naz

Dr. Robert Oconell

Dr. John Gahl

Dr. Ronald McGarvey

Dedicated to

The awaited Saviour,

Mohammad Ibn Al-Hasan Al-Mahdi

ACKNOWLEDGMENTS

Firstly, I want to express my gratitude to my creator, Almighty Allah, for his blessing and benefactions. I would never be willing to accomplish my PhD without his consistent support and care. My Lord, I am unable to accomplish thanksgiving since my thanking is a lack of appreciation!

I'd also want to thank my sponsor, Iraq's ministry of higher education and scientific research, for their support. MOHSER has given me a six-year scholarship to pursue my PhD in the United States, and has provided me with consistent funding despite Iraq's economic, political, and military problems. Thank MOHSER, for allowing me to pursue my passion of studying overseas!

Prof. Naz, research adviser is deserving of my appreciation and admiration. Working with Dr. Islam has proven to be one of the best decisions I've ever made. Dr. Islam is a firm believer in me and sees my accomplishment as his own (or fail). He has motivated me to put in more effort in order to generate more innovative ideas and become a successful researcher. He has always taken the time to listen to my academic and personal issues. Being under his direction is something I will always be proud of. Thank you, Dr. Islam!

I'd like to thank Dr. Robert O'conell, Dr. John Gahl, and Dr. Ronald McGarvey for their contributions to my PhD committee. Thanks to the Electrical and Computer Engineering department, as well as the caring instructors and staff at MU, my PhD studies have been beneficial and pleasurable. Thank you for your support, Mizzou!

During my stay at MU, I met many colleagues and true friends who have assisted me and created a warm and friendly environment around me. I owe a huge debt of appreciation to my lovely friends in Kumal's, Imam Hasan's, and "Duhaa and Zeara" groups, who have shared in my joys, sorrows, and other life events. Participating in their weekly get-together has helped me relax and concentrate on my studies while also improving my

personal attributes. They deserve a big thank you!

Finally, words will never be able to adequately explain my feelings about my large family, which includes my wife and children. My PhD studies have been made much easier and effective because of their love, care, understanding, patience, and unwavering support. Dr. Sura, my life partner, deserves a special thank you for walking alongside me on this difficult path!

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Acronyms

ACK/NAK	ACKnowledgment and Negative AcKnowledgment signaling protocol
AI	Alternate-Iterate optimization
AI-FDPS	Alternate-Iterate Feasible Direction Projected Subgradient
CC	Compound Capacity
CDMA	Code-Division Multiple Access
CR	Cognitive Radio
CSCG	Circularly Symmetric Complex Gaussian distribution
CSI	Channel State Information
DSA	Dynamic Spectrum Access
EVD	Eigenvalue Decomposition
FBF	Fractional Bayes Factor
FDD	Frequency-Division Duplexing
GLRT	Generalized Likelihood Ratio Test
HSDPA	High Speed Downlink Packet Access communication system
IT	Interference Temperature
INR	Interference-to-Noise Ratio
LMI	Linear Matrix Inequality
LR	Leakage Rate
LTE-A	Long Term Evolution-Advanced communication system
MI	Mutual Information
MIMO	Multiple-Input Multiple-Output
NMSPG	Non-Monotone Spectral Projected Gradient
OFDM	Orthogonal Frequency-Division Multiplexing
PU	Primary User

QoS	Quality of Service
ROC	Receiver Operating Characteristic
SDP	Semi-Definite Programming
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio
SU	Secondary User
TDD	Time-Division Duplexing

ABSTRACT

A cognitive radio technique has the ability to learn. This system not only can observe the surrounding environment, adapt to environmental conditions, but also efficiently use the radio spectrum. This technique allows the secondary users (SUs) to employ the primary users (PUs) spectrum during the band is not being utilized by the user. Cognitive radio has three main steps: sensing of the spectrum, deciding and acting. In the spectrum sensing technique, the channel occupancy is determined with a spectrum sensing approach to detect unused spectrum. In the decision process, sensing results are evaluated and the decision process is then obtained based on these results. In the final process which is called the acting process, the scholar determines how to adjust the parameters of transmission to achieve great performance for the cognitive radio network.

Cognitive radio processes are affected due to different conditions that have an effect on the channel such as shadowing, multi-path fading, and uncertainty conditions. Thus, the measurements obtained by the SUs through the sensing step could be uncertain data. Decisions are determined according to observation at the secondary side utilizing knowledge, which may have been affected by the uncertainty. For this reason, decisions can be wrong. Therefore, the cognitive radio system can fall into the wrong actions. It's clear that uncertainty distribution directly influences cognitive radio's achievement, and overcoming it is considered a crucial process in this technique.

Wideband spectrum sensing is considered the principal challenge of the CR technique. Existing spectrum sensing techniques can be attempted based on observations data sampled by using an ADC at the Nyquist rate. Since there are hardware limitations are regrading the sampling technique, that approach can sense the spectrum only one spectrum channel on time. Moreover, in order to perform the sensing of a wideband spectrum, the spectrum needs to be split into multiple frequency narrow bands. On the SUs side, each point of the spectrum is sensing constantly from the RF frontends, and this can be per-

formed at a fast time with computational complexity, and hardware cost.

To overcom this issue, the signal sampling time needs to be fast, even with high dimensional signals but the sensing step is still influenced by the uncertainty situation as mentioned earlier. Spectrum sensing is proposed to solve the unuseful use of the spectrum in a low-cost solution with low processing time and increase the scanning process. This spectrum sensing technique can be utilized with a low number of samples that are required which can maintain the essential information. This dissertation aims to propose an efficient and fast spectrum sensing technology to improve spectrum detection overusing the sensing solutions. In this dissertation, a review of using the different spectrum sensing techniques that can be utilized to sense the signal with fewer data samples, then compare them with the proposed techniques in terms of the probability of detection and the probability of false alarm. In this work, we propose the use of a Bayesian approach to estimate the unknown parameters after numerically obtaining the threshold value to perform the detection technique. we also propose an approach that combines the framework advantages of estimation and detection part using the Bayesian models to perform sensing under uncertainty. In addition, a Bayesian sensing technique based on the maximum posterior probability is proposed to handle the channel uncertainty and lately enhance the spectrum sensing efficiency.

The estimation approach also use the convex optimization method using to estimate the parameters of the wideband spectrum under certain uncertainty. All proposed approaches in this dissertation were simulated. Their results were compared to the existing proper techniques based on a number of metrics, like probability of detection, probability of false alarm, signal to noise ratio (SNR), number of samples, sensing threshold, required number of measurements, the SNR, and complexity.

Chapter 1

Introduction

This chapter discusses the basic scope of the spectrum sensing concept for cognitive radio networks under various settings and issues, as well as the spectrum sensing challenges for various static spectrum allocations. We also discuss how we employed spectrum sensing in collaboration with an estimation and detection approach to overcome some of these challenges. The remainder of the chapter, however, can be structured as follows: section

1.1 and Section 1.2 shows the spectrum management issue as well as the cognitive radio network; section 1.3 presents the MIMO wireless network; 1.4 describes the cognitive radio steps; Section 1.5 describes the dissertation objectives; Section 1.6 and Section 1.7 describes the different contributions with the list of all the published papers in international conferences and indexed journals. Finally, the dissertation organization is described in section 1.8 .

1.1 Spectrum Management in Cognitive Radio

One of the essential techniques is called the radio spectrum, which limited resources established by regulations and approved by agencies, like the Federal Communications Com-

mission (FCC) and the national agency for the legalization of communications (ANRT) that is located in the US. Modern spectrum management is assigning the channel spectrum to particular users with licenses to have particular wireless services and technologies. Information and wireless system traffic have recently increased rapidly over the past decade, which ended in a critical need for resources of the signal spectrum [11][4]. To precisely share the spectrum sensing, the authorized users should have totally access to that spectrum to send the data information, while other users are considered unauthorized users and not allowed even though the spectrum hole is free [4].

In the United States, studies recently have shown that the spectrum utilization ranges between 15%- 85% based on the used spectrum (FSA) policy[4]. FCC computations also clarified the most part of the channels are heavily used, while the rest of the channel is sparsely utilized as shown in Figure (1.1) [5].

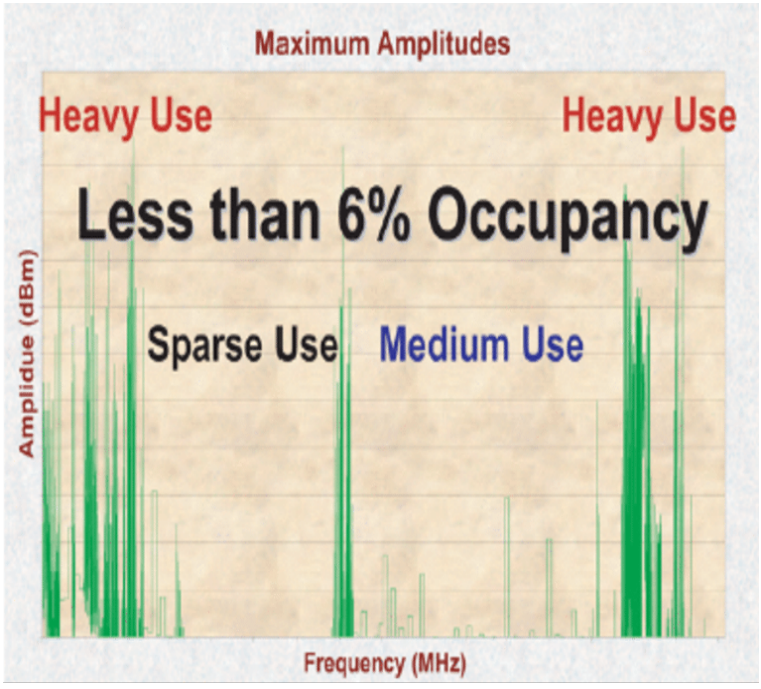

Figure 1.1: Range of spectrum occupancy [4] [5]

Those studies show that allocated spectrum parts are not utilized all the time by the users, this allocated user is called by the primary end-users (PUs) that can use the spectrum space. A white space, also named by spectrum hole, is a spectrum bandwidth that is

allocated to an end-user but not being utilized at a certain location and at a specific time. Thus, it can be seen that the radio frequency band in wireless communication is being utilized in an inefficient way [12] [13][[8,9]. Therefore, the inefficiency use and scarcity of the spectrum is required an urgent demand to achieve an enhancement for radio spectrum usage *and consequently enhance the network achievement.* A more reliable method to defeat the spectrum deficiency is dynamically solved by involving unused spectrum to be used by unlicensed users, secondary users (SUs), but these methods externally causes an interference issue with the other PUs signals. Two different access are achieved, the first method is known by opportunistic spectrum access (OSA), and the second is known by dynamic spectrum access (DSA), those techniques are proposed to study the spectrum allocation problems. Those techniques FSA and DSA allow that licensed and non-licensed users can share the spectrum band. In other words, the spectrum should be allocated into various bandwidths, then those spectrum bands are assigned to different dedicated users[14] [15].

1.2 Cognitive Radio

According to this research paper "Mitola", cognitive radio [16][17] is a smart radio frequency transceiver technique designed to recognize the available holes in a wireless spectrum and adjust parameters of the transmission, allowing more reliable communications and promoting radio processing performance [14,15]. It is totally a new technique to improve wireless networking. This technique allows the radio device to be awarded to its environment and also provides the ability of the device system to adjust and achieve its parameters autonomously. A cognitive radio network can be learned and can observe from its around the environment to be adapted to the environmental situation, then make decisions to establish an efficient use for the radio spectrum. the main point for this technique is to provide permission for the SUs to use the allocated radio spectrum of the PU

when it is temporally not being used as shown in Fig (1.2).

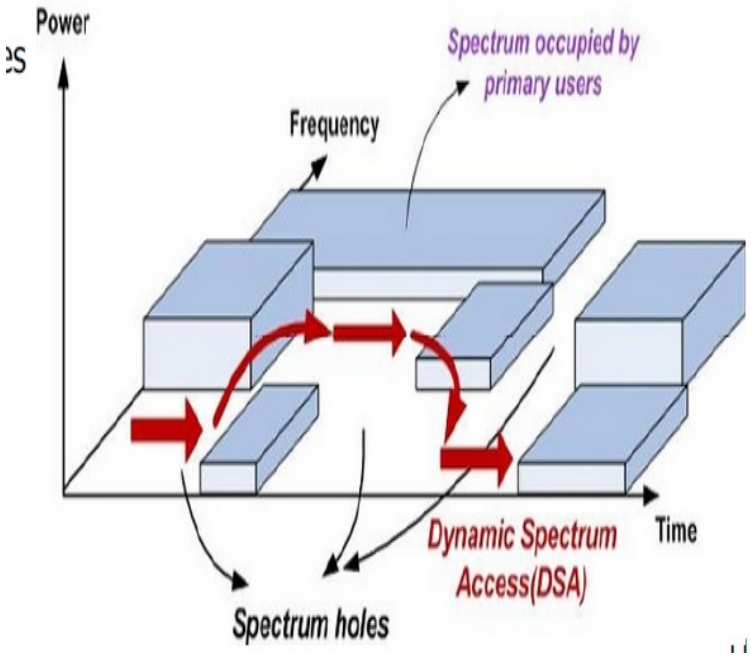

Figure 1.2: Dynamic band access [4][6]

Cognitive radio has been highlighted as a possible way to overcome spectrum resource constraints, which is a requirement of modern wireless communication technology, such as the 5*G* wireless communication structure. Wireless communication systems are among the 5*G* production. Wide wireless networks will be linked by providing reliable data rates and services. The IEEE 802.22 standard is defined as the first dependable and high-performance cognitive radio network resolution to allow SUs to reuse TV white spaces in the VHF and UHF spectrums, for example [2].

The rapid increase in the number of remote users in close proximity to the radio spectrum's static management resulted in a lack of radio range [1]. Within the next five years, almost fifty billion smart devices will be connected. Those devices are most likely responding to a desire for Internet access [2].

The management of the radio spectrum will not be adequate to enable get these devices into the services. With this challenge, it can be seen that a few portions of the radio range

are intensely utilized whereas a few others are seldom utilized. Sharing the radio spectrum among users can lead to unwelcome benefit denial scenarios, but it's to be expected that there will be a number of issues that need to be resolved. The scarcity of radio range has recently become one of the most urgent concerns, prompting future plans to enquire about obtaining new algorithms to address the problems. As a result, new methodologies have been implemented to overcome the spectrum difficulties. To solve this problem, several ways are being investigated, one of which is the use of cognitive radio modification [3].

To achieve service quality criteria while consuming the least amount of energy, cognitive radio innovation is used in wireless systems to assign user spectrum, take into account the status of the channel spectrum range, and analyze communication parameters [4]. These devices can now use unlicensed spectrum bands in addition to licensed spectrum as long as their approved users are not active.

1.3 Multiple-Input Multiple-Output (MIMO)

Wireless network employs several transmitters and receivers to exchange more extra data simultaneously, these devices can transfer data using multiple antennas on both sides like MIMO technique [18] [19] as shown in Fig (1.3). The MIMO technique supports all devices with 802.11n standers to gain very high speed for sending and transmitting the data more than outcomes without 802.11n. For implementing MIMO network, the network station such as a mobile network devices or the access point (AP) should be supported by this technique. To reach an optimal achievement and spectrum range, both networks such as the station and AP should be sustain MIMO technology. In MIMO technology, the natural radio wave aspect that is described by the multipath can appear. With multipath methods, transmitted data suffering from ceilings, walls of the building, and another object coming to the receiving antenna at different times and several points. Within the past, multipath considered the main reason to cause interference and moderated down wireless signals.

In multipath, the MIMO innovation utilizes different intelligent transmitters and receivers of the data with an included spatial measurement, expanding the range and the performance of the method. [20] [21].

MIMO methods increment the observation signal because this technique allows the ra-

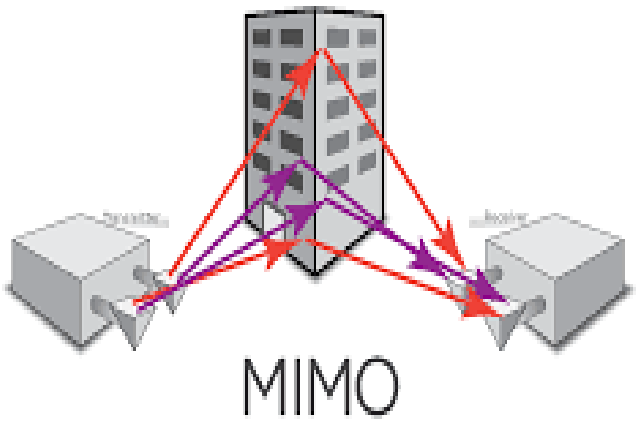

Figure 1.3: Multi-input and Multi-output [7]

dio to combine the receiving information from different places at different times. Smart antennas utilize spatial differing qualities technology, which puts excess radio wires to great use. When an antennas network has spatial streams, the smart antenna can add to the receiver differences in order to raise the range. More antennas in the MIMO technique have usually raised the speeds [22] [23]. MIMO technique is used to increase data transfer speed. For example, the adapter with a certain number of the antennas like three can increase the signal speed to 600 Mbps, while a wireless adapter with a specific number can have a speed of 300 Mbps, two antennas in this case. MIMO technique can be also utilized in router devices with the full support of all specialties of 802.11standers to possibly reach the greatest speed. In addition, legacy wireless network usually apply Single Input Single Output (SISO) method becuase this network can only send or receive one spatial

data stream at the same time [7] [24].

1.4 Cognitive Radio steps

The cognitive radio framework shows a three-process cycle: spectrum sensing step, deciding, and acting step. Fig. (1.4) shows the cognitive radio cycle[6]. The primary procedure is crucial because it organizes the challenge, estimates the parameters, and performs spectrum detection. Since there are different conditions such as multipath fading, shadowing, or changing channel like the channel uncertainty [25] [2], this step can be influenced by these situations. Within the observation preparation step, measurements chosen by the SUs are moreover unknown. Within the other method, SUs produce a choice based on what must already be recognized according to the information source, and this case might have been influenced by the uncertainty within the identified estimations, driving to the off-base opportunities. In the final preparation, the uncertainty value increases during the cognitive radio network cycle, and in some cases, incorrect actions are considered. In this way, uncertainty distribution has impact over all the radio network spectrum steps to decrease the cognitive radio achievement[3]. Therefore, this seems that the demand process is required to handle the uncertain difficulties in the CR network by accurately sensing the spectrum to reach the right activity.

1.5 Dissertation Objectives

This dissertation aims to develop efficient spectrum sensing based on a Bayesian an approach that can enhance the wideband radio spectrum scanning and deal with uncertainty for the cognitive radio systems. To achieve the main goal, the following objectives were pursued:

- Develop new techniques for efficient spectrum sensing in wireless communication.

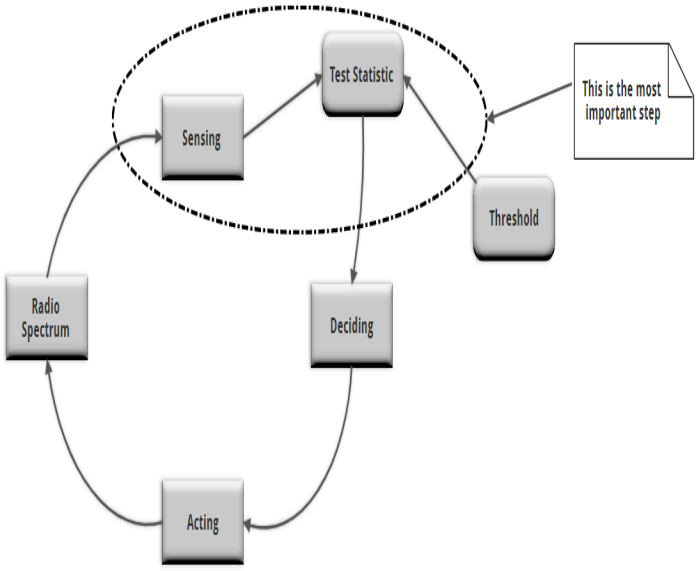

Figure 1.4: Cognitive radio cycle [6]

- Develop efficient estimation techniques for model parameters where few signal samples are exits.
- Implement an extensive framework for estimation and detection of the proposed techniques through simulations.

1.6 Dissertation Contributions

The contributions of the dissertation as:

1) First contribution: First assumption, the parameters of the noise are unknown. In this part, the below contribution shows the steps that have been don for this assumption. This work was **published in the IET Journal, IEEE, Dec. 2018.**

- We developed a method based on the Bayesian model and optimization methods. The suggested method is performed and extensively tested. The simulation outcomes are examined and analyzed to other techniques like the state-of-arts algo-

rithms such as energy detection. To assess the proposed method achievements, multiple metrics used, data samples, the SNR required.

- Spectrum sensing techniques have been contributed in this work by proposing improved new framework techniques based on dynamic sensing threshold to enhance the sensor detection. This technique is evaluated and compared with the state-of-the- arts techniques based on different values of SNR, sensing threshold, and a number of samples. The approach of selecting the sensing threshold allows better assessing the sensing method's performance than with a fixed threshold.

- We showed deep research of this technique that handles an unknown parameters issue in the cognitive radio. The proposed models and methods for dealing with estimation noise parameters were presented. These methods were and carefully investigated in this technique for spectrum sensing in the meaning of cognitive radio networks correlated with other state-of-the-arts techniques in spectrum sensing, among them probabilistic theory such as the Bayesian theory.

- We provided deep research of optimization techniques that deal with parameters estimation in the cognitive radio network. We examined several methods such as convex optimization to deal with uncertainty. possibility theory like the Bayesian theory is utilized to prov id the prior distribution for the unknown parameters. These techniques were compared and deeply analyzed in the context of cognitive radio networks. Metrics for this comparison include robustness to noise, complexity.

- Second contribution: The second assumption is that neither the channel nor the noise characteristics are known. The contribution below shows the procedures that have been taken for this assumption. This paper was approved for publication in the MDPI journal, Electronics, in 2021.

- We proposed a framework for estimation and detection methods using convex optimization for parameters estimation under this assumption. The proposed method is executed and extensively examined. The simulation events are studied and compared to the results of several state-of-the-art methodologies. Various metrics, such as the likelihood of detection and the probability of misdetection in terms of sample number, the SNR, and the accuracy of the proposed method are obtained. The total number of receiving antennas as well as the total number of sending antennas.

- In this work, Spectrum sensing techniques are provided by proposing framework techniques based on dynamic sensing threshold to enhance the sensor detection. This technique is evaluated and compared with the state-of-the-arts techniques based on different values of SNR, sensing threshold, and a number of samples. The approach of selecting the sensing threshold allows better assessing the sensing method's performance than with a fixed threshold.

- We have performed a proposed approach to handles an **uncertainty** issue in the cognitive radio. We presented a proposed model to deal with the noise and the channel uncertainty. Again, using the probabilistic theory such as the Bayesian theory to analyze the signal with the presence of the channel uncertainty. these proposed approaches were and deeply analyzed in this technique for cognitive radio networks spectrum sensing compared with other state-of-the-arts techniques in spectrum sensing.

- In this case, we implemented an extensive approach of optimization procedures that can deal with the estimation of the parameters in the cognitive radio network. We utilized convex optimization to deal with uncertainty. Since this problem is more complicated to directly and can not be directly solved by using the convex optimization. Thus, we perfomed an extensive mathematics contribution to solve this optimization problem. These proposed approaches were are analyzed and simulated

in the context of cognitive radio networks.

1.7 List of Publications

1. Muthana Al-Amidie, Ahmed Al-Asadi, Athanasios C Micheas, Naz E Islam“Spectrum sensing based on Bayesian generalised likelihood ratio for cognitive radio systems with multiple antennas,” IET communication journal, IEEE, October 2019.

2. Ahmed Al-Asadi, Muthana Al-Amidie, Athanasios C Micheas, Ronald G McGarvey, Naz Eas “Worst case fair beamforming for multiple multicast groups in multicell networks” IET communication journal, IEEE, October 2019.

3. Muthana Al-Amidie, Ahmed Al-Asadi, I. Naz,"Robust B-GLRT for Spectrum Sensing in MIMO Cognitive Radio Networks Under CSI Uncertainty", Electronics, March, 2021

4. Ahmed Al Asadi, Muthana Khallil Al-Amidie, Ronald McGarvey, Naz I.„"Robust Underlay Cognitive Network Download Beamforming In Multiple Users,Multiple Groups Multicell Scenario", Manuscript ID: COM-2020-0116, IET communication journal, 2021.

5. Muthana Al-Amidie, etal ."Modeling of OWAs with Missing Inputs", Submitted to information science, Elseiver, 2021

1.8 Dissertation Organization

The remaining of this dissertation can be seen as follows.

1. In Chapter II, we present a deep review of statistical concepts, including techniques of unknown parameters estimation and challenges. We also analyze the existing

assumption techniques that can be used in our approach in spectrum sensing and we compare their efficiencies such as proper prior distribution based on the Bayesian approach.

2. In Chapter III, we propose a Bayesian sensing approach based on GLRT sensing. We analyze and compare the efficiency of the proposed technique with the existing techniques based on a number of samples that cover most of the GLRT sensing aspects. We also present a deep review of sensing detectors including energy tests.

3. In Chapter IV, we propose a robust Bayesian-GLRT in spectrum sensing technique with the presence of channel uncertainty detection with simulated threshold estimation. We compare its efficiency to the existing spectrum sensing techniques.

4. In Chapter V, We conclude with the dissertation objectives and contributions and highlight the dissertation's future works. In Appendix A, we analyze the use of Dirichlet-Multinomial for the Bayesian approach to dealing with uncertainty in cognitive radio networks. In Appendix B, we presented the derivation of estimation and detection of the framework using a synthetic data set. In Appendix C, we provide a general derivation of Bayesian-GLRT that is solved using optimization algorithms.

Chapter 2

Probability Processes and Bayesian Inference

This chapter includes the following subjects:

- Probability Theory
- Theories and techniques of Bayesian inference.
- Bayesian theory testing and model identification.
- Bayesian computation via variations inference.
- Spectrum Sensing.
- Spatial Diversity
- Convex Optimization.

2.1 Probability Theory

Basic Definitions

A scientific set of numbers changes and a set of events sample space and appointed to every event can be called the foundation of the probability hypothesis. The important fact of the thought is the opportunities of the system of assignment. An *event* is a set of

sample points ω_i obtained by using algebraic laws incorporated with the Boolean algebra rules is formally considered a *sample space*. Let us assume A and B important events, the laws are displayed as follows [1],

$$A \cup B = \{\omega : \omega \in A \text{ or } \omega \in B\} \text{ (union)}$$
$$A \cap B = \{\omega : \omega \in A \text{ and } \omega \in B\} \text{ (intersection)}$$
$$\overline{A} = \{\omega : \omega \notin A\} \text{ (complement)}$$

$$\overline{A \cup B} = \overline{A} \cap \overline{B}.$$

$\emptyset$- considered the null set is recognized as the complement of Ω. The data samples are assumed to be *mutually exclusive* [26][1] [27] when there is no part well-familiar to both samples: $A \cap B = \emptyset$. $\Pr[A_i]$ is the measurements of the associated events with each event A_j , seldom described by π_i, that follows the *probability*.

$$\bullet \Pr[A_j] \geq 0$$
$$\bullet \Pr[\Omega] = 1$$
$$\bullet \text{ If } A \cap B = \emptyset, \text{ then } \Pr[AUB] = \Pr[A] + \Pr[B].$$

The probabilities set Pr allocated to a data point, considered as the *a priori probabilities*. According to assuming, Boolean expressions likelihood can be easily determined. Next, The simple directions ($A \cup B = A \cup (\overline{A}B)$ and $AB \cup \overline{A}B = B$) direct to

$$\Pr[AUB] = \Pr[A] + \Pr[B] - \Pr[A \cap B].$$

Assume knowing that the likelihood of B has occurred with $\Pr[B] \neq 0.$; what is then the likelihood of event A can occur? A given event B, this consideration is identified by the *conditional probability* and is indicated as $\Pr[A|B]$. The conditional probabilities can be completed, were considered a B to be the sample space not only Ω. Obtaining a

distribution where these conditions are regular occurs with these probability assumptions [26]. Then, it can be seen.

$$\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]}.$$

statistically independent of B can be considered the event if $\Pr[A|B] = \Pr[A]$: therefore an existence of the B event will not have an impact on the A likelihood . For an events are independent, their intersection of the events $\Pr[A \cap B]$ can be obtained by the multiplying of the *a priori* probabilities [28] [29] of the events such that $\Pr[A] \cdot \Pr[B]$. This most importan feature , it is an enough for the independency of the several events. it can be seen $\Pr[B|A] = \Pr[A \cap B]/\Pr[A]$ and $\Pr[A|B] = \Pr[A \cap B]/\Pr[B]$, the it can determine the *Bayes' Rule*.

$$\Pr[B|A] = \frac{\Pr[A|B] \cdot \Pr[B]}{\Pr[A]}$$

2.1.1 Probability Density Functions and Random Variables

The number distribution of each sample point (a real or complex) in sample scope can be written by A *random variable* X; mathematically, $X : \Omega \mapsto \mathbb{R}$. Thus, a function with a set of domains that can describe a random variable is named as its range, and this will be a subset of the real data. The discrete-valued presents the range also (particularly if the domain Ω is assumed to be discrete). In this example, the symbolic-valued can be the random variable, while in another case, the symbols are used to point to the integers. Then, the random variable values can then be specified. This can have a continuous-valued random variable when there is a continuous range of the information, it shows an interval on the real-line data. In some situations, a mixed random variable can be a random variable value like *continuous-valued*. The*distribution function* or named *cumulative* can be assigned for continuous, and in case as a discrete (when an ordering available), and combined random variables such as [1] [26].

$$P_X(x) \equiv \Pr[X \leq x].$$

Now, x present the argument of the model distribution function while X presentes the random variable. The Probability distribution functions can be raising the functions: if $A = \{\omega : X(\omega) \leq x_1\}$ and $B = \{\omega : x_1 < X(\omega) \leq x_2\}, \Pr[A \cup B] = \Pr[A] + \Pr[B] \Rightarrow P_X(x_2) = P_X(x_1) + \Pr[x_1 < X \leq x_2]^*$,i.e, we means that $P_X(x_2) \geq P_X(x_1)$, $x_1 \leq x_2$.

When integrated used for the distribution function, it obtains the *probability density function* $p_X(x)$ as follows [1]:

$$P_X(x) = \int_{-\infty}^{x} p_X(\alpha) d\alpha$$

The random variable is in both cases such as mixed, it can be as distribution functions to be discontinuous. Furthermore, the density functions can non-negative because **it has an increasing integral**

2.1.2 The Random Variable function

Assume [1] using a real-valued function when random variables (real). Let $Y = f(X)$; in particular, the sample maps can be considered $f : \Omega \mapsto \mathbb{R} \mapsto \mathbb{R}$ similar to the transferring from sample space Ω to the real line. Making to achieve that Y is a random variable when using the map of this kind from the definition of a random variable. Now, the problem can be for the Y's probabilistic characteristics function. The solution is to obtain the probability density parameters by calculating of the mean and variance , using the *distribution* function.

Now, [1] assuming that f is a monotonical function, the probability distribution func-

tion of Y is

$$\begin{aligned} P_Y(y) &= \Pr[\mathrm{Y} \leq y] \\ &= \Pr[\mathrm{Y} \leq y] \\ &= \Pr[X \leq f^{-1}(y)] \\ &= P_X(f^{-1}(y)) \end{aligned} \tag{2.1}$$

Eq. (2.1) is the principal step; where, f^{-1} is the inverse function. it can be seen from the above question that the underlying part of sample range according to $Y \leq y$ has to be equal to that corresponding to $X \leq f^{-1}(y)$, since f is a clearly monotonically function. Thus, Y's function can be performed by abating the derivative.

$$p_y(y) = \frac{df^{-1}(y)}{dy} Px(f^{-1}(y))$$

This derivation indicates the monotonically reducing functions. The variation is that the set according to $Y \leq$ y now refer to the inverse function $X \geq f^{-1}(x)$. Thus, $P_Y(\mathrm{y}) = 1 - P_X(f^{-1}(y))$. Developing of the probability density function a random variable can be described as [30]:

$$p_y(y) = |\frac{1}{f'(f^{-1}(y))}| Px(f^{-1}(y))$$

Therefore, X is assumed to becomes an monotonically function such as exponential probability density: $p_X(x) = e^{-x}\mathrm{u}(x)$, where $\mathrm{u}(x)$ is a unit-step function. By having $Y = X^2$, it can be seen that

$$p_Y(y) = \frac{1}{2\sqrt{y}} e^{-\sqrt{y}},\ y > 0.$$

Over the positive real line [26], it can be square and also monotonic.

2.1.3 Predicted Values

The random variable *expected to value* [31] X can be assigned as

$$\mathrm{E}[f(X)] = \int_{-\infty}^{\infty} f(x)p_X(x)dx.$$

It can be seen that there are Various important numbers considered as crucial values in the density with specific information for the function f

•$f(X) = X$.

The value of the probability density center can be identified by *expected value* of a random variable. It refers to the expected value such as m_X or m when the density function distribution is clear. The main thing is that the expected value is a number that can not be explained by $(p_X(m))$ and the *linearity* represents a crucial feature of the expected value *linearity*: $\mathrm{E}[aX] = a\mathrm{E}[X], a$ where a scalar[32] [33].

•$f(X) = X^2$.

Where $\mathrm{E}[X^2]$ is the *means squared value* of X, notifying the random variable power.

•$f(X) = (X - m_X)^2$.

This is considered the second characteristic of a random variable, known by the *variance*, normally indicated by σ_X^2. This information for the variance can be considered as $\sigma_X^2 = \mathrm{E}[X^2] - \mathrm{E}^2[X]$ that represents the variance data samples. The *standard deviation* is The square root of the variance σ_X introduces the range of the variable distribution of X. $(X-c)^2$ is the second difference with the minimum value occurring at $c = m_X$ (c obtained by evaluating the derivative and equal it to zero).

•$f(X) = X^n$.

$\mathrm{E}[X^n]$ is expected value of the n^{th} represents the *moment* of the random variable, where $\mathrm{E}[(X - m_X)^n]$ the n^{th} represents the value of the central moment.

•$f(X) = e^{juX}$.

The random variable The *function* is essentially form of the Fourier Transform for the

likelihood density function.

$$\mathrm{E}[e^{jvX}] \equiv \Phi_X(jv) = \int_{-\infty}^{\infty} p_X(x) e^{jvx} dx$$

The random variable values can be defined by deriving the characteristic function, and then at the origin can be evaluated [34] [35].

$$\mathrm{E}[X^n] = j^{-n} \frac{d^n \Phi_X(jv)}{dv^n}$$

$$v = 0$$

2.1.4 Random Vectors

Random variables distribution can be defined as *random vector* X, $\mathrm{X} = \mathrm{co1}[X_1, \ldots, X_L]$. The random variables density function is established similarly to the combinations of random variables. Random vector elements can be presented as expected vector values as shown [36]:

$$\mathrm{E}[\mathrm{X}] = \int_{-\infty}^{\infty} \mathrm{x} p_{\mathrm{X}}(\mathrm{x}) d\mathrm{x} = \mathrm{co1}[\mathrm{E}[X_1], \ldots, \mathrm{E}[X_L]]$$

The *covariance matrix* K_X is a square matrix including every possible variance between each random vector's sample. it can be described as:

$$\mathrm{K}_{ij}^{X} = \mathrm{cov}[X_i, X_j] = \mathrm{E}[X_i X_j^*] - \mathrm{E}[X_i]\mathrm{E}[X_j^*] \; i, j = 1, \ldots, L$$

According to the matrix representation [37], the distribution covariance matrix is represented as $\mathrm{K}_X = \mathrm{E}[(\mathrm{X} - \mathrm{E}[\mathrm{X}])$ (X-E [X])']. it can be seen several points such as this covariance matrix is an asymmetric matrix and, its value is positive-definite when the random vector does not have a zero element. When the real-valued is assigned to the random variable, the diagonal values of the matrix should be the same as the variances of the parts:

$K_{ii}^{X} = \sigma_{X_i}^2$. When the *Circular* random vectors are generally considered to be real-values, complex-values and uncorrelated with identical likelihood, then the $E[|X_i|^2] = 2\sigma_{X_i}^2$ and $E[X_i^2] = 0$. Generally , $\sigma_{X_i}^2$ indicates the variance of the real (or imaginary) element. A "real-values" specific function can be defined to be [35]

$$\Phi_X(jv) = E[e^{jv^t X}]$$

2.1.5 The Gaussian Random Variable

A random element named by *Gaussian random variable** [38] [39] and it will have a probability density function which can be determined as follows:

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x-m)^2}{2\sigma^2}\}$$

Where m is the mean of a distribution, while the variance of the distribution is defined as σ^2. This information can be described as $x \sim\| (m, \sigma^2)$. The principal function of a Gaussian random element Φ_X is provided as:

$$\Phi_X(jv) = e^{jmv} \cdot e^{-\sigma^2 v^2/2}$$

Since this expression is totally dependent on the distribution parameters, Therefore no specific expression for a Gaussian random variable. For example, when the mean is a zero with a unit-variance such that Gaussian distribution ($\Lambda'(0,1)$) the probability *exceed* the value x , defined by $Q(x)$.

$$\Pr[X > x] = 1 - P_X(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\alpha^2/2} d\alpha \equiv Q(x)$$

A system of Q is shown in Fig. 2.1 [31]. Here, the Gaussian distribution where the mean is a real value, and also the variance of the distribution is real values "non-unit", generally

the distribution probability can be given as follows in terms of Q

$$\Pr[X > x] = Q(\frac{x-m}{\sigma}) \ , \ X \sim \Lambda'(m, \ \sigma^2)$$

by Integrating the above equation with Q is bounded (for $x > 0$) , it can be rewritten

$$\frac{1}{\sqrt{2\pi}} \cdot \frac{x}{1+x^2} e^{-x^2/2} \leq Q(x) \leq \frac{1}{\sqrt{2\pi}x} e^{-x^2/2} \tag{2.2}$$

Note, the Gaussian distribution is well identified by *normal* random distribution.

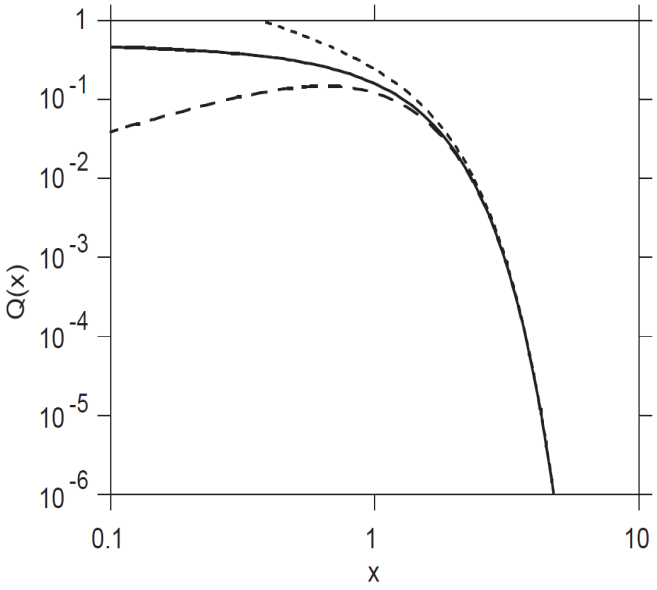

Figure 2.1: The function Q is described in logarithmic spaces. this function reduces quite rapidly [4]

2.1.6 The Central Limit Theorem

Assuming random variable sequence are an independent sequence of the elements, identically distributed, with zero means and finite variances (σ^2), then the Central Limit Theorem can be utilized to explain that the sum $\sum_{l=1}^{L} X_l/\sqrt{L}$ converges into a random with

Gaussian distribution [40] [9].[19]

$$\frac{1}{\sqrt{L}} \sum_{l=1}^{L} X_{l\rightarrow}^{L\rightarrow\infty} \mathcal{N}(0,\ \sigma^2)$$

Generally, this theory is normally applied to simplify computations including *finite* quantities of all random variables except Gaussian random variable. However, this concept is rarely given to the close to the CL rate.

Theorem. Kolmogorov theory , the important currently mathematician theory, is approved that, "The Central Limit Theorem is a active rule for the of studies.

Let σ^2equal to 1, the significant conclusion is that the magnitude difference for the $P(x)$, determined to be the probability of the sum values that exceed x and $Q(x)$. The probability distribution for the variables with a unit-variance passes x, can be contained a number that is inversely related to the square root values of L [41] [37].

$$|P(x) - Q(x)| \leq c \cdot \frac{\mathrm{E}[|X|^3]}{\sigma^3} \cdot \frac{1}{\sqrt{L}}$$

The constant c is identified to be about 0.8 [41]. The skew is defined to be the ratio of the third absolute value of the X_l over standard deviation (cube value), and these values are assigned by 7α, which is totally affected by the distribution of X_l and is not subject to the scale.

As 2.2 can be seen, the right side in Fig. 2.2 , of this equation that is presented as monotonical function. For more detail about the derivation please see.

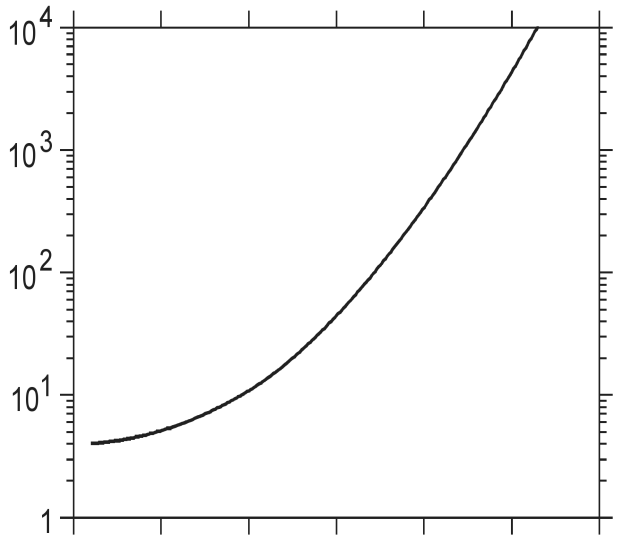

Figure 2.2: The validity limitions for applying the Central Limit Theorem [4]

Figure 2.2 presents the boundaries of valid quantity for implementing the central theorem, where the limited information of the destiny range. To explore the limited numbers, the quantity $L\epsilon^2/2\pi c^2\gamma_X$ is calculated in [42], where ϵ symbolizes the desired approximate percentage error in this Theorem with the number of observations L. This value is obtained on the vertical axis and set the value of x maintaining it, but still be obtained as the normalized value with upper limit on an L-term, means ($x = 1$ implies unit variance). Now, a curve progression can be noticed. In this case, to have an accurate approximation, huge amounts of data are expected.

2.2 Bayesian inference

Bayesian Inference definition, to understand the definition, it can be seen that two critical methods for the statistical signal processing: classical methods named by Frequentist and *Bayesian* approach. To be clear, it is very necessary to state both methods. The difference can be seen as below:

Frequentist method versus Bayesian approach [34] [1]

- In the first approach, the probabilities are determined based on the large-run fre-

quencies. Thus, The principal purpose is to plan systems, which should have long-run frequency guarantees.

• In the Bayesian method, the probability is qualified as subjective degrees of knowledge (believe). The principal purpose is to explain and clarify the expectations.

Morever, the differences between the Bayesian and frequentist methods can be seen are as follows:

	Frequentist	**Bayesian**
Probability is:	**Limiting Relative Frequency**	**Degree of Belief**
Parameter θ is a:	**Fixed Constant**	**Random Variable**
Probability statements are about:	**Procedures**	**Parameters**
Frequency guarantees?	**Yes**	**No**

Table 2.1: Frequentist versus Bayesian Methods [1]

To show the difference, assume that [38] the following example. Consider that $X_1, \ldots, X_n \sim N(\theta,\ 1)$. Then, it needs to clarify the interval of the estimated values C for θ. In the Frequentist method, the interval of the confidence means as follows [43] [44]:

$$C = [\overline{X}_n - \frac{1.96}{\sqrt{n}},\ \overline{X}_n + \frac{1.96}{\sqrt{n}}]$$

Then

$\mathbb{P}_\theta(\theta \in C) = 0.95$ for every $\theta \in \mathbb{R}$.

The variable interval C is the subject of the probability information. Because it's still a data function, the interval can be random as well. The parameter θ is recognized as a fixed value, despite the fact that it should be unknown. C can find the real value with a probability of 0.95, according to the statement.

Furthermore, imagine repeating this experiment a number of times. In fact, it allows users to alter θ at any time. This is how it looks in this case: Natural choose $\theta 2 \rightarrow$ can provide n data points from $N(\theta 2, 1) \rightarrow$ Statistician calculates $C2$ confidence interval

Now, the interval C_j is assigned for the following parameter θ_j, 95 percent of the time.

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I(\theta_i \in C_i) \geq 0.95 \tag{2.3}$$

almost for any sequence $\theta_1, \theta_2, \ldots$.

Bayesian Approach. The Bayesian treats likelihood and it is not depending on frequencies. The unknown parameter θ states the prior distribution $\pi(\theta)$ describing this subjective beliefs about θ. After observing the data $X_1, \ldots, X_n$, then determines the posterior distribution for θ given the data applying Bayes theory [45][46]:

$$\pi(\theta|X_1, \ldots, X_n) \propto \mathcal{L}(\theta)\pi(\theta) \tag{2.4}$$

where $\mathcal{L}(\theta)$ is the likelihood function. Next, it can be found an interval C such that

$$\int_C \pi(\theta|X_1, \ldots, X_n)d\theta = 0.95.$$

With,

$$\mathbb{P}(\theta \in C|X_1, \ldots, X_n) = 0.95.$$

This is a degree-of-belief probability description of the θ presented in the data. the results seem not the same as2.3. If it is replicated within this experiment several times, the confidence intervals will not catch the true value of 95 percent.

At described schemes with frequency, the frequentist technique [40] [47]is obtained. The Bayesian approach is concerned with stating and constructing the beliefs of objects. When employing $F(C)$ to indicate frequency probability and express degree-of-belief, a portion of the confusion is to be eliminated. These are two distinct entities, and there's no reason to think its the same. Strangely, it is common to use the same symbol to indicate both types of probability, such as $\mathbb{P}$, which may also cause confusion.

Summary[48] [16]: The Frequentist approach gives plans with a probability of frequency, but Bayesian inference is a mechanism for sustaining and regenerating beliefs.

The confidence interval C can satisfy the following conditions:

$$\inf_{\theta} \mathbb{P}_{\theta}(\theta \in C) = 1 - \alpha$$

where the probability designate to a unplanned interval C. its call $\inf_{\theta} \mathbb{P}_{\theta}(\theta \in C)$ the coverage of the interval C. A Bayesian confidence interval C satisfies

$$\mathbb{P}(\theta \in C | X_1, \ldots, X_n) = 1 - \alpha$$

where the probability points to θ. Following, concrete instances can be provided wherever the posterior probabilities are very different.

Note, There are several characteristics of subjective Bayesians who define probability surely as individual belief levels. Objective Bayesians attempt assumes the prior distributions indicate unity with the information that the posterior is, in some cases. The empirical Bayesians approach obtained the prior distribution from the observations data. When the statistics indicate that the posterior has a greater frequency behavior than the frequency, Bayesians frequently use the Bayesian approaches separately. As a result, the comparison between the "Bayesian" and "frequentist" approaches can be a little ambiguous. The aforementioned leads to serious misunderstandings in statistics, machine learning, and mathematics [49].

2.2.1 The Mechanics of Bayesian Inference

Let $X_1, \ldots, X_n$ be n observations data sampled from a probability density $p(\mathrm{x}|\theta)$. It can be denoted $p(\mathrm{x}|\theta)$ where assuming θ to a random variable, while a $p(\mathrm{x}|\theta)$ is considered to be the conditional probability of X given θ. Howevere, this can be write $p_{\theta}(\mathrm{x})$ if viewing θ as a deterministic variable. Bayesian approach usually can be executed as following [50] [51] :

Bayesian Procedure [42] [49]

1. The probability density can be defined as $\pi(\theta)$ – announced the prior distribution. This represents the beliefs for the parameter θ before takign any observation data.

2. then, a statistical model $p(\mathrm{x}|\theta)$ need to be defined to reflects the expectation for the x condition of θ.

3. After observing the data $\mathcal{D}_n = \{X_1, \ldots, X_n\}$, expectation is updated, and the posterior distribution $p(\theta|\mathcal{D}_n)$ can be calculated. By using the Bayes' method, the posterior probability distribution is redefined as:

$$p(\theta|X_1, \ldots, X_n) = \frac{p(X_1, \ldots, .X_n|\theta)\pi(\theta)}{p(X_1, .., X_n)} = \frac{\mathcal{L}_n(\theta)\pi(\theta)}{c_n} \propto \mathcal{L}_n(\theta)\pi(\theta) \tag{2.5}$$

where $\mathcal{L}_n(\theta) = \prod_{i=1}^{n} p(X_i|\theta)$ is the likelihood function and

$$c_n = p(X_1, \ldots, X_n) = \int p(X_1, \ldots, X_n|\theta)\pi(\theta)d\theta = \int \mathcal{L}_n(\theta)\pi(\theta)d\theta$$

this constant is the normalizing value, it is also named by the evidence. Utilizing a Bayesian estimation by measuring the center of the posterior, it also could be done by using the mean or mode of the posterior distribution, which can be shown in the following [20] [52].

$$\overline{\theta}_n = \int \theta p(\theta|\mathcal{D}_n)d\theta = \frac{\int \theta\mathcal{L}_n(\theta)\pi(\theta)d\theta}{\int \mathcal{L}_n(\theta)\pi(\theta)d\theta}.$$

A Bayesian interval estimate then can determined. For example, for $\alpha \in (0, 1)$,it can get a and b such that

$$\int_{-\infty}^{a} p(\theta|\mathcal{D}_n)d\theta = \int_{b}^{\infty} p(\theta|\mathcal{D}_n)d\theta = \alpha/2.$$

Let $C = (a,\ b)$. Then

$$\mathbb{P}(\theta \in C|\mathcal{D}_n) = \int_a^b p(\theta|\mathcal{D}_n)d\theta = 1 - \alpha,$$

so C is a $1 - \alpha$ this named by credible interval or Bayesian posterior interval. If the parameter θ has multi dimensions, the case is simple and straightforward and a credible region can obtained . Moreover, let $\mathcal{D}_n = \{X_1,\ \ldots,\ X_n\}$ where $X_1, \ldots, X_n$ ~Bernoulli (θ) . Assume we have the uniform distribution $\pi(\theta) = 1$ as a prior, then the posterior [53] is

$$p(\theta|\mathcal{D}_n) \propto \pi(\theta)\mathcal{L}_n(\theta) = \theta^{S_n}(1-\theta)^{n-S_n} = \theta^{s_n+1-1}(1-\theta)^{n-S_n+1-1}$$

where $S_n = \sum_{i=1}^{n} X_i$ is the successes number. Again, a random variable θ within the random interval $(0, 1)$, means, this parameter consist of a Beta distribution with sub parameters α and β as following [54]:

$$\pi_{\alpha,\beta}(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma^{\urcorner}(\alpha)\Gamma^{1}(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

after simplifying the above expression, It can be seen that the posterior distribution for θ can be a Beta function with parameters $S_n + 1$ and $n - S_n + 1$. which is,

$$p(\theta|\mathcal{D}_n) = \frac{\Gamma(n+2)}{\Gamma^{1}(S_n+1)\Gamma(n-S_n+1)}\theta^{(S_n+1)-1}(1-\theta)^{(n-S_n+1)-1}$$

it can be addressed as

$$\theta|\mathcal{D}_n \sim \text{Beta}(S_n + 1,\ n - S_n + 1)\ .$$

Later then [42] [55], it have been reached out the normalizing fixed without correctly explaining the integral $\int \mathcal{L}_n(\theta)\pi(\theta)d\theta$. Since a density function combines to one, we can see,

$$\int_0^1 \theta^{S_n}(1-\theta)^{n-S_n} = \frac{\Gamma^{1}(S_n+1)\Gamma(n-S_n+1)}{\Gamma(n+2)}.$$

The Beta distribution mean $(\alpha,\ \beta)$ is $\alpha/(c\ell + \beta)$. Thus, the Bayes posterior estimator can be:

$$\overline{\theta} = \frac{S_n + 1}{n + 2}.$$

It is common to write $\overline{\theta}$ as

$$\overline{\theta} = \lambda_n \hat{\theta} + (1 - \lambda_n)\overline{\theta}$$

where $\hat{\theta} = S_n/n$ is the maximum estimation of the likelihood , $\overline{\theta} = 1/2$ is the prior mean and $\lambda_n = n/(n + 2) \approx 1$. A 95 posterior interval percent can be obtained by numerically calculating a and b such that $\int_a^b p(\theta|\mathcal{D}_n)d\theta = .95$.

Now, assuming utilizing a nonuniform prior, using the prior $\theta \sim \text{Beta}(\alpha,\ \beta)$. If the steps above are repeated, then it can be found that $\theta|\mathcal{D}_n \sim \text{Beta}(\alpha + S_n,\ \beta + n - S_n)$, means this flat prior distribution is a special case has equality between $\alpha\ \beta$ =1. The posterior mean for this case can be shown as following:

$$\overline{\theta} = \frac{\alpha + S_n}{\alpha + \beta + n} = (\frac{n}{\alpha + \beta + n})\hat{\theta} + (\frac{\alpha + \beta}{\alpha + \beta + n})\theta_0$$

where $\theta_0 = \alpha/(\alpha + \beta)$ is the prior mean.

A representation of this model is shown in Figure 12.1. It can be seen the Bernoulli model to create $n = 15$ data with parameter $\theta = 0.4$. it can be realized that $s = 7$. Thus, maximum likelihood estimation is $\hat{\theta} = 7/15 = 0.47$, which is higher than the true parameter value that is 0.4. Figure 12.1, on the left, shows a prior Beta distribution of $(4, 6)$ with a posterior mode of 0.43, whereas Figure 12.1, on the right, shows a prior Beta distribution of $(4, 2)$ with a posterior mode of 0.67.

In this diagram, the prior distribution is indicated by the red-solid, the likelihood function is indicated by the blue-dashed, and the posterior distribution is indicated by the black-dashed. The true parameter value $\theta = 0.4$, which is shown by the vertical line, can be observed.

Let's see another example, $X \sim \text{Multinomial}\ (\text{n},\ \theta)$ where $\theta = (\theta_1,\ \ldots,\ \theta_K)^T$ be a K-

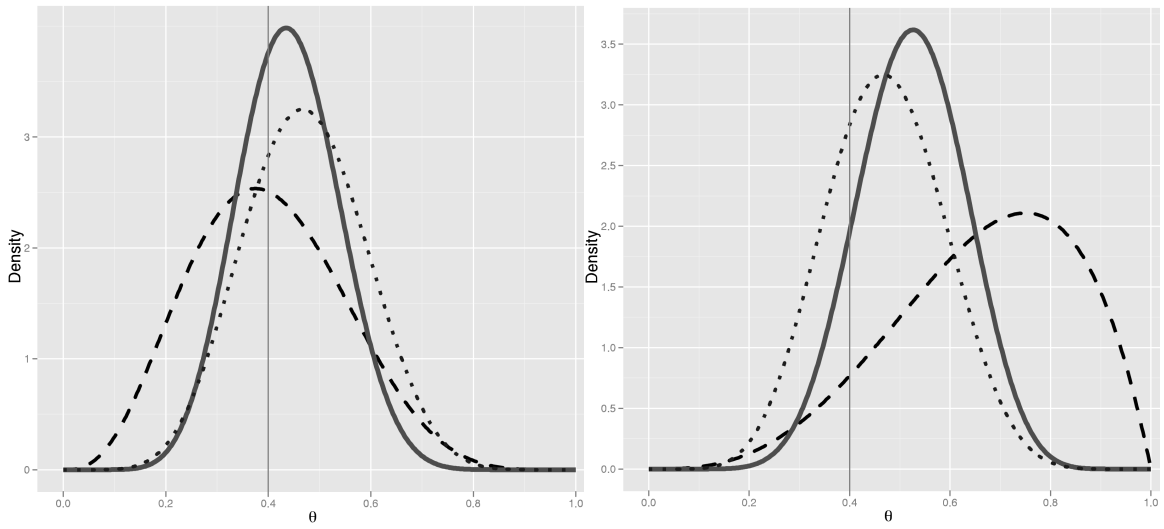

Figure 2.3: Example of Bayesian inference using the Bernoulli data that have different priors [8] [9]

dimensional space parameter ($K > 1$) . A Dirichlet prior is designed with the multinomial, which is recognized as a generalized model of the Beta prior for the Bernoulli distribution to the previous example. The Dirichlet model for K events is notified to be the exponential family distribution on the $K - 1$ dimensional probability single $\triangle_K$ given by

$$\pi_\alpha(\theta) \;=\; \frac{\Gamma(\sum_{j=1}^{K} \alpha_j)}{\prod_{j=1}^{K} \Gamma(\alpha_j)} \prod_{,J=1}^{K} \theta_j^{\alpha_j - 1}$$

The probability simplex $\triangle_K$ can be defined as

$$\triangle_K = \{\theta = (\theta_1, \ldots, \theta_K)^T \in \mathbb{R}^K | \theta_i \geq 0 \text{ for all } i \text{ and } \sum_{i=1}^{K} \theta_{\mathrm{i}} = 1\}.$$

where $\alpha = (\alpha_1, \ldots, \alpha_K)^T \in \mathbb{R}_+^K$ is a positive values of computing components , which are the model parameters. instantly, The sample space of the multinomial with K events, as the group of vertices elements of the K- multidimensional hypercube $\mathbb{H}_K$, may be observed to consist of vectors with 1 and the remainder elements 0:

$$\mathbf{x} = \underbrace{(0, 0, \ldots, 0, 1, 0, \ldots, 0)^T}.$$

Let $X_x = (X_{i1}, \ldots, X_{iK})^T \in \mathbb{H}_K$. If

$\theta \sim \text{Dirichlet}(\alpha)$ and $X_{i'}|\theta \sim \text{Multinomial}(\theta)$ for $i = 1, 2, \ldots, n$,

The posterior satisfies [56] [42]

$$p(\theta|X_1, \ldots, X_n) \propto \mathcal{L}_n(\theta)\pi(\theta) \propto \prod_{i=1}^{n}\prod_{j=1}^{K}\theta_j^{X_{ij}}\prod_{j=1}^{K}\theta_j^{\alpha_j-1} = \prod_{j=1}^{K}\theta_j^{\sum_{i=1}^{n}X_{ij}+\alpha_j-1}$$

It should be noted that the posterior is a Dirichlet distribution as well [49]:

$$\theta|X_1, \ldots, X_n \sim \text{Dirichlet}(\alpha + n\overline{X})$$

where $\overline{X} = n^{-1}\sum_{i=1}^{n}X_i \in \triangle_K$.

For the Dirichlet distribution mean $\pi_\alpha(\theta)$ can be represented by

$$\text{E}(\theta) = (\frac{\alpha_1}{\sum_{i=1}^{K}\alpha_i}, \ldots, \frac{\alpha_K}{\sum_{i=1}^{K}\alpha_i})^T$$

Then, the multinomial posterior mean with Dirichlet prior distribution showing as:

$$\text{E}(\theta|X_1, \ldots, X_n) = (\frac{\alpha_1 + \sum_{i=1}^{n}X_{i1}}{\sum_{i=1}^{K}\alpha_i + n}, \ldots, \frac{(y_K + \sum_{i=1}^{n}X_{iK}}{\sum_{i=1}^{K}\alpha_i + n})^T$$

The posterior mean is shown as giving around the maximum likelihood estimation by choosing some designed probability size for the low-frequency analyses. The parameters $\alpha_1, \ldots, (y_K$ seem as practical counts", which is not the same as the preserved data.

Remark, [57] From the earlier examples, a Dirichlet distribution is prior and posterior. Note, When the case that has the prior and the posterior are in the same family, it can be seen that the prior is conjugate concerning;

2.2.2 Improper,Flat, and Noninformative's Priors

[34][35]

How does the subjective offer the prior $\pi(\theta)$? This is a main key in the Bayesian technique. first, subjective estimations of *theta* that the prior distribution can be considered the object attitude about (before collecting the data). This method is suitable in some situations, but it is challenging in others, particularly when essential parameters are needed. In addition, including subjective opinion into the evaluation as a means of developing scientific ideas as exactly as possible.

To clarify some of noninformative prior, A simple example for a noninformative distribution prior can be used to achieve a flat prior $\pi(\theta) \propto$ constant. For example, in the Bernoulli case, taking $\pi(\theta) = 1$, leads to $\theta|\mathcal{D}_n \sim \text{Beta}(S_n + 1,\ n - S_n + 1)$ as it can be seen earlier, which seems fair. But the unfettered control of flat priors suggests some cases.

Improper Priors. Let $X \sim N(\theta,\ \sigma^2)$ with σ known. it can justify $\mathcal{D}_n = \{X_1,\ \ldots,\ X_n\}$ as the known data. Choosing a flat prior $\pi(\theta) \propto c$ where $c > 0$ is to be a constant values. Note that $\int \pi(\theta)d\theta = \infty$ so this is not a valid probability density, it can tell such a prior an improper prior. Although, by establishing the prior and the likelihood, it may still employ the Bayes' theorem to estimate the posterior density function [35]:

$$p(\theta|\mathcal{D}_n) \propto \mathcal{L}_n(\theta)\pi(\theta) \propto \mathcal{L}_n(\theta)\ .$$

Let $\overline{X} = \sum_{i=1}^{n} X_i/n$. This gives $\theta|\mathcal{D}_n \sim N(\overline{X},\ \sigma^2/n)$, this provide a Bayesian point and confidence interval match accompanied by equality of the frequentist. Generally, unuseful priors are not a hard situation in case the final posterior is defined by the probability distribution.

This flat prior shows the reduction of details of θ before doing the experiment. Now let $\psi = \log(\theta/(1 - \theta)$, this is a change of θ and can be determined the final distribution for ψ, namely

$$p(\psi) = \frac{e^{\psi}}{(1 + e^{\psi})^2},$$

It seems that this prior is not flat. But θ is uninformed, then it would also be confused about

the ψ, so utilizing a flat prior for ψ is demand. Since a flat before reaching a parameter does not represent a flat prior to a changing characteristic of this parameter, the concept of a flat prior is incorrectly set.

Jeffreys' Prior. Harold Jeffreys brightened up who makes it a rule by defining the prior parameter range can set priors distribution, which is invariant transformation. These parameters need to be compatible with the Fisher information, which means the square root of the determinant.

$\pi(\theta) \propto \sqrt{|I(\theta)|}$ where $I(\theta) = -\mathrm{E}[\frac{\partial^2 \log p(X|\theta)}{\partial\theta\partial\theta^T}\theta]$ is the Fisher information.

There are several reasons to believe that this prior distribution can be beneficial prior, however no characteristics are required here. It gives the feature of transformation invariant in the next theory.

Theorem The Jeffreys' prior [56] is transformation invariant.

Proof: Providing the probability function be $p(\mathrm{x}|\theta)$, then ψ match a conversion of θ, it experience that $\pi(\psi) \propto \sqrt{|I(\psi)|}$. This is a simple conclusion based on the variable exchange theory. The fact that the matrix product determinant is the same as the determinant outcome [58].

To state this case, assuming the model of Bernoulli (θ).

$$I(\theta) = \frac{1}{\theta(1-\theta)}.$$

Jeffreys' rule uses the prior

$$\pi(\theta) \propto \sqrt{I(\theta)} = \theta^{-1/2}(1-\theta)^{-1/2}$$

From above it can be seen that a Beta (1/2,1/2) density, which is totally according to a uniform distribution. From those, it can be concluded that Jeffreys' prior is invariant transformation, not expect this is noninformative" more further studies have tried to produce further satisfactory noninformative priors such as in reference priors [9, 7]. When

the model has only one parameter, the reference prior matches Jeffrey's prior. Others can be challenging for popular multi-parameter models.

2.2.3 Conjugate Priors

It can previously be identified about the conjugate priors model, with the multinomial/Dirichlet families and the binomial/Betas. Here, the first form at conjugacy from more further common characteristics, and followed by additional examples [26] [59] [58].

Note, A prior distribution is conjugate prior as long as it is achieved below the sampling method. Supposing that $\mathcal{P}$ is a set of the prior distributions, in which each θ has a $p(\cdot|\theta) \in \mathcal{F}$ over a sample space $\mathcal{X}$ is ready, the posterior distribution is:

$$p(\theta|\mathrm{x}) = \frac{p(\mathrm{x}|\theta)\pi(\theta)}{\int p(\mathrm{x}|\theta)\pi(\theta)d\theta}$$

it is beforehand given that the family $\mathcal{P}$ should be within sampling distributions $\mathcal{F}$ when it takes a conjugate. Note, the group $\mathcal{P}$ is adequately unsatisfactory including it typically supposed to be a demanding parametric group.

The conjugate priors for general exponential family representations are taken into consideration. Considering that $p(\cdot|\theta)$ is a natural exponential model, with μ as a positive measure, the former as shown [41] [60]:

$$p(\mathrm{x}|\theta) = \exp(\theta^{T\mathrm{x}} - A(\theta)) \tag{2.6}$$

where the parameter $\theta \in \mathbb{R}^d$ is d-dimensional, and as well as the parameter space $\Theta \subset \mathbb{R}^d$ is accessible, with

$$\int \exp(\theta^T \mathrm{x} - A(\theta))d\mu(\mathrm{x}) < \infty.$$

The producing result by taking the logconstant $A(\theta)$, then normalizing is shown as follows

[26]:

$$A(\theta) = \log \int \exp(\theta^T \mathrm{x} - A(\theta)) d\mu(\mathrm{x}) \ .$$

The density can be found where a conjugate prior belong to the exponential family: 2.6

$$\pi_{\mathrm{x}_0, n_0}(\theta) = \frac{\exp(n_0 \mathrm{x}_0^T \theta - n_0 A(\theta))}{\int \exp(n_0 \mathrm{x}_0^T \theta - n_0 A(\theta)) d\theta}$$

where $\mathrm{x}_0 \in \mathbb{R}^d$ is a vector and also $n_0 \in \mathbb{R}$ is a scalar.

Then the conjugate can be,

$$p(\mathrm{x}|\theta)\pi_{\mathrm{x}_0, n_0}(\theta) \ = \ \exp(\theta^T \mathrm{x} - A(\theta)) \exp(n_0 \mathrm{x}_0^T \theta - n_0 A(\theta))$$

$$= \ \exp((\mathrm{x} + \mathrm{x}_0)^T \theta - (1 + n_0) A(\theta))$$

$$= \ \exp((1 + n_0)(\frac{\mathrm{x}}{1 + n_0} + \frac{n_0 \mathrm{x}_0}{1 + n_0})^T \theta - (1 + n_0) A(\theta))$$

$$\propto \pi_{\frac{\mathrm{x}}{1+n_0} + \frac{n_0 \mathrm{x}_0}{1+n_0}, 1+n_0}(\theta) \ .$$

Assuming [39] that the prior is as incorporating n_0 observations (virtual)which can be scalar $\mathrm{x}_0 \in \mathbb{R}^d$, Then the posterior parameters can be obtained after imagination one not virtual " observation x are then $n_0' = 1 + n_0$ and

$$\mathrm{x}_0' = \frac{\mathrm{x}}{1 + n_0} + \frac{n_0 \mathrm{x}_0}{1 + n_0}$$

This seems that it is a combination of real and virtual observations. Frequently, if there is n observations $X_1, \ldots, X_n$, then the posterior becomes the form

$$p(\theta|X_1, \ \ldots, \ X_n) \ = \pi_{\frac{n\overline{X}}{n+n_0} + \frac{n_0 \mathrm{X}_0}{n+n_0}, n+n_0}(\theta)$$

$$\propto \exp((n + n_0)(\frac{n\overline{X}}{n + n_0} + \frac{n_0 \mathrm{x}_0}{n + n_0})^T \theta - (n + n_0) A(\theta)) \ ,$$

where $\overline{X} = \sum_{i=1}^{n} X_i/n$. Thus, the parameters of the posterior are $n_0' = n + n_0$ and the mixture

$$\mathrm{x}_0' = \frac{n\overline{X}}{n+n_0} + \frac{n_0\mathrm{x}_0}{n+n_0}.$$

Now, let π_{x_0,n_0} be defined by

$$\pi_{\mathrm{x}_0,n_0}(\theta) = \exp(n_0\mathrm{x}_0^T\theta - n_0 A(\theta)) ,$$

so that

$$\nabla\pi_{\mathrm{x}_0,n_0}(\theta) = n_0(\mathrm{x}_0 - \nabla A(\theta))\pi_{\mathrm{x}_0,n_0}(\theta) .$$

Since

$$\int \nabla\pi_{\mathrm{x}_0,n_0}(\theta)d\theta = \nabla(\int \pi_{\mathrm{x}_0,n_0}(\theta)d\theta) = 0,$$

from which it follows that

$$\mathrm{E}[\nabla A(\theta)] = \int \nabla A(\theta)\pi_{\mathrm{x}_0,n_0}(\theta)d\theta = \mathrm{x}_0 - \frac{1}{\eta_{t0}}\int \nabla\pi_{\mathrm{x}_0,n_0}(\theta)d\theta = \mathrm{x}_0,$$

where the expectation is with respect to $\pi_{\mathrm{x}_0,n_0}(\theta)$. More generally

$$\mathrm{E}[\nabla A(\theta)|X_1, \ldots, X_n] = \frac{n\overline{X}}{n_0+n} + \frac{n_0\mathrm{x}_0}{n_0+n} .$$

According to proper consistency conditions, so the linearity of

$$\mathrm{E}(\nabla A(\theta)|X_1, \ldots, X_n)$$

This is suitable for conjugacy; Diaconis (1979) results in the continuous case.

Theorem[39] [61] Considering a $\Theta \subset \mathbb{R}^d$ is not limited, then assume that X be a size sample to be one from the exponential group $p(\cdot|\theta)$, aμ has an open interval. Now assume

that θ has a prior probability $\pi(\theta)$ with only a use single point. Then the posterior mean of $\nabla A(\theta)$ provided only one observation X which is a linear,

$$\mathrm{E}(\nabla A(\theta)|X) = aX + b,$$

if and only if the prior $\pi(\theta)$ is given by

$$\pi(\theta) \propto \exp(\frac{1}{a}b^T\theta - \frac{1-a}{a}A(\theta))$$

A similar result holds when the μ is supposed to be a discrete value, as clarified in the multinomial family

Both continuous and discrete distributions's models and conjugate priors can be provided in the table. 2.2 and table. 2.3.

Sample Space	Sampling Dist.	Conjugate Prior	Posterior
$\mathcal{X} = \{0, 1\}$	Bernoulli(θ)	Beta(α, β)	Beta($\alpha + n\bar{X}, \beta + n(1 - \bar{X})$)
$\mathcal{X} = \mathbb{Z}_+$	Poisson (λ)	Gamma(α, β)	$operatornameGamma(\alpha + n\bar{X}, \beta + n)$
$\mathcal{X} = \mathbb{Z}_{++}$	Geometric (θ)	Gamma(α, β)	Gamma($\alpha + n, \beta + n\bar{X}$)
$\mathcal{X} = \mathbb{H}_{++}$	Multinomial (θ)	Dirichlet (θ)	Dirichlet($\alpha, n\bar{X}$)

Table 2.2: Conjugate priors for discrete exponential family distributions[2] [3].

Sampling Dist.	Conjugate Prior	Posterior
Uniform (θ)	Pareto (v_0, k)	Pareto $\left(\max\left\{v_0, X_{(n)}\right\}, n + k\right)$
Exponential (θ)	Gamma(α, β)	Gamma($\alpha + n, \beta + n\bar{X}$)
$N\left(\mu, \sigma^2\right)$, known σ^2	$N\left(\mu_0, \sigma_0^2\right)$	$N\left(\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}\left(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{X}}{\sigma^2}\right), \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}\right)$
$N\left(\mu, \sigma^2\right)$, known μ	InvGamma(α, β)	InvGamma $\left(\alpha + \frac{n}{2}, \beta + \frac{n}{2}\overline{(X-\mu)^2}\right)$
$N\left(\mu, \sigma^2\right)$, known μ	ScaledInv- $\chi^2\left(v_0, \sigma_0^2\right)$	ScaledInv- $\chi^2\left(v_0 + n, \frac{v_0\sigma_0^2}{v_0+n} + \frac{n\overline{(X-\mu)^2}}{v_0+n}\right)$
$N(\boldsymbol{\mu}, \Sigma)$, known Σ	$N\left(\boldsymbol{\mu}_0, \Sigma_0\right)$	$N\left(\mathbf{K}\left(\Sigma_0^{-1}\mu_0 + n\Sigma^{-1}\bar{X}\right), \mathbf{K}\right), \mathbf{K} = \left(\Sigma_0^{-1} + n\Sigma^{-1}\right)^{-1}$
$N(\boldsymbol{\mu}, \Sigma)$, known $\boldsymbol{\mu}$	InvWishart (v_0, S_0)	InvWishart $\left(v_0 + n, \mathrm{S}_0 + n\overline{\mathrm{S}}\right), \overline{\mathrm{S}}$ sample covariance

Table 2.3: Conjugate priors for some continuous distributions [2][3].

2.2.4 Bayesian Hypothesis Testing

Supposing need to check the current hypothesis [62] [45]:

$H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$

where $\theta \in \mathbb{R}$. The Bayesian method to measure with putting a prior on H_0 include the parameter θ , then finding $\mathbb{P}(H_0|\mathcal{D}_n)$. This method is to utilize the prior $\mathbb{P}(H_0) = \mathbb{P}(H_1) = 1/2$ (although this is not necessary). Under H_1, it wants a prior distribution for θ, which is expressed by $\pi(\theta)$ .Then, the prior model is assumed to have a point mass 0.5 at $\theta_0$and associated with a continuous function. Now, Bayes' theorem can be given as:

$$\mathbb{P}(H_0|\mathcal{D}_n) = \frac{p(\mathcal{D}_n|H_0)\mathbb{P}(H_0)}{p(\mathcal{D}_n|H_0)\mathbb{P}(H_0) + p(\mathcal{D}_n|H_1)\mathbb{P}(H_1)}$$

$$= \underline{p(\mathcal{D}_n|\theta_0)}$$

$$p(\mathcal{D}_n|\theta_0) + p(\mathcal{D}_n|H_1)$$

$$p(\mathcal{D}_n|\theta_0)$$

$$p(\mathcal{D}_n|\theta_0) + \int p(\mathcal{D}_n|\theta)\pi(\theta)d\theta$$

$$\mathcal{L}(\theta_0)$$

$$\mathcal{L}(\theta_0) + \int \mathcal{L}(\theta)\pi(\theta)d\theta$$

The frequentist and Bayesian approaches [61] [42] achieved similar results in the estimate context where the prior is not necessary. This is true not only in hypothesis testing, but also in measurement, because an incorrect prior results in an infinite value in the denominator of the above experiment. If Bayesian measurement is applied, then it requires very precisely determine the prior $\pi(\theta)$.

2.2.5 Bayesian Linear Models

Following a Bayesian framework, a number of frequentist techniques are used as the maximal a posterior (MAP) estimator. Gaussian linear regression is a prime example [63]:

$$Y = \beta_0 + \sum_{j=1}^{d} \beta_j X_j + \epsilon,\ \epsilon \sim N(0,\ \sigma^2)\ .$$

Here consider that σ is given. Let $\mathcal{D}_n = \{(X_1,\ Y_1),\ \ldots,\ (X_n,\ Y_n)\}$ be observation data samples. Then, The conditional density of $\beta = (\beta_0,\ \beta_1,\ \ldots,\ \beta_d)^T$ as

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} p(y_i|x_{x'},\ \beta) \propto \exp(-\frac{\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{d} \beta_j x_{ij})^2}{2\sigma^2})\ .$$

Using a Gaussian prior $\pi_\lambda(\beta) \propto \exp(-\lambda\|\beta\|_2^2/2)$, the posterior of β can be written as

$$p(\beta|\mathcal{D}_n) \propto \mathcal{L}(\beta)\pi_\lambda(\beta)\ .$$

The MAP estimator $\hat{\beta}^{\text{MAP}}$ takes the form

$$\hat{\beta}^{\text{MAP}} = \arg_\beta \max p(\beta|\mathcal{D}_n) = \arg\min_\beta \{\sum_{i=1}^{n}(Y_i - \beta_0 - \sum_{j=1}^{d} \beta_{J'} X_{ij})^2 + \lambda\sigma^2\|\beta\|_2^2\}.$$

This is the same the regression system, specially when the regularization parameter $\lambda' = \lambda\sigma^2$ is used. Adopting the Laplacian prior $\pi_\lambda(\beta) \propto \exp(-\lambda\|\beta\|_1/2)$, then Lasso framework estimator can be determined.

$$\hat{\beta}^{\text{MAP}} = \arg\min_\beta \{\sum_{i=1}^{n}(Y_i - \beta_0 - \sum_{j=1}^{d} \beta_j X_{i'_j})^2 + \lambda\sigma^2\|\beta\|_1\}.$$

A Bayesian examination of the entire posterior probability distribution $p(\beta|\mathcal{D}_n)$ is used instead of the MAP estimation approach. Because $p(\beta|\mathcal{D}_n)$ does not have an analytic pattern, the simulation approach is used to approximate the posterior in this scenario

[64].

2.3 SPECTRUM SENSING

2.3.1 MIMO Model

Spectrum sensing, which can be accomplished by a cognitive radio network, is one of the most important activities. This approach allows end-users (SUs) to obtain radio network circumstances by determining the availability of PU signals using one or more processes and selecting the frequency band to send the data samples [11]. This general model can be obtained as:

$$y(n) = \begin{cases} w(n)\ \mathrm{H}_0 :\ \text{PU is absent} \\ h * x(n) + w(n),\ \mathrm{H}_1 :\ \text{PU is present} \end{cases} \tag{2.7}$$

where $y(n)$ is the SU received signal and $n = l \ldots N$ is the number of points in the signal. N denotes the data sample, $x(n)$ represents the PU signal, $y(n)$ represents the SU received signal, $x(n)$ represents the PU signal, h represents the sensing channel gain, and $w(n)$ represents the additive white Gaussian noise (AWGN), which has the zero mean and variance. H_0 and H_1 indicate sequentially whether PU is exists with the network or not. The signal detection of the PU band can be performed by utilizing spectrum sensing approach to select one of the hypotheses H_0 and H_1. This can also determined by using the test statistic method, related to a threshold values to obtain the decision of the sensing in order to find the presence of observated data. The sensing decision is achieved as follows:

$$\begin{cases} \text{if } T \geq \gamma,\ H_1 \\ \text{if } T < \gamma,\ H_0 \end{cases} \tag{2.8}$$

Where γ notes the sensing threshold and T indicates the test statistic of the detector.

The SU can locate the PU channel as long as the PU signal is absent. Otherwise, the end-user cannot select that channel on time. Fig. (2.4) performs the common spectrum sensing model [3] [44].

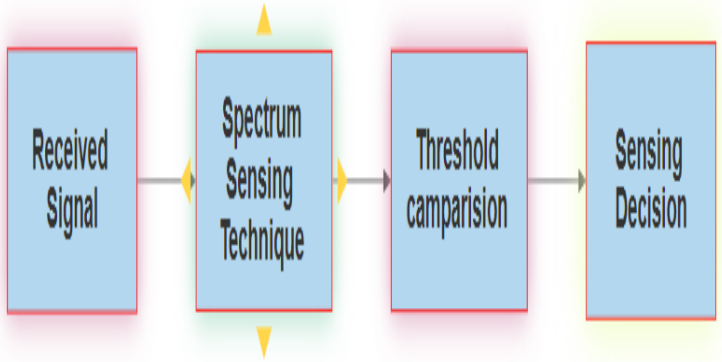

Figure 2.4: The steps of the spectrum sensing network [10]

As mentioned, these techniques can be divided into two categories: cooperative sensing and non-cooperative sensing. in Fig. (2.5) [10][29].

Each end-user makes their own decisions in the non-cooperative method, recommended local sensing, which may subsequently be transferred to an account for other users. To put it another way, because there is no communication among the multiple SUs seeking to detect the same frequency band, spectrum sensing decisions are made locally. This strategy is not only not required for high-complexity hardware components, but it also saves time. This method, however, is sensitive to inaccuracies due to fading, shadowing, noise uncertainty, and interference. It's frequently utilized when only one sensing limit user is available or when secondary users have no way of communicating.

The SUs can coordinate and cooperate with one another using an account of the objectives in the cooperative spectrum sensing approach, and each SU can finalize shared

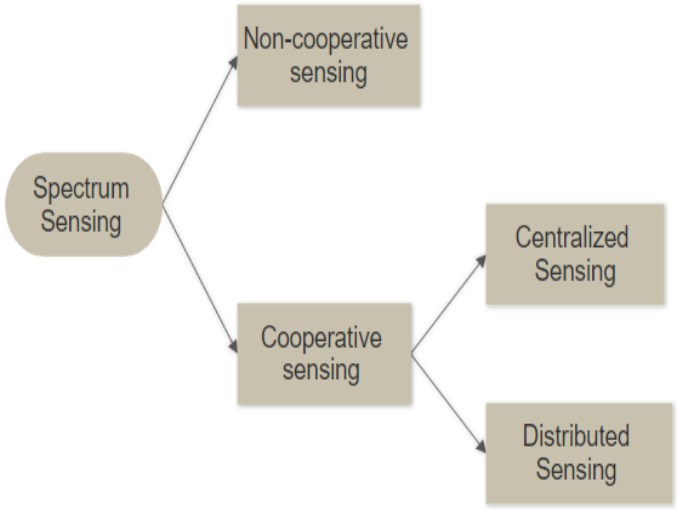

Figure 2.5: Spectrum sensing classification [29]

decision within the same account. This cooperative approach among various SUs can be classified into two types: shared and centralized systems. SUs displace their own local observations and then deliver the results in the distributed or shared technique. Each SU performs its own decision based on the situation of the other SUs in the same frequency band. Because the final decision will be managed by the SUs, no public infrastructure is necessary for this strategy. A centralized system, in which all SUs transmitting their sensing findings should send to a central unit, is another technique. The fusion center is the term of this unit, as depicted in Fig.(2.6).

Spectrum access-based can be decided the fusion center regrading to the received observations data. The decision can be one of the different ways hard or soft utilizing AND/OR laws. Under the final spectra sensing classes selected by the fusion center, SUs can make the sensing utilizing a spectrum sensing. [65][30][10].

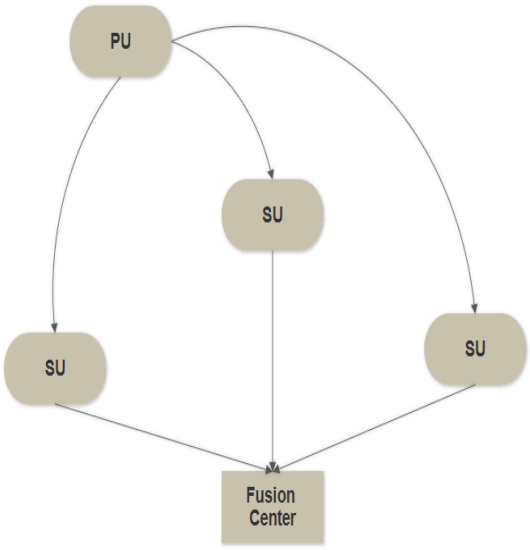

Figure 2.6: Centralized Cooperative spectrum sensing [25]

2.3.2 Spectrum Sensing Methods

In cognitive radio, several spectrum sensing approaches are proposed to verify for the presence of the transmitted user PU signal. These strategies allow SUs to use additional spectrum while maintaining the PU signal's, in case need to sharing the same spectrum. Models of these techniques are presented in Fig. (2.7) shows briefly description for the model of approaches, namely energy, Euclidian distance, autocorrelation model, wavelet based sensing, and matched filter [65].

2.3.3 Energy detection

The energy detector is considered the simplest sensing method compared with other approaches since it does not need any data about the PU signal to proceed with the sensing spectrum. This process starts with examining the observed data energy and analyzing it with a threshold value. The threshold can be obtained corresponding only to the noise power. In this technique, the statistic decision of can be obtained from the squared mea-

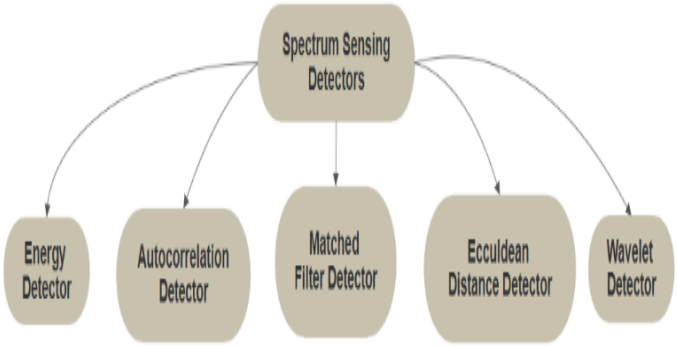

Figure 2.7: Examples of spectrum sensing techniques [13]

sure of the FFT divided by r N samples of the observation signal "SU received signal", as illustrated in Fig. (2.8) [65][3] [44].

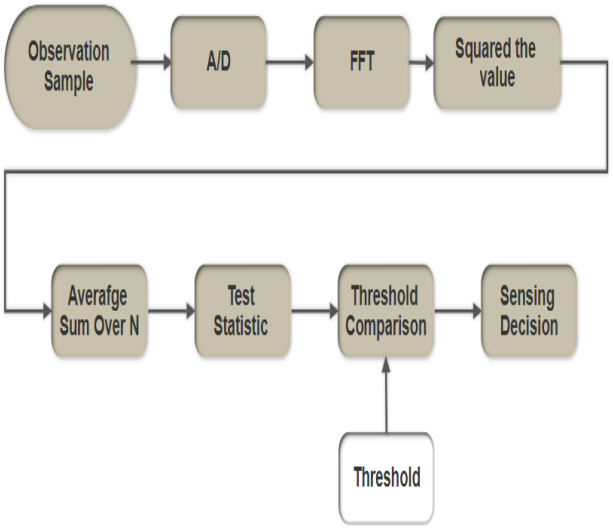

Figure 2.8: Energy detection approach[17]

The detector decision is obtained according to the received signal energy is given by

$$\mathrm{T_{ED}} = \sum_{\mathrm{n=0}}^{\mathrm{N}} \mathrm{y(n)}^2 \; n = l \ldots . N \tag{2.9}$$

Where $y(n)$ is considered as the SU of the received data, N is the number of the data sample, and T_{ED} represents the test method. Thus, the energy detector can be obtained as:

$$\begin{cases} \text{If } T_{ED} \geq \lambda, \text{ PU signal is present} \\ \text{If} T_{ED} < \lambda_l \text{ PU signal is absent} \end{cases} \tag{2.10}$$

The detecting threshold value is indicated by λ. A Gaussian random distribution is assumed for the signal received. As mentioned earlier in this chapter in the central limit theorems, if the number of data samples reaches or exceeds 250 ($N > 250$), the statistic detector will have a central chi-square distribution with degrees of freedom N for the null hypothesis $H0$, but a non-central chi-square model with N degrees of freedom for the $H_1 1$ hypothesis [28]. As a result, the approximated test statistic, T_{ED}, can be generated by:

$$\begin{cases} H_0 : T_{ED} \sim \mathrm{N}(N\delta_w^2, 2N\delta_w^4) \\ H_1 : \tau_{ED} \sim \mathrm{N}(N(\delta_w^2 + \delta_s^2),\ 2N(\delta_w^2 + \delta_s^2)^2) \end{cases} \tag{2.11}$$

Where δ_s^2 indicators PU signal variance, δ_w^2 indicates the variance of the noise distribution and the normal distribution. can be considered by N. The evaluation metrics such as detection probability, P_d, and probability of false alarm, P_{fd} for AWGN channel is defined respectively as:

$$Pd = Q\left(\frac{\lambda - \mathrm{N}((6_{\mathrm{w}}^2 + 6_{\mathrm{s}}^2))}{\sqrt{2\mathrm{N}((\delta_{\mathrm{w}}^2 + \delta_{\mathrm{s}}^2))^2}}\right),\ Pfd = Q\left(\frac{\lambda - \mathrm{N}6_{\mathrm{w}}^2}{\sqrt{2\mathrm{N}\delta_{\mathrm{w}}^4}}\right) \tag{2.12}$$

Where $Q(\cdot)$ indicates Q-function and λ denotes the threshold [25][30]. P_d and P_{fd} are formulated as a function *of SNR*

$$Pd = Q(\frac{\overline{\lambda} - N(1+\gamma)}{\sqrt{2N(1+\gamma)^2}}) \tag{2.13}$$
$$Pfd = Q(\frac{\lambda - \mathrm{N}\delta_{\mathrm{IV}}^2}{\sqrt{2\mathrm{N}\delta_{\mathrm{w}}^4}})$$

where γ denotes to the SNR while $\overline{\lambda}$ denotes the average threshold value, $\overline{\lambda} = \lambda/\delta_w^2$. Here, the threshold value is related the noise power calculation, and this values can be expressed for a target P_{fd} as:

$$\lambda = (Q^{-1}(P_{fd})\sqrt{2N} + N)\delta_w^2 \tag{2.14}$$

Each threshold value related to a pair of $(P_d,\ P_{fd})$,which can be used to describe a receiver's operating curve (ROC). For various threshold settings, this curve represents the graphing of true probability in terms of false detection values. [45][66] [67][58][68] [69].

The threshold value for this type of energy detector is a critical parameter; if a detector cannot correctly retrieve its real threshold values, its performance for the spectrum sensing problem decreases. [64][70] [71] [72] [73] [74] [75] [57][38-44]. Since the error of the noise power significantly impacts the sensing achievement, changing estimation methods for the noise power is suggested in [64]. These suggested several ways: Adaptive threshold methods are achieved with the linear arrangement on the threshold based on SINR. This method achieves a better SU rather than the static threshold way but it continues undesirable outcomes of false alarms [71]. [72] presented an adaptive method to obtain the threshold value when unknown AWGN distribution and the false alarm rate assumed to be under any noise level. According to dedicated noise estimating, this technique means the SUs obtain only noise. Enhancing the energy detector is introduced in [73] where misdetection of PU observation signal is available because of an unexpected

decrease in the power of the PU signal, which is managed by using an additional renewed list of the fixed number of sensing events. This approach is appropriated to obtain an average test statistic value. A double-threshold approach is suggested in [74] [75] with the additional goal of acquiring and restricting the detection framework where narrowband signals exist. Another method is proposed for sensing signal power levels across various frequency ranges, improving SU's opportunistic throughput while reducing interference to PU., this approach is based on wideband spectrum sensing.

Since the error of the noise power significantly impacts the sensing achievement, changing estimation methods for the noise power is suggested in [64]. These suggested several ways: Adaptive threshold methods are achieved with the linear arrangement on the threshold based on SINR. This method achieves a better SU rather than the static threshold way but it continues undesirable outcomes of false alarms [71]

2.4 MIMO Channel State Information(CSI)

MIMO starts a new era in wireless communication systems by developing the quality of service and the data rate at the end-users[76], which can maintain the high-quality multimedia transmission audio and video as well as support internet services. The MIMO enhances these new applications due to the main two essential characteristics contributed by MIMO technology which are [58]

2.4.1 Spatial Multiplexing

With this advantage, MIMO can transmit multiple streams of data simultaneously instant by same frequency within the same phase through several transmit antennas and can be received by several receiving antennas. This innovation develops the capacity of the wireless communication system and leads to meaningful improvement in data rate because the

data rate is directly proportional to the system capacity

$$\mathbf{C} = \mathbf{B}\log_2(1 + SINR) \tag{2.15}$$

Where **C**is the system capacity, **B** is the system bandwidth that is related to the system rate, and SINR is the signal to inference noise ratio at the receiver.

2.4.2 Spatial Diversity

Determining the proper antenna geometry at the transmitter and the receiver sides can significantly diminish the effect of correlation between the instantaneous channel coefficient seen by the transmitters, and this step can improve the signal quality at the receiver side. With the reduction in channel correlation and applies some advanced techniques, the signals at the receiver sides can be easily separated, which is one of the essential applications of MIMO technology. orthogonally polarizing the antenna at both the receiver and transmitter can reduce the effect of correlation between channel coefficients. The channel state information (CSI) severely affects the MIMO system's performance. To attain a relabel and an efficient algorithm in the MIMO system for efficient utilization of MIMO benefits, a rigid design of the MIMO channel should be specified. The MIMO channel for a wireless communication system with M transmit and N receive antenna can be described mathematically by[76]:

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_{11} & \cdots & \cdots & \cdots & \mathbf{h}_{1M} \\ \mathbf{h}_{21} & \ddots & \ddots & \ddots & \mathbf{h}_{21} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{h}_{1M} & \ddots & \ddots & \ddots & \mathbf{h}_{NM} \end{bmatrix} \tag{2.16}$$

Where $\mathbf{h}_{nm}$ represent the channel gain between the mth transmute antenna and nth re-

ceive antenna as shown in the above matrix. The received signal vector at the receiver is given by:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + n \tag{2.17}$$

Where $\mathbf{y} \in \mathbf{C}^{N\times 1}$ is the received signal vector at the receiver, $\mathbf{H} \in \mathbb{C}^{N\times M}$ is is the channel matrix, and $\mathbf{n} \in \mathbb{C}^{N\times 1}$is the noise vector. Because of its essential of CSI a lot of techniques and algorithms has been investigated to estimate it and the most popular approach to estimate the CSI are: 1) Training based CSI acquisition:
In this process, the transmitter and/or receiver send a pilot signal to the receiver and receive feedback from the receiver. This method is used in a conventional wireless network and the secondary network in the wireless cognitive network. This method cannot estimate the CSI between the secondary and primary users in the cognitive network if there is no agreement between the primary and secondary users. Different algorithms have been requested using this method[77, 56, 78, 79, 80].
2) Blind CSI acquisition:
In this method, the secondary user in the cognitive network recognizes the control signal sent by the primary user and estimates some related CSI between the secondary and primary user. This method is appropriate if the primary user performs CDMA, HSDPA, Wi-Fi, and LTE-A. as a part of its operations [81, 46].

The CSI can be estimated deterministically or statistically. The deterministic estimation is applicable if the measurement or estimation time is small as compared with channel coherence time[82, 83, 52]. The deterministic estimation is suitable for slow fading channels. In fast fading channel, the statistical CSI estimation are used where the measurement

over a long time is averaged [84]

2.5 CSI Uncertainty

Due to channel estimation error, limited feedback, and/or the instantaneous nature of the wireless channel, the estimated CSI is subject to error [85, 29, 28] this led to an inefficient gain of MIMO benefit. The error in CSI can be mathematically represented as:

$$\mathbf{H} = \widehat{\mathbf{H}} + \mathbf{\Delta H} \tag{2.18}$$

Where $\mathbf{H}$ is the actual CSI,$\widehat{\mathbf{H}}$ is the estimated CSI, and $\mathbf{\Delta H}$ is the error in CSI. The error or which is normally described as the uncertainty can be modeled stochastically or deterministically. The designer seeks to develop the outage probability in stochastic representation and reduce the uncertainty effect through some sophisticated mathematical processes. In a deterministic model, the quality of service should be improved either by maintaining a specific signal-to-noise ratio at the receiver side and minimizing the transmitter's transmitted power or improving the signal to interference noise ratio at the receiver for some limited transmitted power by the transmitter. In deterministic models, which of concern in this work, different mathematical set are available such as:

1) Polyhedron set.

2) Ellipsoid, spectral, and Frobenius norm sets.

3) Schatten norm and unitary-invariant sets.

$$\mathbf{H} = \widehat{\mathbf{H}} + \mathbf{\Delta H} \tag{2.19a}$$

$$\|\mathbf{H}\|_F \leq \epsilon \tag{2.19b}$$

Where ϵ represents the error in value in CSI and must be limited to be not more than one.

2.5.1 Interweave Cognitive Network

In this mode [86], the cognitive network sensing the spectrum holes and using some sophisticated signal processing algorithm to use one of the holes. In this mode, the network continuously observes the primary user to release the used frequency to the primary user when it restarts to utilize it. Various works have been achieved to create suitable algorithms for spectrum sensing in a cognitive wireless network.

2.5.2 Underlay Cognitive Network

In this mode of hierarchical sharing network, the secondary user shares the same frequency used by the primary user instantaneously under the rule that the interference from the secondary user to the primary user is under a small predefined level. In this mode, the secondary user must know the interference CSI between the second user and the primary user in order not to affect the quality of service of the primary user.
Using the MIMO technique can significantly improve hierarchical sharing in a wireless network. Precisely, applying MIMO to the secondary user in cognitive overlay networks achieves rigid spectrum sensing by investing in spatial diversity. Also, using MIMO to the primary user does not need to know the CSI between the primary and secondary users. A cognitive network involving MIMO [2] at the secondary user can significantly decrease

the secondary user's interference power to the primary user in an underlay method. In an underlay cognitive network, the secondary user needs to estimate the CSI [30] from the secondary transmitter to the secondary receiver and the CSI from the secondary transmitter to the primary receiver if the primary transmitter resides in the other network.

2.5.3 CSI Accuracy

Different techniques are used to estimate CSI, and different algorithms were invoked for each technique. For an actual CSI estimation equation, multiple simple approaches can be directly used to determines the channel gain matrix. But due to the practical limitations such as the dynamic characteristics of the channel(estimation time larger than the coherence time), limited feedback, and/or estimation error, the gathered CSI is subjected to the error that can be modeled using different mathematical models. In the case that the estimated CSI is not accurate, the conventional methods cannot apply directly to determine the channel gain matrix. Thus, in the channel error technique, two types of solution were developed, which are: 1) Sub-optimal solution: in this solution, the channel gain vectors are extracted under worst-case conditions (lowest quality of service at the receiver) but with less computational complexity [34]. 2) Robust solution: in this solution, the channel uncertainty vectors are extracted robustly by mitigating the effect of CSI error through some sophisticated math manipulations. In this solution, the rapidly generated variable can achieve a better quality of service at the receiver, but the method requires high computational complexity than the sub-optimal method.

2.6 Convex Optimization

Convex optimization is a mathematical topic because of the integration of three mathematical subjects [87]:

1) Convex analysis [88, 89].

2) Optimization [41, 90].

3) Numerical computation [60, 60]

Now it considers as an important tool in Engineering where it can be used to find an efficient and reliable solution for large problems.

The formal problem in convex optimization takes the form:

$$\underset{x}{\text{minimize}}.\mathbf{f}_0(x) \tag{2.20a}$$

$$\text{s.t.}\mathbf{f}_i(x) \leq 0 \forall i = 1, \ldots, m \tag{2.20b}$$

$$\mathbf{f}_k(x) = 0 \forall k = 1, \ldots, n \tag{2.20c}$$

Where x is the variable vector need to find ,$\mathbf{f}_0(x)$ is the objective function and$\mathbf{f}_i(x) \leq 0 \forall i = 1, \ldots, m$ and $\mathbf{f}_k(x) = 0 \forall k = 1, \ldots, n$ is the inequality and equality constraints.

Suppose the objective function and the inequality constraints are convex functions and the equality constraints are affine. In that case, the optimization problem is called a convex optimization problem in the condition that the setting in which the optimization variable lie is a convex set.

The set is convex if a line segment whitens; it contains all the points of the line as shown in figure 2.9. for more details, see [91]

2.7 Successive Convex Approximation

The function is convex if it satisfy the following criteria:

$$\mathbf{f}(\lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2) \leq \lambda \mathbf{f}(\mathbf{x}_1) + (1 - \lambda)\mathbf{f}(\mathbf{x}_2) \tag{2.21}$$

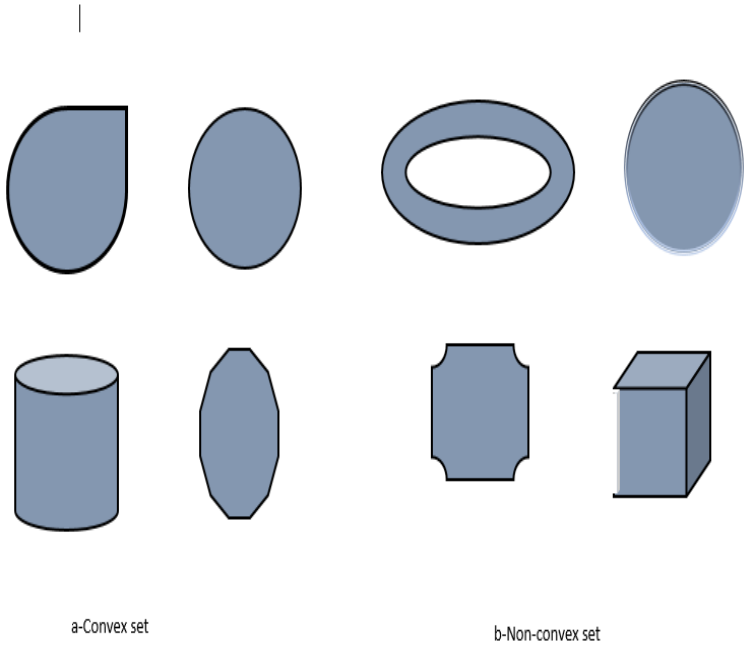

Figure 2.9: Convex and Non-convex set[59].

Where $\lambda \in [0, 1]$ as shown in figure

A function can be concave, and if f1 is concave, then -f1 is convex for more details see [**?**]

2.8 Lagrange duality

Lagrange duality is a method to solve an optimization problem by finding a solution to an equivalent problem. The first problem is known as the primal problem, while the second is called the dual problem [41]. This procedure is followed when there is difficulty in solving the primal problem.

$$\text{minimize.}\mathbf{f}_0(x) \tag{2.22a}$$

$$\text{s.t.}\mathbf{f}_i(x) \leq 0 \quad \forall i \in (1,\ldots,m) \tag{2.22b}$$

$$\mathbf{f}_k(x) = 0 \quad \forall k \in (1,\ldots,n) \tag{2.22c}$$

Where $\mathbf{f}_0(x)$ is the objective function, $\mathbf{f}_i(x)$ is the inequality constraint , and $\mathbf{f}_k(x)$ is the quality constraint. The above is the primal problem, and the equivalent dual problem is:

$$\mathrm{d}(x, \lambda, \mu) = \inf_{\forall i \in (1,\ldots,m), \forall i \in (1,\ldots,m)} (\mathbf{f}_0(x) + \sum_{i=1}^{m} \lambda_i \mathbf{f}_i(x) + \sum_{k=1}^{n} \lambda_i \mathbf{f}_k(x)) \tag{2.23}$$

Where $\lambda_i \forall i \in (1, \ldots, m)$and $\mu_k \forall k \in (1, \ldots, n)$ are the Lagrange multiplier associated with inequality and equality constraints, respectively

The resultant dual function is always convex, even if the original primal function is not convex. Depending on the resultant optimal value from the dual function(d) and the actual optimal value of the primal function(p) there are two types of duality [48]:

1) Strong duality.

2) Weak duality.

These kinds of duality depend on the duality gap between the dual value(d) and primal value(p). If the duality gap is more extensive than or equal to zero (i.e., the objective value of the dual problem is less than or equal to the objective value of the primal problem), then the situation has weak duality; otherwise, if the objective value of the dual and simple problems are the same, then the problem has strong duality as shown in figure 2.10.

2.9 Successive Convex Approximation

Several obstacles in engineering are formulated as non-convex where the objective function is non-convex, and the constraints are convex or non-convex but, most of them obtaining the global solution are computationally expensive since they have no closed-form

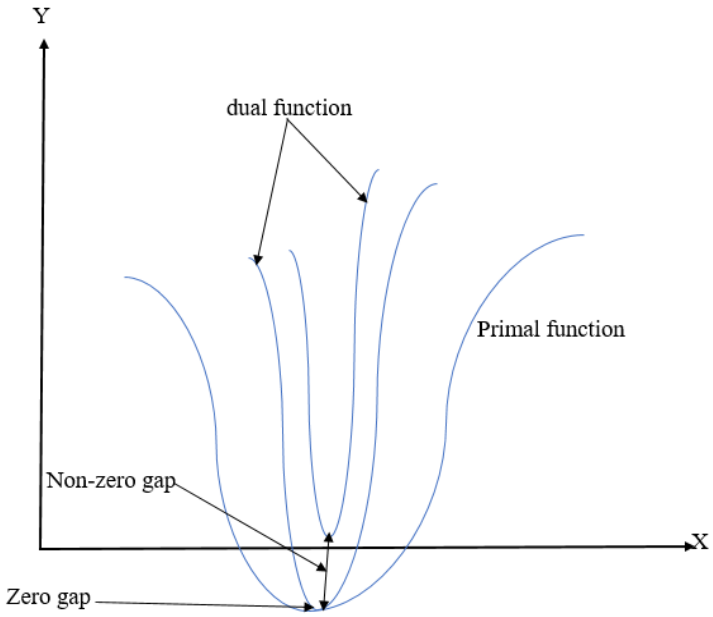

Figure 2.10: strong and weak duality explanation[60].

solutions are available. The goal is to find a method that can be proved to be simple and easy to implement. The successive convex approximation (SCA) method is a powerful method to solve such a problem [92].

The militarization-minimization structure is a remarkable example of the SCA method and can be considered the base to design the SCA method. In this problem, the general is to consider the following non-convex optimization problem.

$$\underset{\mathbf{x}}{\text{minimize}}.\mathbf{f}_0(x) + \mathbf{g}_0(x) \tag{2.24a}$$

$$\text{s.t.}\mathbf{f}_i(x) + \mathbf{g}_i(x) \quad \forall i \in (1,\ldots,m) \tag{2.24b}$$

The function $\mathbf{f}_i(x)$ is non-convex while $\mathbf{g}_i(x)$ is convex. The above problem can be solved using Successive convex approximation method algorithm [21].

For more details about the assumption that needs to be satisfied by the above algo-

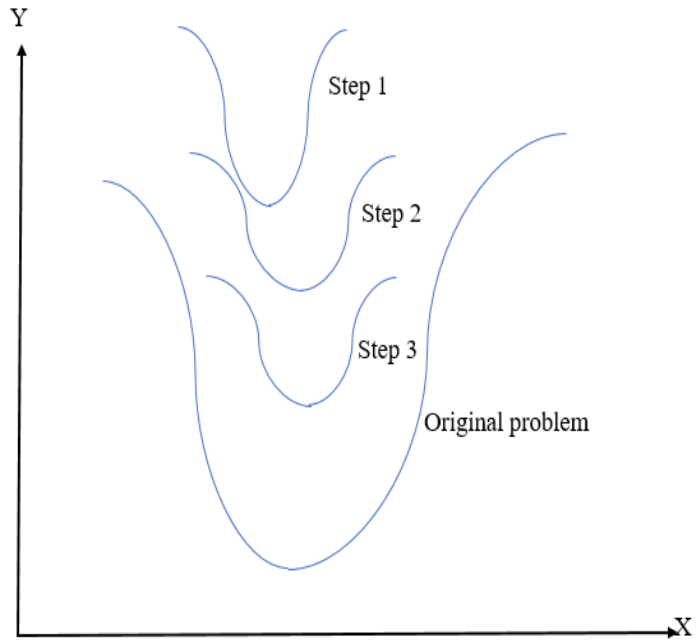

Figure 2.11: convergence of successive convex approximation method[20].

rithm please refer to [32].

2.10 bisection search method

The synonym of this method is the binary search method, interval halving, or dichotomy [33]. The method is a numerical method to find the roots of continuous non-linear function by splitting the interval of the function in half iteratively, as shown in the figure2.12. the main advantages of this method are:

1) It has high speed as compared with another numerical method because it halves the interval each time of iteration.

2) It has higher accuracy in determining the roots of the polynomial as compared with the incremental method.

On the other side, the disadvantages of this method are: 1) If the roots' function is determined are of complex values, then the method fails to estimate these roots.

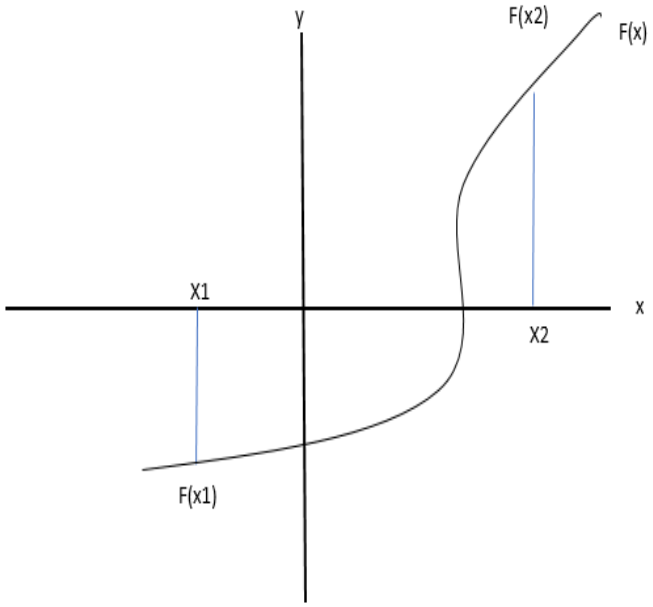

Figure 2.12: Bisection search method starting interval[63].

2) If the roots of a function to be determined lie on the axis the method fails to determine these roots.

3)If the function to which the roots have to be determined is a singularity function, the bisection search method falls to estimate the exact or approximate roots.

The method divides the intervals of a polynomial iteratively to extract the roots, the final intervals should shrine to zero to estimate the exact roots, but this means that the number of iteration must go infinite. To avoid an infinite number of iterations, the iteration must be stopped when the interval reaches a small predefined value, as shown in the figure, 2.13or the number of iteration reach a maximum predefined value.

the method work as follows[93]:

1) let we have function $\mathbf{F}(\mathbf{x})$ as shown in figure below:

2)Chose two value $\mathbf{F}(\mathbf{x_1})$ and $\mathbf{F}(\mathbf{x_2})$ such that the gap between $\mathbf{x}_1$ and $\mathbf{x}_2$ are large enough.

3) Determine $F(\mathbf{x}_3)$ such that:

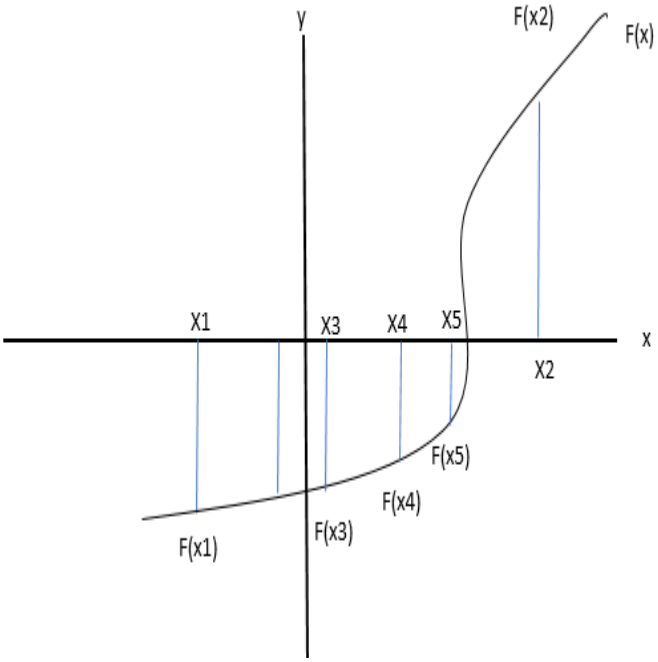

Figure 2.13: Bisection search end interval[64].

$$\mathbf{x}_3 = \frac{\mathbf{x}_1 + \mathbf{x2}}{2} \tag{2.25}$$

4) If $F(\mathbf{x}_3)$ is equal to zero, then the method fails to determine the roots it must restart again by going to step 2 otherwise:

a) If $F(\mathbf{x}_3) < F(\mathbf{x}_2)$then $\mathbf{x}_2 = \mathbf{x3}$

b)If $\mathbf{x1} < \mathbf{x3}$then $\mathbf{x}_1 = \mathbf{x3}$

5) If $F(\mathbf{x}_3)$ is less than or equal to the predefined error or the number of iteration reaches the maximum number of iteration stop the algorithm or go to step 3 and repeat.

Chapter 3

Spectrum sensing based on Bayesian generalised likelihood ratio for cognitive radio systems with multiple antennas

3.1 Introduction

We solve the spectrum sensing problem in cognitive radio systems by deploying multiple antennas when the noise and the principal user signal are independent. We look at the problem of optimal detection and provide an estimating framework where the principal user signal is a complex zero-mean Gaussian distribution. We investigate the generalized likelihood ratio detector (GLRT) in a multiantenna framework using received signal statistics, channel information, and noise prior information to design the optimal GLRT detector for the spectrum sensing problem. The scenario assumes that the noise variance is unknown, but that the secondary user's channel matrix gain is known (SU). The simulation results show that even with a limited number of samples, the proposed Bayesian GLRT detector is optimal, and we show that it outperforms existing state-of-the-art de-

tectors.

Index Index Terms: Bayesian generalized likelihood ratio test (B-GLRT), Bayesian generalized likelihood ratio test (B-GLRT), Bayesian generalized likelihood ratio test (B-GLRT), Bayesian generalized likelihood ratio test (BG-MIMO).

3.2 Literature Review

The GLRT detector is recognized as one of the most critical solutions to the hypothesis testing issue of eq. (3.10).However, only when a sample is a large number, the GLRT give an optimal asymptotically clarification. For reaching a robust test for limited sample size, we succeed **[25]** which links the Bayesian procedure in the estimation of the model parameters based on the Neyman-Pearson (NP) lemma. The resulting detector is known as the Bayesian GLRT (B-GLRT), and it is a reliable solution for solving problems involving a finite number of samples. $f(\mathbf{Y}|\Theta_0)$ and $f(\mathbf{Y}|\Theta_1)$ denote the density model under H_0 and H_1, respectively, whereΘ_0 and Θ_1 are the unknown model parameters. Furthermore, assuming that the unknown distribution parameters have previous distributions of $f(\Theta_0)$ and $f(\Theta_1)$. The framework of the detection estimation problem is illustrated following **[25]** as

$$\xi = \left\{\delta(H_1|\mathbf{Y}), \delta(H_0|\mathbf{Y}), f(\hat{\Theta}_1|\mathbf{Y}, H_1), f(\hat{\Theta}_0|\mathbf{Y}, H_0)\right\}, \tag{3.1}$$

where $f(\hat{\Theta}_0|\mathbf{Y}, H_0)$ and $f(\hat{\Theta}_0|\mathbf{Y}, H_0)$ define the probability density functions for the the two hypotheses H_0 and H_1 respectively. The ξ is recognized the discriminator between the two hypotheses $\delta(H_1|\mathbf{Y})$, $\delta(H_0|\mathbf{Y})$ and the two density functions $f(\hat{\Theta}_1|\mathbf{Y}), f(\hat{\Theta}_0|\mathbf{Y})$. Now let $c_{\mathrm{ji}}(\hat{\Theta}_\mathrm{j}, \Theta_\mathrm{i})$ clarify the cost of estimating the distribution parameters using $\hat{\Theta}_\mathrm{j}$ (i.e we favor H_j when the true values are Θ_i that is, H_i is the true state of the nature). Then the conditional risk for each hypothesis is designated as the average

$$\zeta(\xi|H_i) = \int \left(\delta(H_0|\mathbf{y}) \int f(\hat{\Theta}_0|\mathbf{y}) \mathbb{A}_{0i}(\hat{\Theta}_0, \mathbf{y}) d\hat{\Theta}_0 \right) d\mathbf{y} + \int \left(\delta(H_0|\mathbf{y}) \int f(\hat{\Theta}_1|\mathbf{y}) \mathbb{A}_{0i}(\hat{\Theta}_1, \mathbf{y}) d\hat{\Theta}_1 \right) d\mathbf{y}. \tag{3.2}$$

where

$$\mathbb{A}_{ji}(\hat{\Theta}_j, \mathbf{y}) = \int C_{ji}(\hat{\Theta}_j, \Theta_i) f(\mathbf{y}|H_i, \Theta_i) f(\Theta_i) d\Theta_i. \tag{3.3}$$

Following [29], the cost function can be defined as follows:

$$\begin{aligned} & C_{01}(\hat{\Theta}_0, \Theta_1) = C_{10}(\hat{\Theta}_1, \Theta_0) = 1; \\ & C_{00}(\hat{\Theta}_0, \Theta_0) = C_{11}(\hat{\Theta}_1, \Theta_1) \\ & = \begin{cases} 0 & \left\| \Theta_i - \hat{\Theta} \right\|^2 \leq \Delta << 1 \\ 1 & \text{otherwise} \end{cases}. \end{aligned} \tag{3.4}$$

consequently, the optimization problem can be clarified as follows.

$$\inf_{\xi} \zeta(\xi|H_1), \quad \text{subject to} \quad \zeta(\xi|H_1) \leqslant \alpha, \tag{3.5}$$

where α is the maximum of type I error, reaches to a B-GLRT which uses the MAP techniques to estimate the model parameters (instead of Maximum Likelihood Estimation (MLE)) and is considered as an optimal method under a finite sample size. In particular, the B-GLRT is given by **[25]**

$$\mathbf{LR} = \frac{f\left(\mathbf{y}|H_1, \hat{\Theta}_{MAP_1}\right) f\left(\hat{\Theta}_{MAP_1}\right)}{f\left(\mathbf{y}|H_0, \hat{\Theta}_{MAP_0}\right) f\left(\hat{\Theta}_{MAP_0}\right)} \begin{cases} H_0 & < \\ H_1 & \geq \end{cases} \eta \tag{3.6}$$

where $\hat{\Theta}_{MAP_0}$ and $\hat{\Theta}_{MAP_1}$ denote the MAP estimators of the parameters under H_0 and H_1, respectively. They are defined as

$$\begin{aligned}\hat{\Theta}_{MAP_0} &= \underset{\Theta_0}{\operatorname{argmax}} f(\mathbf{Y}/H_0, \Theta_0) f(\Theta_0), \\ \hat{\Theta}_{MAP_1} &= \underset{\Theta_1}{\operatorname{argmax}} f(\mathbf{Y}/H_1, \Theta_1) f(\Theta_1),\end{aligned} \tag{3.7}$$

denote the MAP technique of the parameters estimation under H_0 and H_1, respectively. The MAP method are the modes of the following posterior distributions:

$$f\left(\hat{\Theta}_{MAP_0}|\mathbf{Y}, H_0\right) \propto f_{H_0}\left(\mathbf{Y}|\hat{\Theta}_{MAP_0}\right) f\left(\hat{\Theta}_{MAP_0}\right), \tag{3.8a}$$

$$f\left(\hat{\Theta}_{MAP_1}|\mathbf{Y}, H_1\right) \propto f_{H_1}\left(\mathbf{Y}|\hat{\Theta}_{MAP_1}\right) f\left(\hat{\Theta}_{MAP_1}\right). \tag{3.8b}$$

In the next section, we derive the optimal values of the unknown model parameters in order to obtain the finite-sample optimal detector based on the Bayesian GLRT approach, when the variance of the noise is unknown for the SU.

The expanding request for great quality and reliable high-speed information administrations which is required by users has displayed numerous challenges that got to be illuminated in arrange to realize the details for the end client. For more details see [94]. The obvious shortage of the wireless spectrum in expansion to the foremost designated spectrum bands being seldom utilized, propel finding modern innovation which is called the Cognitive Radio (CR) paradigm [95][2].

CR [86] [96] [97] [98] [99] is a novel and attractive technology for wireless communications, which can be utilized to improve the allocation of valuable natural resources by exploiting the spectrum band efficiently and actively[100] [95]. In cognitive network, the end user may avoid interference with the essential user by altering the parameter of the transmission or gathering of the cognitive radio to perform dependable communication, and CR is considered as the most critical development in wireless communications that can be utilized to move forward spectrum utilization for future wireless frameworks [101]. Moreover, CR approach facilitate amazing spectrum sensing [102] [103][104] [105] [106], which is a essential step in the process to detect a free spectrum band that can be

filled by a secondary user (SU) [16].

For spectrum sensing, many different approaches have been proposed in [4]. Energy detector (ED) is considered one of the most critical spectrum sensing detectors, which is affected by the error of the noise parameters such as the variance. From ED, it can be seen that the signal to noise ratio (SNR) must be more than a specific threshold to provide a credible performance for primary signal signal detection. This serious challenges was addressed by obtaining an approach that significantly overcome the noise uncertainty [25].

In this method, sharing multiple antennas at the SU's side has become one of the most important methods to establish a reliable approach in signal transmission and spectrum sensing. Moreover, using the sharing methods for multiple antennas is considered the best promising method that can utilized to achieve a good performance for the spectrum sensing, instead of trying to reduce the noise uncertainty by exploiting the free band of observations in a specific domain. In other words, sharing of the spectrum can utilize the multi input and multi output (MIMO) network that has been used by multiple authors in spectrum sensing research area, such as in [107], beside using the optimal Neyman Pearson (NP) and sub-optimal GLRT detectors that are derived for orthogonal frequency division multiplexing (OFDM) method from multiple antennas process.

In [59], the researcher unitized an orthogonal frequency-division multiple access (OFDMA) design with the presence of a varied number of affected eavesdroppers in cognitive radio networks (CRNs). In this paper, an assumption is considered, in which each frame-based transmitted signal is split into two consecutive slots each has an equal duration of time. It can be shown from the first time slot, the eavesdroppers and secondary users (SUs) are observing and listing for the primary base station (PBS) transmission, while in the second time slot, the primary users (PUs) receive PBS signal from the selected SUs. This problem is described as an optimization point and the simulation begins with the proposed strategy to appear expanded secondary normal secrecy rate compared to the conventional method.

In [108] authors present a good performance of the energy detector when using MIMO -CRs compared with applying the single receiving e antenna.

Moreover, [108] shows that the use of MIMO base CRs has a credible performance at low SNRs when compared to the use of a single antenna scheme. In [101], the authors proposed a approach for both signal detection and estimation of the parameters for the secondary user signal in a combined MIMO-CR approach based on Hypothesis testing (HT).

When there are more than two statistical models available in a situation, the HT approach can be used to determine which model best describes the observed data. HT has been proposed by a number of authors for a variety of purposes, including spectrum sensing [109], linear regression [110], and many more. The test statistic of the HT is simply known the detector.

The main point of the detector is for raising the probability of detection compared with reducing the probability of false alarm. Several detectors can be used in HT, such as generalized likelihood ratio test (GLRT) [111].

In GLRT-based detectors, the problem of detection consists of two frames; the first one involves the estimation of the unknown parameters and the second frame is usually proposed to detect the derivation signal. Since the estimation and the signal detection subproblems of the GLRT detector has to the optimal solutions , the GLRT detector becomes one of the best detectors for spectrum sensing using HT approach [47].

In the GLRT of [101] and [112], most of the model parameters can be unknown. Consequently, we require to obtain a proposal that can be a framework to estimate the observation parameters. The Bayesian approach is a recent way for estimating these parameters, as shown in [12], where the authors utilize the Bayesian approach to explain the unknown model parameters by selecting prior distributions for these parameters and then using the estimated values to form the mathematical test. Combing the GLRT and the Bayesian framework is proposed by various authors in determining the signal such

as in [104]. In those articles, the authors proposed the analytical method of the B-GLRT scheme to detect observation signal and estimated the parameters that are obtained by using the Bayesian technique for the given prior distributions related to these parameters.

Valuable detectors according to the B-GLRT approach can be simply created depending on the choice of an appropriate prior model [26]. Prior distribution option is supposed to be a additional step in any Bayesian analysis, since prior distribution consist of entertained *a priori* statistical information for unknown parameters, In the presence of such information about the distribution parameters we can use non-informative priors [54] [113].

In [112], the authors have considered that the B-GLRT utilizing prior knowledge on the unknown model parameters to detect the signal, for two separate scenarios. In the first scenario, only the unknown parameter to the secondary user is channel gain to the SU, while in the second scenario, the model noise variance is unknown to SU. In [104], the authors derived a mathematical approach of the B-GLRT technique for the signal detection and parameter estimation based on binary hypothesis testing, where the Gaussian distribution is considered for the observation data point under two different cases. Maximum *a posteriori probability* (MAP) method has been proposed to estimate the unknown parameters. MAP is considered one of the best methods in spectrum sensing problem when the secondary use receive small number of signal samples [114].

In comparison with art of state methods in the literature, this chapter makes the following contribution: the assumption of this work consider that the noise vector is assumed to be a complex Gaussian distribution with a zero-mean. In addition that a diagonal covariance matrix is defined unknown values. This is a main point of the different assumption from the typical one, namely, that the noise covariance matrix is eaqual to to $\sigma^2 I$, where σ^2 is the noise variance and I is the identity matrix, eg. [112]. The channel gain vallues that is used in our approach is a matrix rather than a vector as assumed defined

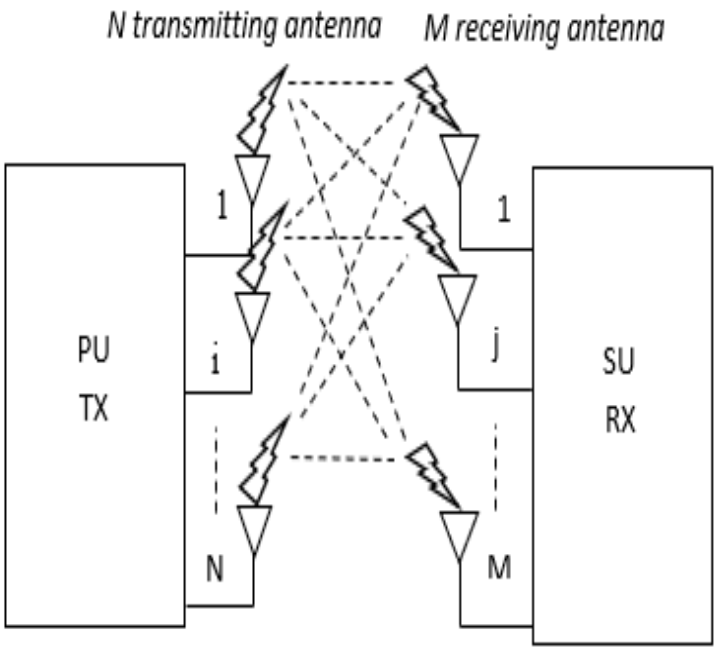

Figure 3.1: MIMO system on the uplink with N transmitting antennas and M receiving antennas

in [104] [112].

3.3 System Model and Problem formulation

In this section, it is assumed that the SU is associated with M receiving user antennas and L is considered samples for each antenna. We define a sequence of i.i.d. data point where each $\mathbf{y}(j)$ is considered a vector of complex numbers, means that $\mathbf{Y} = [\mathbf{y}(1), \ldots, \mathbf{y}(L)] \in \mathbb{C}^{M \times L}$ represent the data point as a complex matrix that consist of the data values at the receiving M antennas. This work further assume that the noise distribution and the primary user (PU) signal are independent.

Since the channel gain can be directly affected by scaling of the signal for primary user , without loss of generality, we consider that the PU signal has unit power. In the HT technique we assume below, the presence and absence of the PU signal is defined by H_0 and H_1, respectively. Thus, the following binary hypothesis detector of the MIMO PU

spectrum sensing detection can be shown as:

$$\mathbf{y} = \begin{cases} \mathbf{n} & H_0 \\ \mathbf{Hx} + \mathbf{n} & H_1 \end{cases} \tag{3.9}$$

Where $\mathbf{H} \in \mathbb{C}^{M \times N}$ present the channel gain, which is complex values between the transmitter and receiver users, $\mathbf{x} \in \mathbb{C}^{N \times 1}$ is the data samples of the transmitted signal, and $\mathbf{n} \in \mathbb{C}^{M \times 1}$ is the values of additive noise samples which is considered to be a zero-mean complex Gaussian distribution with covariance diagonal matrix Σ_0. As in [12] [25], we assume $\mathbf{x}$ is independently distributed from $\mathbf{n}$. Moreover, $\mathbf{x}$ is complex random vector with a zero-mean and the identity covariance matrix for the PU signal. These are the essential assumptions in this chapter that can be utilized to lead to tractable results, i.e., it is straightforward to find the optimal detector under H_0 and H_1.

As a result, the received data sample under each hypothesis can be written as.

$$\begin{aligned} H_0 &: \mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \Sigma_0) \\ H_1 &: \mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \Sigma_1) \end{aligned} \tag{3.10}$$

where $\mathcal{CN}$ is assigned to be the complex Gaussian distribution.

3.4 Proposed B-GLRT detector with unknown noise variance

Here the channel gain matrix is assumed to be known and the noise variance is assumed to unknown for the SU. However, we assume that the SU has entree to the prior statistical distribution of the noise variance. In particular, assume a diagonal matrix $\Sigma_0 = diag(\sigma_{0,1}^2, \ldots, \sigma_{0,M}^2)$, and assume the conjugate prior distributions for $\sigma_{0,i}^2$, $i = 1, \ldots, M$. More precisely, $\sigma_{0,i}^2$ is modeled *a priori* by an Inverse-Gamma model with the shape parameter α_i and the scale parameter β_i. That is, $\sigma_{0,i}^2$ follows an Inv-Gamma(α_i,β_i), $i = 1, \ldots, M$

as clarified in next section.

MAP estimation under H_0

In this section, the joint conjugate prior distribution of the unknown parameters is considered for the spectrum sensing hypothesis. For more elastic calculation, we choose the joint prior for the noise variance given by

$$f\left(\sigma_{0,1}^2,\ldots,\sigma_{0,M}^2\right) = f(\Sigma_0) = \prod_{i=1}^{M} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} \left(\sigma_{0,i}^2\right)^{-\alpha_i-1} \exp\left(\frac{-\beta_i}{\sigma_{0,i}^2}\right), \tag{3.11}$$

In order to determine the B-GLRT detector of eq. (4.23), we require the MAP estimator of the noise variance parameters under H_0, that is,

$$\Sigma_{\mathrm{MAP}_0} = \underset{\Sigma_0}{\operatorname{argmax}} \left[\log(f(\mathbf{Y}|H_0,\Sigma_0)) + \log(f(\Sigma_0))\right]. \tag{3.12}$$

The received data sample vector $\mathbf{y} = [y_1,\ldots,y_M]^T$ has zero-mean circular complex normal distribution as shown as follows,

$$f_{H_0}(\mathbf{y}|\Sigma_0) = \pi^{-M}\left|\Sigma_0^{-1}\right| \exp\left\{-\mathbf{y}^H\Sigma_0^{-1}\mathbf{y}\right\} \tag{3.13}$$

$$= \pi^{-M}\left(\prod_{i=1}^{M}\sigma_{0,i}^2\right)^{-1} \exp\left\{-\sum_{i=1}^{M}\frac{y_i^* y_i}{\sigma_{0,i}^2}\right\}$$

, (3.14)

Since $\Sigma_0^{-1} = \mathrm{diag}(1/\sigma_{0,1}^2,\ldots,1/\sigma_{0,M}^2)$, the data matrix $\mathbf{Y}$ has density :

$$
\begin{aligned}
f_{H_0}(\mathbf{Y}|\Sigma_0) &= \prod_{j=1}^{L} f_{H_0}(\mathbf{y}_j|\Sigma_0) \\
&= \pi^{-ML}\left(\prod_{i=1}^{M}\sigma_{0,i}^2\right)^{-L}\exp\left\{-\sum_{i=1}^{M}\sum_{j=1}^{L}\frac{y_{ij}^* y_{ij}}{\sigma_{0,i}^2}\right\} \\
&= \pi^{-ML}\prod_{i=1}^{M}\left[(\sigma_{0,i}^2)^{-L}\exp\left\{\frac{-1}{\sigma_{0,i}^2}\sum_{j=1}^{L}y_{ij}^* y_{ij}\right\}\right] \qquad (3.15)
\end{aligned}
$$

where $\mathbf{y}_j = [y_{1j}, \ldots, y_{Mj}]^T$. As a result the MAP method of $\sigma_{0,i}^2$ is obtained by maximizing the function:

$$
\log(\sigma_{0,i}^2) = -L\log(\sigma_{0,i}^2) - \frac{1}{\sigma_{0,i}^2}\sum_{j=1}^{L}y_{ij}^* y_{ij} + \log f(\sigma_{0,i}^2|\alpha_i, \beta_i) \qquad (3.16)
$$

with respect to $\sigma_{0,i}^2 > 0$, $i = 1, \ldots, M$. Thus, it is straightforward to see that the MAP of $\sigma_{0,i}^2$ is given by.

$$
\sigma_{0,i}^2 = \frac{1}{L + \alpha_i + 1}\left(\sum_{j=1}^{L}y_{ij}^* y_{ij} + \beta_i\right), \qquad (3.17)
$$

for $i = 1, \ldots, M$ (See Appendix 1 for the proof).

Map estimation under H_1

The received data point vector under H_1 is considered to be complex normal with zero mean and covariance matrix $\Sigma_1 = \mathbf{H}\mathbf{H}^H + \Sigma_0$, where $\Sigma_0 = diag(\sigma_{01}^2, \ldots, \sigma_{0M}^2)$ and Σ_0 is modeled *a priori* using eq. (4.16), following the approach of the previous section. Now the distribution of $\mathbf{y}$ is given by

$$
\begin{aligned}
f_{H_1}(\mathbf{y}|\Sigma_0) &= \pi^{-M} |\Sigma_1|^{-1} \exp\left\{-\mathbf{y}^H \Sigma_1^{-1} \mathbf{y}\right\} \\
&= \pi^{-M} \left|\mathbf{H}\mathbf{H}^H + \Sigma_0\right|^{-1} \\
&\times \exp\left\{-\mathbf{y}^H (\mathbf{H}\mathbf{H}^H + \Sigma_0)^{-1} \mathbf{y}\right\}, \qquad (3.18)
\end{aligned}
$$

So that the gain matrix under H_1 has distribution

$$
\begin{aligned}
f_{H_1}(\mathbf{Y}|\Sigma_0) &= \pi^{-M} \left|\mathbf{H}\mathbf{H}^H + \Sigma_0\right|^{-L} \\
&\times \exp\left\{-\sum_{j=1}^{L} \mathbf{y}_j^H (\mathbf{H}\mathbf{H}^H + \Sigma_0)^{-1} \mathbf{y}_j\right\} \\
&= \pi^{-M} \left|\mathbf{H}\mathbf{H}^H + \Sigma_0\right|^{-L} etr\left\{-\mathbf{R}(\mathbf{H}\mathbf{H}^{\mathbf{H}} + \Sigma_{\mathbf{0}})^{-\mathbf{1}}\right\}, \qquad (3.19)
\end{aligned}
$$

where $\mathbf{R} = \sum_{j=1}^{L} \mathbf{y}_j \mathbf{y}_j^H$. Now the MAP estimator of the noise variance under H_1 is given by solving the following optimization problem:

$$
\begin{aligned}
P1 : \Sigma_{MAP_1} &= \underset{\Sigma_1}{\operatorname{argmax}} \left[\log(f_{H_1}(\mathbf{Y}|\Sigma_1)) + \log(f(\Sigma_0))\right] \\
&= \underset{\Sigma_0 : diagonal}{\operatorname{argmax}} (g(\Sigma_0)), \qquad (3.20)
\end{aligned}
$$

where:

$$
\begin{aligned}
P1 : g(\Sigma_0) &= \underset{\Sigma_0}{\operatorname{argmax}} \left[\log f_{H_1}(Y|\Sigma_1) + \log f(\Sigma_0)\right] \\
&= \underset{\Sigma_0 : diagonal}{\operatorname{argmax}} (L \log \left|\Sigma_1^{-1}\right| - tr[\mathbf{R}\Sigma_1^{-1}] \\
&+ (\alpha + 1) \log \left|\Sigma_0^{-1}\right| - tr[\mathbf{B}\Sigma_0^{-1}]), \qquad (3.21)
\end{aligned}
$$

where $\mathbf{B} = diag(\beta_1, \ldots, \beta_M)$ is the diagonal matrix for the hyperparameters of the con-

jugate priors and ($\alpha_1 = \ldots = \alpha_M = \alpha$) are the hyperparameters of the conjugate priors. P1 is appeared as convex problem and Y is perfectly known for the secondary user. Reaching the solution to P1 is straightforward by using a numerical Semidefinite programming (SDP) optimization package such as CVX [30].

For more details see Appendix 2.

Value of the B-GLRT detector

After finding Σ_{MAP0} and Σ_{MAP1}, the LR function can be calculated using

$$\mathrm{LR} = \frac{f_{H_1}\left(\mathbf{Y}|\Sigma_{MAP_1}\right) f\left(\Sigma_{MAP_1}\right)}{f_{H_0}\left(\mathbf{Y}|\Sigma_{MAP_0}\right) f\left(\Sigma_{MAP_0}\right)},$$

or in the log-scale:

$$\begin{aligned}\mathrm{LLR} = &\log(f_{H_1}\left(\mathbf{Y}|\Sigma_{MAP_1}\right) f\left(\Sigma_{MAP_1}\right)) \\ &- \log(f_{H_0}\left(\mathbf{Y}|\Sigma_{MAP_0}\right) f\left(\Sigma_{MAP_0}\right)),\end{aligned}$$

where

$$\begin{aligned}f_{H_0}(\mathbf{Y}|\hat{\Sigma}_0) f(\hat{\Sigma}_0) = &\frac{1}{\left|\hat{\Sigma}_0\right|^L} \prod_{j=1}^{L} \exp(-\mathbf{y}_j^H \hat{\Sigma}_0^{-1} \mathbf{y}_j) \left|\hat{\Sigma}_0\right|^{-(\alpha+1)} \\ &etr(-\mathbf{B}\hat{\Sigma}_0^{-1})\end{aligned} \tag{3.22}$$

$$f_{H_1}(\mathbf{y}_j|\hat{\Sigma}_1)P(\hat{\Sigma}_1) = \frac{1}{\left|\hat{\Sigma}_1\right|^L}\prod_{j=1}^{L}\exp(-\mathbf{y}_j^H\hat{\Sigma}_1^{-1}\mathbf{y}_j)\left|\hat{\Sigma}_1\right|^{-(\alpha+1)}$$
$$etr(-\mathbf{B}\hat{\Sigma}_1^{-1}) \tag{3.23}$$

By taking the logarithms of eq. (3.22) and eq. (3.23) we have

$$\log(.)|H_0 = -(L+\alpha+1)\log(\left|\hat{\Sigma}_0\right|)$$
$$+ tr(-\mathbf{R}\hat{\Sigma}_0^{-1}) - tr(\mathbf{B}\hat{\Sigma}_0^{-1}) \tag{3.24}$$

and

$$\log(.)|H_1 = -(L+\alpha+1)\log(\left|\hat{\Sigma}_1\right|)$$
$$+ tr(-\mathbf{R}\hat{\Sigma}_1^{-1}) - tr(\mathbf{B}\hat{\Sigma}_1^{-1}) \tag{3.25}$$

and the detector is given by

$$\text{LLR} = -(L+\alpha+1)\log(\frac{\left|\hat{\Sigma}_1\right|}{\left|\hat{\Sigma}_0\right|}) + tr(-\mathbf{R}\hat{\Sigma}_1^{-1})$$
$$-tr(-\mathbf{R}\hat{\Sigma}_0^{-1}) - tr(\mathbf{B}\hat{\Sigma}_1^{-1}) + tr(\mathbf{B}\hat{\Sigma}_0^{-1}) \tag{3.26}$$

3.5 SIMULATION RESULTS

In this section, we clarify the simulation study that demonstrates the performance of the proposed scheme of the GLRT spectrum sensing algorithm. The detector in the previous section is specified for MIMO in CR networks in terms of the primary user detection probability versus average SNR. Assume that the SU is running in a CR network under

a primary network that has a single PU. In the MIMO technique, there are multiple antennas that can be used for wriless communication between the SU and the PU and the SU can use the free spectrum when the PU is not using the spectrum, by applying its peers.

First, we explain how the prior information of the noise parameters affects the performance of the GLRT algorithms, and then provide performance comparisons to other sensing algorithms. We evaluate the performance of the schem detector when the noise is uncorrelated and the unknown noise variance is excited. In other words, the variance of the noise under H_0 is Σ_0 which is a diagonal matrix.

In this simulation results, the number of receiving antennas is considered to be M = 3, and they receive a number of samples L = 25, and transmitting antennas number is assumed to be N=2. The channel gain H is known and the noise variance is simulated from the assumed prior distributions of Section 3. Practically, the model parameter values of the prior distributions are set to be $\alpha_1 = \ldots = \alpha_M = 2$, $\mathbf{B}$ = diag[1.5, 2, 2.5] for the hyperparameter values. The realizations number that is used to generate the probability of false alarm and the probability of detection in order to clarify the performance of detection is 10^5.

The signal-to-noise ratio (SNR) is defined in the detection problem at the SU by,

$$SNR = 10\log\left(\frac{tr(\mathbf{H}\mathbf{H}^H)}{tr(\Sigma_0)}\right). \tag{3.27}$$

In Fig. 2, we illustrate the probability of missed detection (Pm) of the several detectors. The maximum to minimum eigenvalue (MME) υ_{MME} and the energy with minimum eigenvalue (EME) υ_{EME} are considered the heuristic detectors [31] given by

$$v_{EME} = \frac{\lambda_1(\mathrm{S})}{\lambda_M(\mathrm{S})}, \tag{3.28}$$

$$v_{EME} = \frac{\lambda_1(\mathrm{S})}{\lambda_M(\mathrm{S})}, \tag{3.29}$$

where $\lambda_1(\mathrm{S})$ and $\lambda_M(\mathrm{S})$ are the minimum and maximum eigenvalues of the sample covariance matrix $\mathrm{S} \in \mathbb{R}^{M\times M}$ which is defined as.

$$\mathrm{S} = \sum_{j=1}^{L} \left(\mathbf{y}_j - \bar{\mathbf{y}_j}\right)^H \left(\mathbf{y}_j - \bar{\mathbf{y}}\right), \tag{3.30}$$

and the sample mean $\bar{\mathbf{y}} \in \mathbb{R}^{M\times 1}$ is defined as

$$\bar{\mathbf{y}} = \frac{1}{L}\sum_{j=1}^{L} \mathbf{y}_j \tag{3.31}$$

We also make a comparison between the performance of the Hadamard ratio test, the Sphericity test with the proposed detector, as well as other aforementioned detectors. The Hadamard ratio test and Sphericity tests are defined in [13] by:

$$v_{HR} = \frac{\prod_{i=1}^{M} \mathrm{S}_{ii}}{|\mathrm{S}|}, \tag{3.32}$$

$$v_{ST} = \frac{\sum_{i=1}^{M} \mathrm{S}_{ii}/M}{|\mathrm{S}|^{1/M}}, \tag{3.33}$$

Fig. 2 represents the achievement of all detectors when the SNR is developed at a probability false alarm Pfa = 10^{-2}, and $L = 25$. It can be shown that all detectors improve in achievement as the average SNR is increased. Moreover, it can be seen that the proposed optimal detector considerably outperforms the EME and the MME detectors and we can see that the proposed B-GLRT detector has great achievement when matched with the Hadamard ratio test and the Sphericity test.

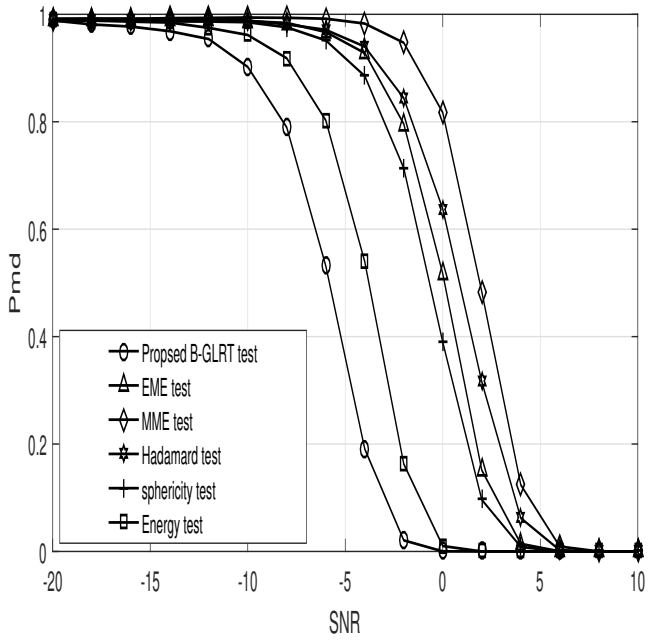

Figure 3.2: Probability of missed detection of different detectors against average SNR for $Pfa = 10^{-2}$, $M = 3$ and $L = 25$.

In Fig. 3 we show the detector respective operating characteristic curves (ROC). In other words, this Figure displays the performance of all detectors in terms of the ROC curves at a false alarm rate of Pfa = 10^{-2}, and an average SNR of -3 dB. We can observe that the proposed B-GLRT schme with the prior that is assumed has ideal performance corresponded with other detectors, such as the EME and MME. The threshold of the case is obtained numerically. Moreover, we can see that the proposed optimal B-GLRT detector defeats the Hadamard ratio test and the Sphericity test.

Next, we study the performance of the proposed detector and the other detectors, when the number of samples of the transmitted signal is changed.

In Fig. 4, we evaluate the probability of missed detection (Pm) for each detector in terms of the number of samples L, at a false alarm rate of Pfa = 10^{-2}, $M = 3$ and the average SNR of -3 dB. From this Figure, we can see that the probability of missed detection Pm of the detectors has an exponential relationship with the number of samples, and also it can be seen that the performance is improved by increasing the number of samples. Moreover,

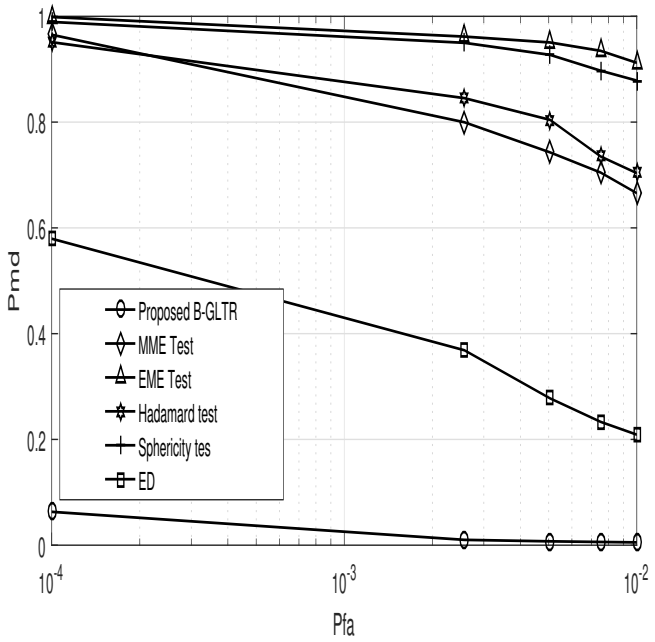

Figure 3.3: Theoretical and simulated ROC performance of the proposed B-GLRT assumed prior at SNR = -3 dB.,M = 3 and L = 25.

the proposed B-GLRT detector performs better than all the other detectors.

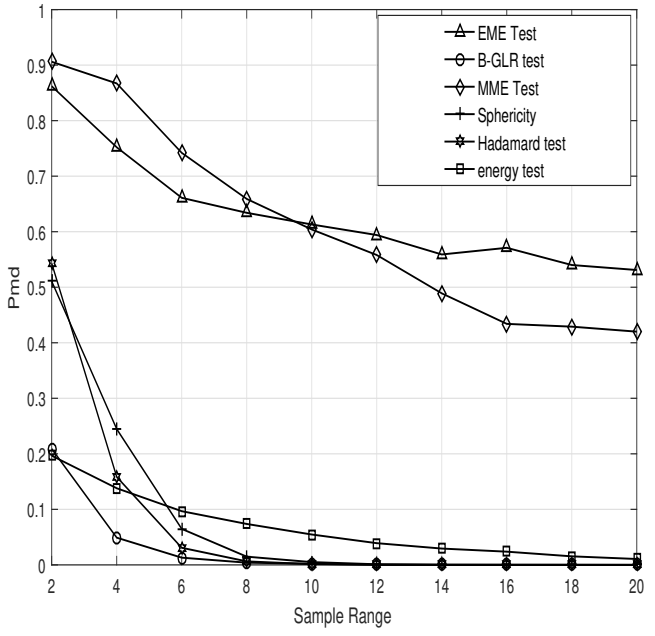

Figure 3.4: Probability of missed detection of the proposed detectors against L for Pfa = 10^{-2}, average SNR = -3 dB.

Fig. 5 shows the performance of all detectors versus the SNR at a probability false alarm Pfa = 10^{-2}, L = 25 and when the hyperparameters have the same values, that is, the shape α = 2 and scale β = 2 for all the hyperparameters. Moreover the noise variance has the same values, i.e. $\sigma_{0,1}^2 = \sigma_{0,2}^2 =, \ldots, = \sigma_{0,M}^2$, where $\sigma_{0,1}^2 \sim$ inv-gamma(α, β). It can be seen that our proposed detector still has a good performance compared with other detectors. Generally, most of the detectors have improved in their performance as the average SNR increased and in comparison with the previous case when noise variance had different values and also **B** were different values.

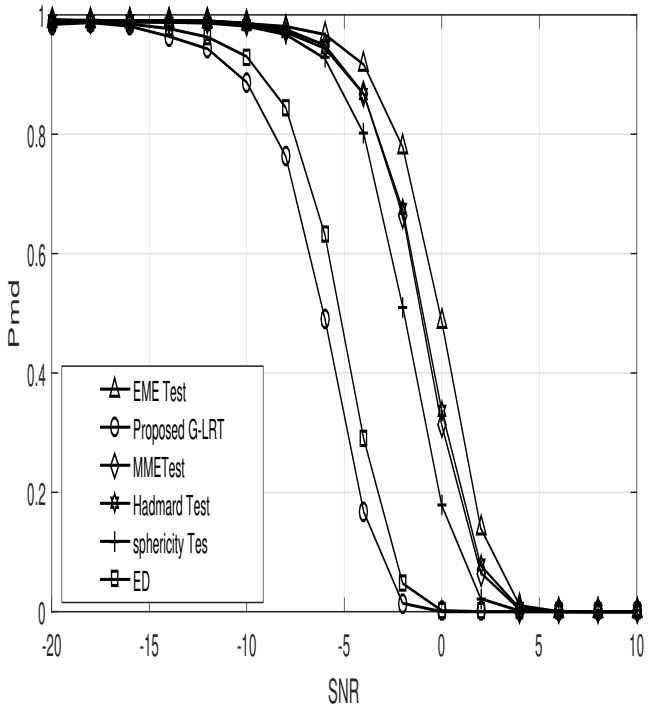

Figure 3.5: Probability of missed detection of the proposed detectors against L when the $\beta_1 = \beta_2 =, \ldots, = \beta_M$ is the same, $\sigma_{01}^2 = \sigma_{02}^2 =, \ldots, = \sigma_{0M}^2$ is the same, Pfa = 10^{-2} and average SNR = -3 dB

.

3.6 Conclusion

In this paper, we show a composite hypothesis problem for spectrum sensing in CR systems by using multiple antennas at the SU. In the presence of unknown model parameters with a finite number of samples, the GLRT is applied to obtain the optimal detectors in the spectrum sensing obstacle, when we need to identify the free spectrum by using the statistics of the obtained signal and conjugate priors for the unknown model parameters of the noise in the presence of the hypothesis test problem. We have derived in detail the corresponding GLRT sensing algorithm cognitive radio. Amazingly, the proposed B-GLRT sensing algorithm offers the best performance under the aforementioned assumptions for unknown parameters of the noise distribution where a small number of signal samples are available. The proposed B-GLRT detector significantly exceeds several state-of-the-art spectrum sensing techniques, such as the EME and the MME, as well as the Hadamard ratio test and the Sphericity detector.

Chapter 4

Robust Spectrum Sensing Detector Based on MIMO Cognitive Radios with Non-Perfect Channel Gain

4.1 Introduction

The spectrum has increasingly become occupied by various wireless technologies. For this reason, the spectrum has become a scarce resource. In the prior chapter 3, we have addressed the spectrum sensing problem by using multi-input and multi-output (MIMO) in cognitive radio systems. We considered the detection and estimation framework for MIMO cognitive network where the noise covariance matrix is unknown with perfect channel state information. In this chapter, we propose a generalized likelihood ratio test (GLRT) for the spectrum sensing problem in cognitive radio where the noise covariance matrix is unknown with non-perfect channel state information. Two scenarios are examined in this study: (i) in the first scenario, the sub-optimal solution of the worst case of the system's performance is considered; (ii) in the second scenario, we present a robust detector for the MIMO spectrum sensing problem. For both scenarios, the Bayesian approach with a generalized likelihood ratio test based on the binary hypothesis problem is

utilized. From the results, It can be seen that our approach provides the best performance in the spectrum sensing problem under specified assumptions. The simulation results also demonstrate that our approach significantly outperforms other state-of-the-art spectrum sensing detectors when the channel uncertainty is addressed.

4.2 Background

Recently, high-speed data services with high quality of service (QoS) are responding to increasing demand by end users and this leads to many challenges in establishing reliable services in current 3G/4G wireless communication systems[115]. Since the spectrum has become a valuable resource for communication applications, it has also become very important to use the spectrum in efficient ways. However, McHenry et al. [115] report an extremely low efficiency for spectrum utilization on the geographic and temporal RF spectrum, and for that reason the demand for a good utilization of RF spectrum has increased and motivated researchers to find the best solutions for this problem. The most important motivation to solve the inefficient spectrum use is the cognitive radio (CR), in which the spectrum will be assigned to the licensed primary user. CR [116, 16, 86, 96, 106, 99, 100, 8] is a attractive and novel communications technology which can be used to enhance the valuable natural resources by efficient use the spectrum. In a cognitive radio network, the spectrum is usually managed by the Federal Communication Commission (FCC) and can be shared between licensed primary users and unlicensed secondary users [62, 50, 117, 118, 119].

Spectrum sensing[120, 121, 122, 123, 124] is considered the operation's key in which the secondary user can identify whether or not the licensed primary user (PU) is using a wireless communication channel in CR networks. A credible spectrum sensing step is a very important stage in detecting a spectrum period that can be subsequently filled by the secondary user (SU). In this case, many different techniques have been produced to

sense the spectrum. Reusing the unoccupied licensed spectrum by unlicensed users in cognitive radio requires an efficient technique to detect the presence of free spectrum without causing interference between the users, as explained with more detail in [125, 63]. Achieving reliable communication also requires an efficient technique for avoiding interference between the unlicensed user and the licensed user. This can be established by adjusting the parameters of the transmission or reception of the cognitive radio [100, 101, 126].

Because of the importance of the problem of spectrum sensing in cognitive radio,many researchers have recently studied this problem. Based on these studies in spectrum sensing, many techniques have been proposed; the most important technique is the energy detector (ED) which represents a simple signal detector in spectrum sensing in cognitive radio [127]. However, ED's performance is still inefficient where there is an error in noise or in channel state information (CSI) compared with other techniques. ED is also considered one of the techniques that depends on hypothesis test (HT)[103, 126].

In spectrum sensing, two or more phases have to be obtained to recognize the signal, so a hypothesis test such as ED is very important in this problem. HT can also be involved in many applications such as multiple model order selection [111, 128, 129, 130, 131], nonlinear regression [110], and many others. Use of a suitable statistic test is considered to be a very important step in HT, i.e. the main goal in achieving an efficient detector is to increase the probability of detection such as [112, 42] in wireless communication. Wireless communication techniques continue to produce the highest data rates. Using multiple transmit and receive antennas can raise the channel capacity without the need for additional power or bandwidth [132]. The first studied was in [53] which clarifies the capacity issue for a single-user MIMO with a Gaussian channel, and then in [133] multiple users has been proposed. In fact, MIMO has recently become the most important technique to achieve good accuracy between users in spectrum sensing wireless transmission and also has widely been exploited in wireless communication [134, 127]. Transmitting a signal

over multiantenna wireless links is affected by an additive interfering signal. Especially when the interferer is found near the transmitter, interference will remain an unknown source for the receiver even though it is learned by the transmitter. Moreover, if there is perfect CSI between the transmitter and receiver, then this technique is well known as Dirty Paper Coding (DPC) in which the interference can be significantly decreased [135, 136, 137, 138].

In this paper [139], a number of optimal detectors have been proposed for non-antipodal signaling spectrum sensing based multiple-input multiple-output cognitive radio (CR) considering uncertainty in the channel. Channel coefficients are modeled as ellipsoidal uncertainty sets while considering the nominal channel estimates as the center of the ellipsoidal. In this case, the detection problem of PU's spectrum is formulated as a second order cone program (SOCP). A proposed robust detector is developed as a closed form solution to solve this problem. This work also present a multicriterion robust detector (MRD) and a relaxed robust detector (RRD) for PU's spectrum sensing in CSI uncertainty scenarios. For these cases, the results show superior performance for the proposed cooperative detectors compared with the traditional detectors such as matched filter (MF) detector. In [140] the authors proposed a novel scheme of spectrum detection in which a robust detector can be derived by existing an uncertainty signal covariance matrix for non-coherent spectrum sensing using multiple-input multiple-output cognitive radio networks. In this work, an eigenvalue perturbation theory is utilized to estimate the signal covariance matrix. Subsequently, the problem is formulated to be an appropriate optimization framework in which the generalized likelihood ratio test (GLRT) [141] based robust test statistic detector (RTSD) and robust estimator-correlator detector (RECD) towards primary user detection is involved. Since this problem is formulated to be a close-form expression for the RTSD and RECD, then it can be efficiently solved by using Karush-Kuhn-Tucker (KKT) conditions. This work demonstrates that the proposed detector has a significant improvement in specified situations compared with the other detectors where the nominal co-

variance matrix is used. In this paper [112], the binary hypothesis problem of spectrum sensing in multiple antenna cognitive radios is addressed using the prior information for unknown parameters. Bayesian technique with proper prior distribution is exploited in order to derive the corresponding detector for three varied scenarios with proper distribution for the unknown parameters. In these cases, the channel uncertainty problem can be seen; besides this, unknown noise variance also exists. The iterative expectation maximization method is used for unknown parameters. Under certain assumptions, this work shows that using Bayesian techniques in addition to the GLRT method gives acceptable results even if the finite number of the observation is available, the results also the proposed outperformance other detectors in multiple antenna sensing spectrum under specified assumptions.

In this paper, we propose a mathematical framework of the GLRT detector for spectrum sensing in cognitive radio. This work assumes an uncertainty exists in the CSI. Two different scenarios are addressed in this paper. Initially, for the first scenario where non perfect CSI is exists with known channel uncertainty while the noise covariance matrix is unknown. In this case, we derive the novel Sub-optimal detector based on binary hypothesis using a generalized likelihood ratio test with a Bayesian approach. Further, in the second scenario where the uncertainty is available in CSI but unknown with unknown noise covariance matrix, we also derive the robust detector against the channel uncertainty based on an optimization framework which can be used to solve a closed form expression for the observed covariance matrix. Maximum *a posteriori* probability (MAP) which is considered one of the most important approaches for spectrum sensing in recent works is utilized in this work for both mentioned cases in order to estimate unknown parameters. Convex optimization is utilized in this paper in order to estimate the unknown variables.

The outline of the rest of the paper is as follows: Section 2 shows the literature review of Bayesian approach for the independent and identically distributed (i.i.d.). Section 3

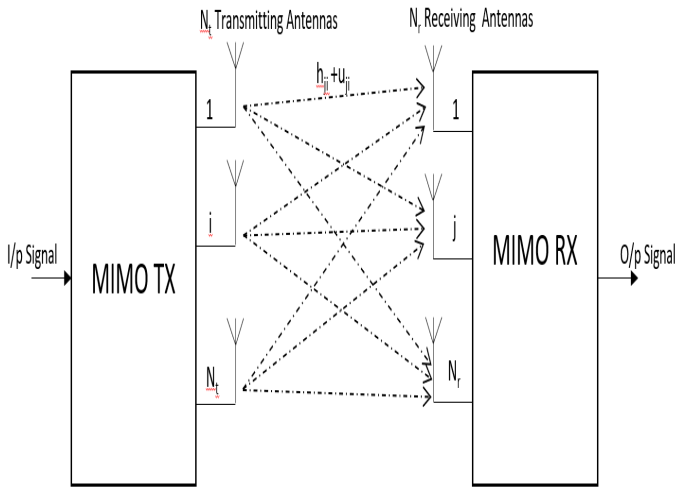

Figure 4.1: MIMO cognitive radio system model consisting of single primary user with N_t transmitting antennas and single secondary user with N_r receiving antennas

shows the system descriptions. Section 4 demonstrate the methodology under different assumptions. In section 5, we show the proposed performance compared with state of art detectors based on the system model. The last section provides concluding remarks.

4.3 Literature Review

The GLRT detector is considered to be an optimal solution in hypothesis testing problem where the signal has large samples. When the signal has a small sample size, the detection problem becomes a big challenge to be solved. Based on [142], a Bayesian approach has been used to solve the detection problem where the signal sample is finite. The framework for Bayesian approach and generalized likelihood ratio test (B-GLRT) is considered an optimal solution to the finite sample problem, which is also considered the combination for estimating the model parameters and detecting the signal. Moreover, $f(\mathbf{X}|\Theta_0)$ and $f(\mathbf{X}|\Theta_1)$ represent the density functions under H_0 and H_1 null and alternative hypothesis respectively [142, 143]. Θ_0 and Θ_1 are considered the unknown parameters. Furthermore, assume that the unknown parameters have prior distributions $f(\Theta_0)$ and $f(\Theta_1)$, respec-

Table 4.1: Table of Notation

Lightface letters	$\triangleq$	a scalar value
Boldface lower-case letters	$\triangleq$	a vector
Boldface upper-case letters	$\triangleq$	a matrix

Variables

$(\cdot)^H$	$\triangleq$	a conjugate transpose
$E[\cdot]$	$\triangleq$	an expected value
diag	$\triangleq$	a diagonal of matrix
$\|\cdot\|$	$\triangleq$	a norm
$tr(\cdot)$	$\triangleq$	a trace
$(\cdot)^*$	$\triangleq$	a conjugate
$\|\cdot\|$	$\triangleq$	a determinant
$etr(\cdot)$	$\triangleq$	an exp $trace(\cdot)$
$\mathcal{CN}$	$\triangleq$	a complex number

tively. The structure of the detection / estimation problem can be defined by

$$\xi = \left\{\delta(H_1|\mathbf{x}), \delta(H_0|\mathbf{x}), f(\hat{\Theta}_1|\mathbf{x}, H_1), f(\hat{\Theta}_0|\mathbf{x}, H_0)\right\}, \tag{4.1}$$

Thus, the conditional risk for each hypothesis can also be defined as

$$\begin{aligned}\delta(\xi|H_i) = \int \left(f(H_0|\mathbf{x}) \int f(\hat{\Theta}_0|\mathbf{x})\mathbb{A}_{0i}(\hat{\Theta}_0, \mathbf{x})d\hat{\Theta}_0\right)d\mathbf{x}+ \\ \int \left(f(H_0|\mathbf{x}) \int f(\hat{\Theta}_1|\mathbf{x})\mathbb{A}_{1i}(\hat{\Theta}_1, \mathbf{x})d\hat{\Theta}_1\right)d\mathbf{x}.\end{aligned} \tag{4.2}$$

where,

$$\mathbb{A}_{ji}(\hat{\Theta}_j, \mathbf{x}) = \int C_{ji}(\hat{\Theta}_j, \Theta_i)f(\mathbf{x}|H_i, \Theta_i)f(\Theta_i)d\Theta_i. \tag{4.3}$$

At the end, the optimization problem becomes:

$$\inf_{\xi}\delta(\xi|H_1), \quad \text{subject to} \quad \delta(\xi|H_1) \leqslant \rho^{'}, \tag{4.4}$$

$C_{ji}(\hat{\Theta}_j, \Theta_i)$ is the cost function and considered to be '1' as defined in [143], the level $\rho^{'}$

is considered the maximum value of error type I under hypothesis H_0. Maximum Likelihood Estimation (MLE) represents the most important methods to estimate the parameters in GLRT detector, while the MAP estimator can be used as an optimal solution under a finite sample size for the model parameters estimation when the B-GLRT is involved to solve spectrum sensing problem. In particular, In [143], the GLRT is given by.

$$\mathbf{LR} = \frac{f\left(\mathbf{x}; H_1, \hat{\Theta}_1\right)}{f\left(\mathbf{x}; H_0, \hat{\Theta}_0\right)} \begin{cases} H_0 & < \\ & \zeta \\ H_1 & \geq \end{cases} \tag{4.5}$$

While the linear ration for B-GLRT can be described as:

$$\mathbf{LR} = \frac{f\left(\mathbf{x}; H_1, \hat{\Theta}_1\right) f\left(\hat{\Theta}_1\right)}{f\left(\mathbf{x}; H_0, \hat{\Theta}_0\right) f\left(\hat{\Theta}_0\right)} \begin{cases} H_0 & < \\ & \zeta \\ H_1 & \geq \end{cases} \tag{4.6}$$

where $\hat{\Theta}_0$ and $\hat{\Theta}_1$ denote the estimation parameters under H_0 and H_1, respectively, ζ is the threshold of detector. These parameters are defined based on the prior distribution as:

$$\hat{\Theta}_t = \underset{\Theta_t}{\operatorname{argmax}} f(\mathbf{x}; H_t, \Theta_t) f(\Theta_t). \tag{4.7}$$

where t refers to the hypothesis case. t can be "0" which refers to null hypothesis or "1" referring to alternative hypothesis, $\hat{\Theta}_t$ denote the MAP estimators of the parameters under H_0 and H_1. The following posterior distributions show the mode's parameter estimators:

$$f\left(\hat{\Theta}_{MAP_t} | \mathbf{X}, H_t\right) \propto f_{H_t}\left(\mathbf{X} | \hat{\Theta}_{MAP_t}\right) f\left(\hat{\Theta}_{MAP_t}\right). \tag{4.8}$$

4.4 System Descriptions

In this section, a multiple-input multiple-output in cognitive radio network with single secondary user is associated with N_r receiving antennas and N_t transmuting antennas as

shown in Fig. 1. In this work, we assume that the noise and PU signal are independent of each other under alternative hypothesis H_1, and also since any scaling of primary user signal can be directly effect channel gain. These assumptions do not lead to loss of generality. In many engineering applications, it can be seen that the detection problem is addressed when an additive Gaussian noise and the channel error are present, such as in [112, 124]. H_1 and H_0 are considered the hypotheses of the presence and absence of the signal respectively.

The basic system model for this scenario at the n^{th} time instant is described as:

$$\mathbf{X} = \begin{cases} \eta & H_0 \\ \mathbf{HS} + \eta & H_1 \end{cases} \tag{4.9}$$

where $\mathbf{H} \in \mathbb{C}^{N_r * N_t}$ is the channel matrix between secondary receiver and the PU transmitter, $\mathbf{S} \in \mathbb{C}^{N_t * L}$ is PU signal samples at transmitter antennas, and $\eta \in \mathbb{C}^{N_r * L}$ is the matrix of additive noise samples at i^{th} antenna which is complex random variables with covariance matrix $\mathbf{R}_\eta$. $\mathbf{X} \in \mathbb{C}^{N_r * L}$ is the observation signal which is the Gaussian distribution with mean and covariance matrix corresponding to the hypothesis. Since the assumption is reasonable and customary, the mathematics and derivation of the problem will be simple. From the cognitive radio system model as described in eq. (4.9), the instant signal for the i^{th} receiving antenna can be equivalently described as,

$$\mathrm{x_k} = \mathbf{Hs} + \eta_\mathbf{k}. \tag{4.10}$$

where $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the MIMO channel matrix between N_r transmitting antennas at the primary side and N_r receiving antennas on the secondary side, $\mathbf{x}_k \in \mathbb{C}^{N_r \times 1}$ is a complex vector that contains the observed values at the instant k corresponding to transmitted signal that is corrupted by the noise vector $\eta(k) \in \mathbb{C}^{N_r \times 1}$ which is additive white Gaussian

L represents the number of input sample **s**. Moreover $X_k = [\mathbf{x}_k(1), \mathbf{x}_k(2), \ldots, \mathbf{x}_k(L)]^H \in \mathbb{C}^{L\times 1}, \quad \forall k \in [1, \ldots, N_r]$, $(h)_k{}^H \in \mathbb{C}^{1\times N_t}, \quad \forall k \in [1, \ldots, N_r]$ represents the row channel matrix such that $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{N_r}]$ and $\eta_k \in \mathbb{C}^{L\times 1} \quad \forall k \in [1, \ldots, N_r]$ is the noise vector whose covariance matrix $\mathbf{R}_\eta = E(\eta_k \eta_k{}^H)$.

Due to several practical limitations, the actual channel matrix can not be accurately gathered at the transmitter. Thus, it is subject to some error such that,

$$\mathbf{H} = \hat{\mathbf{H}} + \mathbf{U}. \tag{4.11}$$

where $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2, \ldots, \hat{\mathbf{h}}_{N_r}]^T \in \mathbb{C}^{N_r\times N_t}$ is the nominal channel matrix that is available at the secondary user and $\hat{\mathbf{U}}_m = [\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \ldots, \hat{\mathbf{u}}_{N_r}] \in \mathbb{C}^{N_r\times N_t}$ is the related uncertainty matrix where $\mathbf{u}_k \quad \forall k \in [1, \ldots, N_r]^T$. Due to the channel uncertainty, the observation signal after plugging eq. (4.11) into eq. (4.1) is reduced to,

$$\mathbf{x}_k = (\widehat{\mathbf{h}}_k + \mathbf{u}_k)\mathbf{s} + \eta_k. \tag{4.12}$$

The objective of the fusion center is to choose the correct model from the following two possible hypotheses:

$$\begin{aligned} \mathbf{x}_k/H_0 &\sim \mathcal{CN}(\mu_0, \mathbf{\Gamma}_0) \\ \mathbf{x}_k/H_1 &\sim \mathcal{CN}(\mu_1, \mathbf{\Gamma}_1) \end{aligned} \tag{4.13}$$

The mean and the covariance matrix of the observation for each i^{th} receiving antenna at alternative hypothesis can be driven as follows:

$$
\begin{aligned}
\mu_1(k) &= E[\mathbf{x}/H_1] = E[\mathbf{Hs} + \eta_k] \\
&= E[\mathbf{Hs}] + E[\eta_k] = \mathbf{H}E[\mathbf{s}] = 0 \\
\Gamma_1(k) &= E[\mathbf{x}\mathbf{x}^H] = E[(\mathbf{Hs} + \eta_k)(\mathbf{Hs} + \eta)^H] \\
&= E[\mathbf{H}\mathbf{s}\mathbf{s}^H\mathbf{H}^H] + E[\eta_k\eta_k^H] + E[\mathbf{Hs}\eta_k^H] + E[\eta_k\mathbf{H}^H\mathbf{s}^H] \\
&= \mathbf{H}\mathbf{H}^H + \Gamma_0.
\end{aligned}
$$

where $\mathbf{\Gamma}_1$ is equal to $\mathbf{H}\mathbf{H}^H + \mathbf{R}\eta \in \mathbb{C}^{N_r \times N_r}$ and $\mathbf{\Gamma}_0$ is equal to $\mathbf{R}_\eta \in \mathbb{C}^{N_r \times N_r}$. The likelihood of the observation X_k corresponding to alternative hypothesis $\mathcal{H}_0$ is given by:

$$
\mathbf{p}(\mathrm{x}_k, \mathcal{H}_0) = \frac{exp(-\mathrm{x}_k\mathbf{\Gamma}_0\mathrm{x}_k^H)}{\pi^{L+N_rN_t}\,|\mathbf{\Gamma}_0|} \tag{4.14}
$$

while the likelihood of the observation X_k corresponding to alternative hypothesis $\mathcal{H}_1$ is given by:

$$
\mathbf{p}(\mathrm{x}_k, \mathcal{H}_1) = \frac{exp(-\mathrm{x}_k\mathbf{\Gamma}_1\mathrm{x}_k^H)}{\pi^{L+N_rN_t}\,|\mathbf{\Gamma}_1|} \tag{4.15}
$$

4.5 Methodology

4.5.1 B-GLRT detector with known channel gains and unknown noise covariance matrix (B-GLRT1)

In this part, we assume the channel uncertainty is known while the noise covariance matrix is unknown. Prior information for the distribution of the noise variance is assumed as known at SU. We assume that the noise variance has an Inverse-Gamma distribution with different values of shape ρ and scale parameter κ . Thus, $\gamma^2 \sim$ Inv-Gamma (ρ, κ) [144, 42, 112]. In the next section, we also assume that the noise has the prior statistical distribution for each noise variance. In other words, assuming a diagonal matrix $\mathbf{\Gamma}_0$ =

$diag\left(\gamma_{0,1}^2,\ldots,\gamma_{0,N_r}^2\right)$, of the noise covariance, $\gamma_{0,i}^2$ is clearly modeled *a priori* by an Inverse-Gamma distribution with the shape parameter ρ_i and the scale parameter κ_i is considered an Inv-Gamma(ρ_i,κ_i).

Now, Bayesian approach can be utilized in this case to determine the model unknown parameters where the observation sample is finite. Here the channel gain matrix has been assumed to be known and the noise variance is unknown for the SU. To achieve more flexible calculation, we used the joint prior for the noise variance parameters as given by

$$\begin{aligned} f\left(\gamma_{0,i}^2\right) &= f(\mathbf{\Gamma}_0) \\ &= \prod_{i=1}^{N_r} \frac{\kappa_i^{\rho_i}}{\Gamma\left(\rho_i\right)} \left(\gamma_{0,i}^2\right)^{-\rho_i-1} \exp\left(\frac{-\kappa_i}{\gamma_{0,i}^2}\right). \end{aligned} \tag{4.16}$$

where $i = [1,\ldots,N_r]$, In order to obtain the values of unknown parameters in null hypothesis, and since these parameters have prior knowledge based on the Bayesian theorem, the maximum *a posterior* probability can be involved in this case. This can be achieved by taking the log of eq. (4.8) followed by taking the maximum as shown in eq. (4.17) to achieve the MAP estimator. We require the MAP estimator of the noise variance parameters under H_0, that is,

$$\hat{\Theta}_0|\mathbf{x}, H_0 = \text{Max}\left\{\log\left(f_{H_0}\left(\mathbf{x}|\Theta_0\right) f\left(\Theta_0\right)\right)\right\} \tag{4.17}$$

Then the posterior distribution of noise variance can be shown as:

$$\hat{\mathbf{\Gamma}}_0 = \underset{\mathbf{\Gamma}_0}{\text{argmax}}\left[\log(f\left(\mathbf{X}; H_0, \mathbf{\Gamma}_0\right)) + \log(f\left(\mathbf{\Gamma}_0\right))\right]. \tag{4.18}$$

In null hypothesis, $\mathbf{\Gamma}_0 = \mathbf{R}_\eta$. Thus, eq. (4.18) become eq. (4.19),

$$\hat{\mathbf{R}}_\eta = \underset{\mathbf{R}_\eta}{\text{argmax}}\left[\log(f\left(\mathbf{X}; H_0, \mathbf{R}_\eta\right)) + \log(f\left(\mathbf{R}_\eta\right))\right]. \tag{4.19}$$

It can be seen from the system model that the observation signal at i^{th} antenna at secondary side is complex normal distribution with density:

$$\begin{aligned} f_{H_0}(\mathbf{X}|\mathbf{\Gamma}_0) &= \prod_{j=1}^{L} f_{H_0}(\mathbf{x}_j|\mathbf{R}_\eta) \\ &= \pi^{-N_r L}\left(\prod_{i=1}^{N_r}\gamma_{0,i}^2\right)^{-L}\exp\left\{-\sum_{i=1}^{N_r}\sum_{j=1}^{L}\frac{x_{ij}^* x_{ij}}{\gamma_{0,i}^2}\right\} \\ &= \pi^{-N_r L}\prod_{i=1}^{N_r}\left[(\gamma_{0,i}^2)^{-L}\exp\left\{\frac{-1}{\gamma_{0,i}^2}\sum_{j=1}^{L}x_{ij}^* x_{ij}\right\}\right]. \end{aligned} \tag{4.20}$$

According to eq. (4.18), the log function can be taken for multiplication between eq. (4.20) and eq. (4.16) followed by taking the derivative with respect to $\gamma_{0,i}^2 > 0$, $i = [1, \ldots, N_r]$, in order to obtain the model parameter. It can be easily seen as:

$$\gamma_{0,i}^2 = \frac{1}{L + \rho_i + 1}\left(\sum_{j=1}^{L} x_{ij}^* x_{ij} + \kappa_i\right). \tag{4.21}$$

for $i = [1, \ldots, N_r]$, (see Appendix 1 in our previous work [42] for more detail about the proof).

Under the alternative hypothesis, we obtain the MAP estimation of noise covariance matrix in hypothesis detection problem where the channel gain and its error are known. Based on eq. (4.17) the MAP estimation equation will be:

$$\hat{\Theta}_1 = \underset{\Theta_1}{\operatorname{argmax}} f(\mathbf{x}; H_1, \Theta_1) f(\Theta_1). \tag{4.22}$$

In alternative hypothesis, $\mathbf{\Gamma}_1 = \mathbf{H}\mathbf{H}^H + \mathbf{\Gamma}_0$ represents the covariance matrix of the observation where the channel has known uncertainty as shown in eq. (4.11). $\mathbf{\Gamma}_0$ is also modeled as *a priori* using eq. (4.16). The density of $\mathbf{x}$ in alternative hypothesis is given by:

$$f_{H_1}(\mathbf{x}; \mathbf{\Gamma}_0) = \pi^{-N_r}\,|\mathbf{\Gamma_1}|^{-1}\exp\left\{-\mathbf{x}^H \mathbf{\Gamma}_1^{-1}\mathbf{x}\right\}$$

where $\Gamma_1 = (\mathbf{H}\mathbf{H}^H + \mathbf{\Gamma}_0)^{-1}$, so that the data distribution under H_1 has density:

$$\begin{aligned} f_{H_1}(\mathbf{X}|\mathbf{R}_\eta) &= \pi^{-N_r} \left|\mathbf{H}\mathbf{H}^H + \mathbf{R}_\eta\right|^{-L} \\ &\quad \times \exp\left\{-\sum_{j=1}^{L} \mathbf{x}_j^H (\mathbf{H}\mathbf{H}^H + \mathbf{R}_\eta)^{-1}\mathbf{x}_j\right\} \\ &= \pi^{-N_r} \left|\mathbf{H}\mathbf{H}^H + \mathbf{R}_\eta\right|^{-L} etr\left\{-\mathbf{R}(\mathbf{H}\mathbf{H}^{\mathbf{H}} + \mathbf{R}_\eta)^{-\mathbf{1}}\right\}. \end{aligned} \tag{4.23}$$

Let $\mathbf{R}_h = \mathbf{H}\mathbf{H}^H$ be the covariance channel matrix that is known for the SU. We also define $\mathbf{R} = \sum_{j=1}^{L} \mathbf{x}_j \mathbf{x}_j^H$. Thus, the MAP estimator of the noise variance under H_1 is given by solving the following optimization problem:

$$P1: \quad \underset{\mathbf{\Gamma}_1}{\operatorname{argmax}} \left[\log(f_{H_1}(\mathbf{X}|\mathbf{\Gamma}_1)) + \log(f(\mathbf{\Gamma}_0))\right]. \tag{4.24}$$

By substituting eq. (4.16) and eq. (4.23) into eq. (4.24), P1 becomes:

$$\begin{aligned} P1: \quad & \underset{\mathbf{\Gamma}_0}{\operatorname{argmax}} \left[\log f_{H_1}(\mathbf{X}|\mathbf{\Gamma}_1) + \log f(\mathbf{\Gamma}_0)\right] \\ & = \underset{\Gamma_0:diagonal}{\operatorname{argmax}} \left(L \log\left|\mathbf{\Gamma}_1^{-1}\right| - tr[\mathbf{R}\mathbf{\Gamma}_1^{-1}]\right. \\ & \quad \left. + (\rho + 1)\log\left|\mathbf{\Gamma}_0^{-1}\right| - tr[\mathbf{K}\Gamma_0^{-1}]\right). \end{aligned} \tag{4.25}$$

where $(\rho_1 = \cdots = \rho_{N_r} = \rho)$ are the hyperparameters of the conjugate priors and $\mathbf{K} = diag(\kappa_1, \ldots, \kappa_{N_r})$ is the diagonal matrix which represents the conjugate priors hyperparameters. Since problem P1 is not convex, we must manipulated it to be a convex optimization problem which can be easily solved by using a numerical semidefinite programming

(SDP) optimization package such as [51]. For more detail, see Appendix (1.1).

4.5.2 B-GLRT detector with unknown channel uncertainty and noise covariance matrix

In this part, we assume the channel uncertainty matrix is unknown and the noise covariance matrix as well at SU. We also use the same prior distribution for the noise variance as mentioned in eq. (4.16). According to the system model as shown in eq. (4.1), the estimated valued of noise covariance matrix in null hypothesis still can be obtained by using eq. (4.21) while in alternative hypothesis both channel uncertainty and noise covariance matrix must be estimated. Thus, the problem can be seen as

$$\begin{aligned} P1: \quad & \underset{\substack{\mathbf{\Gamma}_0:Digonal \\ \mathbf{\Gamma}_1 \geq 0}}{\operatorname{argmax}} \left[\log f_{H_1}\left(\mathbf{X}|\mathbf{\Gamma}_1\right) + \log f\left(\mathbf{\Gamma}_0\right)\right] \\ & = \underset{\mathbf{\Gamma}_0:diagonal}{\operatorname{argmax}} \left(L \log \left|\mathbf{\Gamma}_1^{-1}\right| - tr[\mathbf{R}\mathbf{\Gamma}_1^{-1}]\right. \\ & \left. + (\rho+1)\log\left|\mathbf{\Gamma}_0^{-1}\right| - tr[\mathbf{K}\mathbf{\Gamma}_0^{-1}]\right). \end{aligned} \tag{4.26}$$

Then the objective is to maximize the LR subject to certain constraint, so *P1* becomes.

$$\begin{aligned} P1 \quad : \quad & \underset{\mathbf{\Gamma}_0:Diagonal}{\operatorname{argmax}} \left(L \log \left|(\widehat{\mathbf{R}}_h + \mathbf{\Delta R}_h) + \mathbf{\Gamma}_0\right|^{-1}\right. \\ & -tr[\mathbf{R}\left((\widehat{\mathbf{R}}_h + \mathbf{\Delta R}_h) + \mathbf{\Gamma}_0\right)_1^{-1}] \\ & \left. +(\rho+1)\log\left|\mathbf{\Gamma}_0^{-1}\right| - tr[\mathbf{K}\mathbf{\Gamma}_0^{-1}]\right). \end{aligned} \tag{4.27}$$

Here, we must maximize $f_{H_1}(\mathbf{x}|\mathbf{\Gamma}_0)$ because $f_{H_1}(\mathbf{x}|\mathbf{\Gamma}_0)$ is a function of $\mathbf{\Delta R}_h$ which is the channel error and has an effect on the maximization of LR. To find the value of $\mathbf{\Delta R}_h$ that maximizes $f_{H_1}(\mathbf{x}|\mathbf{\Gamma}_0)$, we can perform the following solutions.

Robust solution, B-GLT2

Back to eq. (4.27), we can see that this problem consists of two optimization variables ($\Delta\mathbf{R}_h$, $\mathbf{\Gamma}_0$) and and since it is difficult to handle as a single optimization problem, it can be solved using alternative optimization technique as following similar procedure in our previous work[30] and also as following in this paper [145]. The optimization problem becomes:

- Solving for $\mathbf{\Gamma}_0$: in this case, we assume that $\mathbf{\Gamma}_0$ is unknown and $\Delta\mathbf{R}_h$ is known and then solve for $\mathbf{\Gamma}_0$.

$$\begin{aligned} P1 \quad : \quad \underset{\Gamma_0}{\operatorname{argmax}} (L \log | \underbrace{(\widehat{\mathbf{R}}_h + \Delta\mathbf{R}_h)}_{\mathbf{R}_h} + \mathbf{\Gamma}_0|^{-1} \\ -tr[\mathbf{R}(\underbrace{(\widehat{\mathbf{R}}_h + \Delta\mathbf{R}_h)}_{\mathbf{R}_h} + \mathbf{\Gamma}_0)^{-1}] \\ +(\rho+1)\log\left|\mathbf{\Gamma}_0^{-1}\right| - tr[\mathbf{K}\mathbf{\Gamma}_0^{-1}]). \end{aligned}$$

  subject to

$$\mathbf{\Gamma_0} \geq 0,$$

  For additional details see Appendix (2.1) .

- Solving for $\Delta\mathbf{R}_h$: now we can assume that $\mathbf{\Gamma}_0$ is known and solve for $\Delta\mathbf{R}_h$. We also define that,

$$\mathfrak{R} = \widehat{\mathbf{R}}_h + \mathbf{\Gamma}_0$$

$$
\begin{aligned}
P1 \quad : \quad & \underset{\Delta \mathbf{R}_h}{\operatorname{argmax}} (L \log |(\mathfrak{R} + \Delta \mathbf{R}_h)|^{-1} \\
& -tr[\mathbf{R} \, (\mathfrak{R} + \Delta \mathbf{R}_h)^{-1}] \\
& +(\rho + 1) \log \left|\mathbf{\Gamma}_0^{-1}\right| - tr[\mathbf{K}\mathbf{\Gamma}_0^{-1}]).
\end{aligned}
$$

subject to

$$
\|\Delta \mathbf{R}_h\|_F \leq \epsilon,
$$

Appendix (2.2) shows mote detail.

Sub-optimal solution, B-GLT3

In this solution we substitute the value of $\Delta \mathbf{R}_h$ that gives the minimum value for $f_{H_1}(\mathbf{x}|\mathbf{\Gamma}_0)$, where $\Delta \mathbf{R}_h$ equal to $-\epsilon \mathbf{I}_M$ or $+\epsilon \mathbf{I}_M$ depending on the position of $\Delta \mathbf{R}_h$ in $f_{H_1}(\mathbf{x}|\mathbf{\Gamma}_0)$ or in case the values of $\Delta \mathbf{R}_h$ in $f_{H_1}(\mathbf{x}|\mathbf{\Gamma}_0)$ that lead to minimum value of $f_{H_1}(\mathbf{x}|\mathbf{\Gamma}_0)$ are both $+\epsilon \mathbf{I}_M$. Thus,

$$
f_{H_1}(\mathbf{x}|\mathbf{\Gamma}_0) = \pi^{-N_r} \left|(\widehat{\mathbf{R}}_h + \epsilon \mathbf{I}_M)^{+} \mathbf{\Gamma}_0\right|^{-L} etr \left\{-\mathbf{R}(\widehat{\mathbf{R}}_{\mathbf{h}} + \epsilon \mathbf{I}_{\mathbf{M}})^{+} + \mathbf{\Gamma}_{\mathbf{0}})^{-1}\right\}, \tag{4.28}
$$

$$
\begin{aligned}
P1 \quad : \quad & \underset{\Gamma_0 : Diagonal}{\operatorname{argmax}} \, (L \log \left|(\widehat{\mathbf{R}}_h + \epsilon \mathbf{I}_M) + \mathbf{\Gamma}_0\right|^{-1} \\
& -tr[\mathbf{R} \left((\widehat{\mathbf{R}}_h + \epsilon \mathbf{I}_M) + \mathbf{\Gamma}_0\right)_1^{-1}] \\
& +(\rho + 1) \log \left|\mathbf{\Gamma}_0^{-1}\right| - tr[\mathbf{K}\mathbf{\Gamma}_0^{-1}]).
\end{aligned} \tag{4.29}
$$

The positive exponent means the bonded matrix must be positive semi definite (PSD). This can be accomplished by setting the negative Eigenvalues of the resultant $\widehat{\mathbf{R}}_h + \epsilon \mathbf{I}_M$ to

zero and this happens due to the Hermitian of $\widehat{\mathbf{R}}_h$. In this case, the problem can be solved based on Appendix (1.2).

After estimating the model parameters in both hypothesis Γ_0 and Γ_1, we can derive the detector formula. According to eq. (4.6), therefore, the log of this equation is:

$$\begin{aligned} \text{LLR} = &\log(f_{H_1}(\mathbf{x}|\boldsymbol{\Gamma}_1) f(\boldsymbol{\Gamma}_1)) \\ &- \log(f_{H_0}(\mathbf{x}|\boldsymbol{\Gamma}_0) f(\boldsymbol{\Gamma}_0)). \end{aligned} \tag{4.30}$$

Then, the log of the posterior distribution can be:

$$\begin{aligned} \log(.)|H_0 = &-(L+\rho+1)\log(\left|\hat{\boldsymbol{\Gamma}}_0\right|) \\ &+ tr(-\mathbf{R}\hat{\boldsymbol{\Gamma}}_0^{-1}) - tr(\mathbf{K}\hat{\boldsymbol{\Gamma}}_0^{-1}). \end{aligned} \tag{4.31}$$

and

$$\begin{aligned} \log(.)|H_1 = &-(L+\rho+1)\log(\left|\hat{\boldsymbol{\Gamma}}_1\right|) \\ &+ tr(-\mathbf{R}\hat{\boldsymbol{\Gamma}}_1^{-1}) - tr(\mathbf{K}\hat{\boldsymbol{\Gamma}}_1^{-1}). \end{aligned} \tag{4.32}$$

Substitute eq. (4.32) and eq. (4.31) in eq. (4.30). Therefore, the detector is given by;

$$\begin{aligned} \mathrm{T}_{\zeta}(\mathrm{B-GLRT}) = &(L+\rho+1)\log(\frac{\left|\hat{\boldsymbol{\Gamma}}_1\right|}{\left|\hat{\boldsymbol{\Gamma}}_0\right|}) + tr(-\mathbf{R}\hat{\boldsymbol{\Gamma}}_1^{-1}) \\ &- tr(-\mathbf{R}\hat{\boldsymbol{\Gamma}}_0^{-1}) - tr(\mathbf{K}\hat{\boldsymbol{\Gamma}}_1^{-1}) + tr(\mathbf{K}\hat{\boldsymbol{\Gamma}}_0^{-1}) \underset{H_0}{\overset{H_1}{\gtrless}} \zeta \end{aligned} \tag{4.33}$$

4.6 Numerical Evaluation

In this part, we demonstrate our approach using simulated data, and clarify the performance of our design of the GLRT spectrum sensing algorithm when the channel uncertainty exists over varied parameter settings for the distribution model. In our simulation, the average received SNR at SU has been defined as [42]:

$$SNR = \log(tr(\mathbf{H}\mathbf{H}^H)) - \log(tr(\mathbf{\Gamma}_0)). \tag{4.34}$$

First, we clarify the effect of the channel uncertainty $\mathbf{U}$ in spectrum sensing where the noise parameters are unknown and uncorrelated. As mentioned, the channel uncertainty is assumed to have deterministic values while prior knowledge is known for the noise parameters. For these assumptions, we show the performance of the GLRT algorithm based on Bayesian approach as known as B-GLR1, and also we show its performance compared with other approaches which are derived in this work. These approaches also compared with other state-of-the-art spectrum sensing under the same assumptions, such as the ED and others.

As mentioned, MIMO technique is considered in this paper with the number of receiving antennas N_r equal to 3 to receive a number of samples L equal to 20, the number of transmitting antennas N_t= 2.

We also set the hyperparameter values of prior distributions parameter as the shape $\left(\rho_1 = \ldots = \rho_{N_r}\right)$ equal to 2, and the scale $\mathbf{K}$ has different arbitrary values as diag [2, 1.5, 3]. The number of realizations is $5 * 10^4$ in which probability of false alarm (pfa) can be generated to eventually obtain the threshold value.

In Fig. 4.2, we illustrate the probability of missed detection (Pm) in terms of SNR for the several detectors such as the maximum to minimum eigenvalue test (MME), Energy with minimum Eigenvalue test (EME) [146, 147] as shown in eq. (4.35) and eq. (4.36) respectively,

$$\mathrm{T}_{\zeta}(\mathrm{MME}) = \frac{\lambda_{max}(\hat{\Gamma})}{\lambda_{min}(\hat{\Gamma})} \underset{H_0}{\overset{H_1}{\gtrless}} \zeta \tag{4.35}$$

$$\mathrm{T}_{\zeta}(\mathrm{EME}) = \frac{\lambda_{av}(\hat{\Gamma})}{\lambda_{min}(\hat{\Gamma})} \underset{H_0}{\overset{H_1}{\gtrless}} \zeta \tag{4.36}$$

where λ_{max}, λ_{min}, and λ_{av} respectively denote the maximum, minimum, average eigenvalues of the covariance matrix ($\hat{\Gamma}$).

This figure also depicts the performance of all detectors when the SNR is within the range [-20 to 10] at a pdf = 5×10^{-1}, with the finite sample number L equal to 20 . It can also be seen that B-GLRT 1 outperforms other proposed detectors since it has a perfect channel compared with other cases.

Respective operating characteristic curves (ROC) can be seen in Fig. 4.3. In this figure, we shows the probability of detection (PD) versus the probability of false alarm using a range of values of pfa, SNR value is -3 dB and number of samples L equal to 10 . It also can be seen an efficient detection performance of the proposed detectors compared with other state-of-the-art detectors such as the Hadamard ratio test and the Sphericity test [105] shown in eq. (4.37) and eq. (4.38) respectively. The threshold of the case is numerically obtained. In addition, the results show that the optimal detector largely outperforms Hadamard ratio test and Sphericity test and we see that the proposed B-GLRT1 detector has good performance compared with other proposed detectors.

$$\mathrm{T}_{\zeta}(\mathrm{HR}) = \left(\prod_{j=1}^{N_r} \gamma_{jj} \right) |\Gamma|^{-1} \underset{H_0}{\overset{H_1}{\gtrless}} \zeta \tag{4.37}$$

$$\mathrm{T}_{\zeta}(\mathrm{ST}) = \left(\frac{\sum_{j=1}^{N_r} \gamma_{jj}}{N_r} \right) |\Gamma|^{-(1/N_r)} \underset{H_0}{\overset{H_1}{\gtrless}} \zeta \tag{4.38}$$

where Γ is the covariance matrix.

Furthermore, we also show the performance of proposed detectors in terms of the

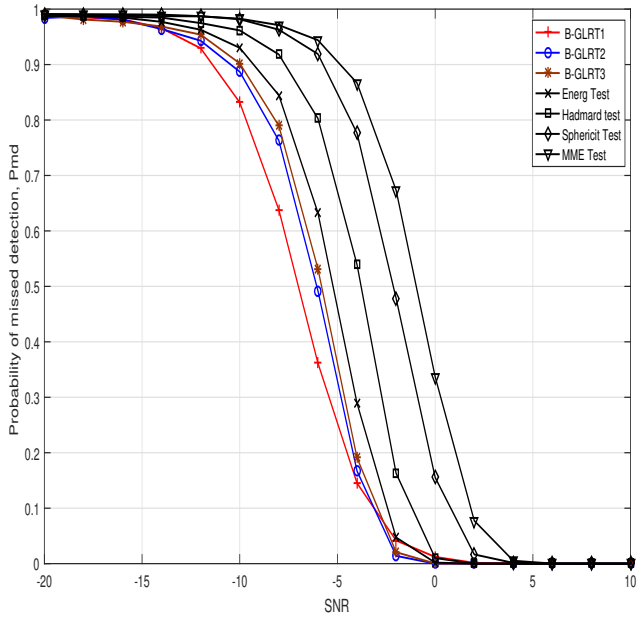

Figure 4.2: Probability of missed detection of different detectors against average SNR for $Pfa = 5 \times 10^{-1}$, $N_r = 3$ and $L = 20$.

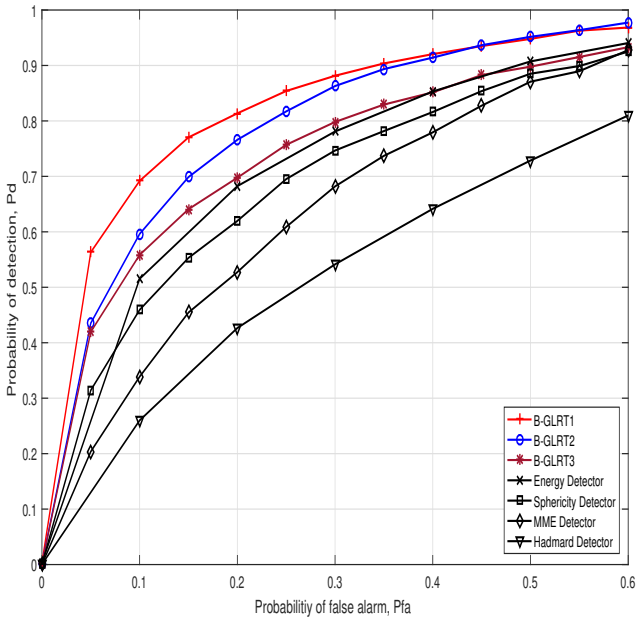

Figure 4.3: Simulated ROC performance of the proposed B-GLRT assumed prior at SNR = -3 dB.,$N_r = 3$ and $L = 10$.

number signal samples. Fig. 4.4, the probability of missed detection (Pm) for each detector can be observed versus the number of samples L . In Fig. 4.4, it can be shown that the proposed detectors have a significantly better performance compared with other detectors where pfa is equal 5×10^{-1}, N_r = 3, N_t = 2 and the average SNR of -3 dB. It can be observed that the proposed detector performance has an extreme increase when the input samples increase. In addition, the proposed B-GLRT1 detector performs better than other detectors since it has a perfect uncertainty.

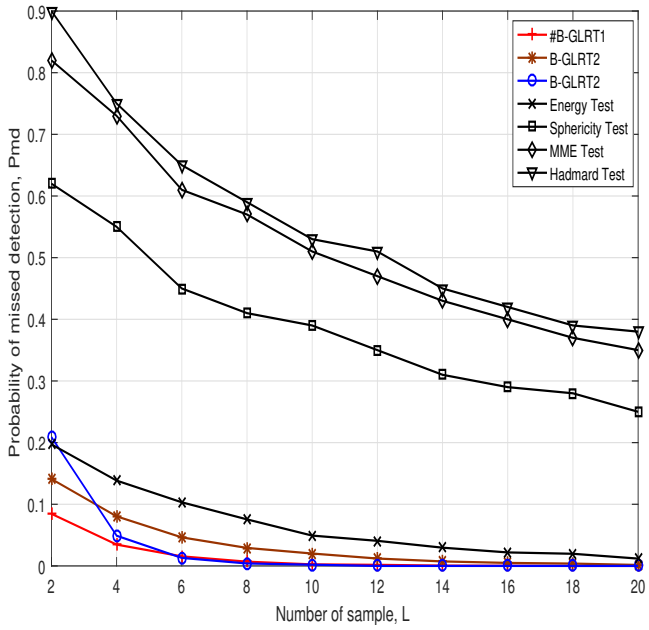

Figure 4.4: Probability of missed detection of the proposed detectors against L for pfa = 5×10^{-1}, average SNR = -3 dB.

We also do another simulation in order to show the performance of the detectors where the signal samples is a large number. In this section, Fig.4.5 clarifies the relation between the probability of detection versus the probability of alarm (pfa). From this figure, it can generally be observed that even the performance of most detectors have a significant improvement but the proposed detectors maintain a better performance compared with other detectors or the propose detectors 'B-GLRT2', 'B-GLRT3' have enough ability to be robust detectors against channel uncertainty. Fig.4.6 shows the performance of the

proposed approach agianst other detectors where the signal samples is a large value.

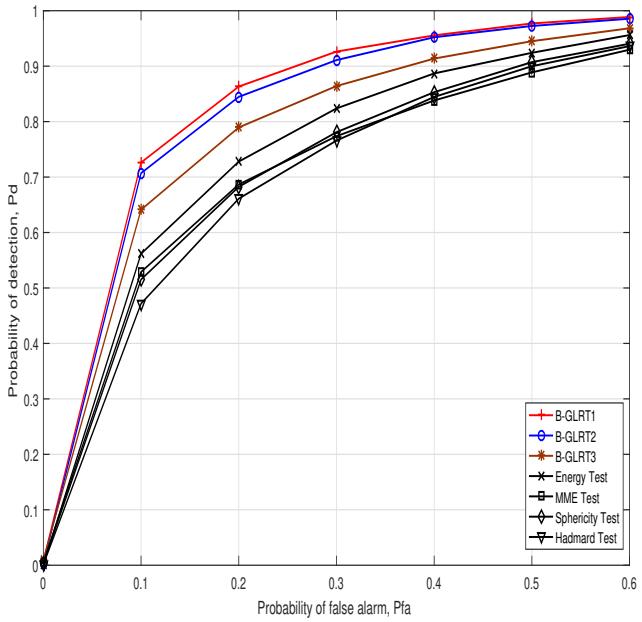

Figure 4.5: Probability of detection versus probability of false alarm at SNR = -3 dB., N_r = 3, N_t = 2 and L =20.

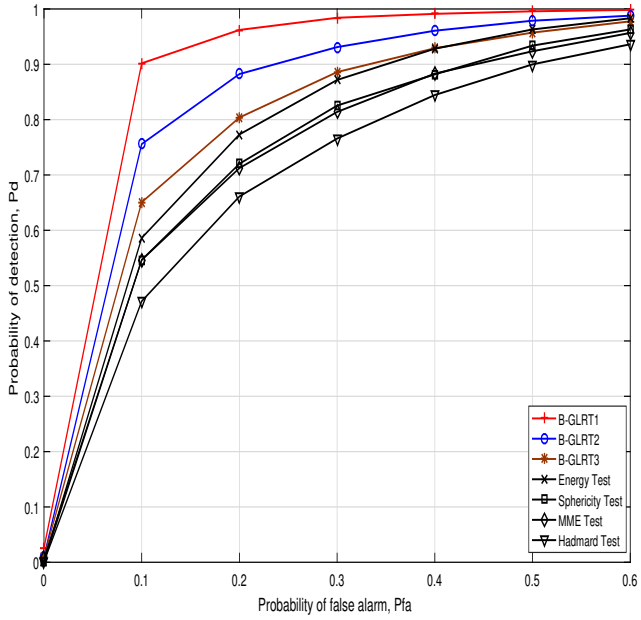

Figure 4.6: Probability of detection against probability of false alarm at SNR = -3 dB., N_r = 5, N_t = 3 and L = 25.

Finally, we show the probability of missed detection of the detectors in terms of SNR where the number of samples is changed as shown in Fig.4.7. From this figure, it can be seen that the proposed detectors have an efficient performance in terms of probability of missed detection when the number of signal samples is increased.

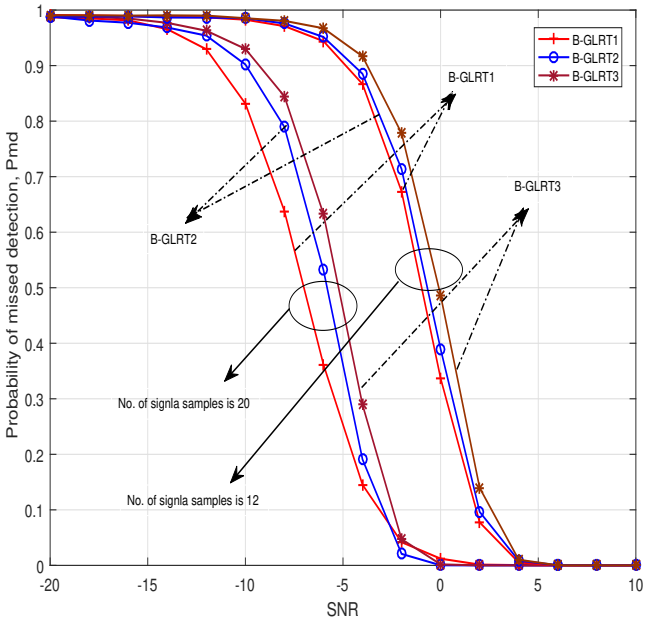

Figure 4.7: Probability of missed detection of different detectors against average SNR for $Pfa = 5 \times 10^{-1}$, $N_r = $, $N_t = 2$.

4.7 Conclusion

This paper presents a novel method for developing spectrum sensing detectors in multi-cell CR network with non-perfect CSI for the SU network. In this context, novel detection schemes such as the B-GLRT1, B-GLRT2, and B-GLRT3 are proposed for different cases when the noise distribution parameters are unknown in the non-perfect channel. For the first case with unknown noise covariance and the perfect channel, we derived a GLRT1 based binary hypothesis detector for spectrum sensing in MIMO CR networks. In this case, our formulation for maximizing likelihood distribution determines the correspond-

ing unknown parameters for each hypothesis within a specified amount of SNR from the SU base stations. We further developed an expression of B-GLRT2 and B-GLRT3, where the CSI uncertainty in MIMO cognitive radio is not perfect. This problem is formulated as an optimization problem that is solved in different ways to obtain a robust and sub-optimal solution for the spectrum sensing CR. Simulation results are presented to demonstrate the proposed detectors' performance where a CSI and noise uncertainty is available with a finite number of observed data samples. The results also illustrate that the proposed B-GLRT detectors achieve significant improvement compared with state-of-the-art spectrum sensing schemes. We present simulation results showing that our solution methods outperform state-of-the-art methods with existing and non-existing channel uncertainty. Further, the theoretical analysis and simulation re- sults indicate that our approaches can offer outstanding spectrum sensing performance when small sample size is available.This work focused on obtaining a robust signal detector at the secondary user based on local decisions received from the cooperating signal PU considering CSI uncertainty in MIMO CR Networks. A promising extension to future work can use the cooperative spectrum sensing at the fusion center side in the presence of the noise and channel uncertainty.

Chapter 5

Concluding remarks and Future Work

This chapter summarizes our developed spectrum sensing techniques for both the noise and the channel uncertainty- in cognitive wireless networks. It also explains some future intended works related to these networks using both Bayesian and convex optimization based on the statistical analysis methods.

5.1 Research Summary

Chapter 3 considered the spectrum sensing problem between the end-users in the conventional wireless network when the links between the primary user and the secondary users are uncertain. Two solutions were developed using the MAP estimation method based on the Bayesian theorem. The B-GLRT solution following our proposed robust solution in several past works. Compared with another art-of-states detector with the presence of the noise uncertainty.

We have presented a novel multiantenna detector for Gaussian distribution signals spatially (with different variances) and uncorrelated noises. In particular, we intend to use the B-GLRT for detecting an observed vector values random variable in presence of

the iid noises, Although this detector does not admit, in general, robust solution, we have shown that in the low SNR our proposed algorithm outperforms other detectors.

Finally, we note that the derivation considered here assumes knowledge of the channel uncertainty. While this may be reasonable in some contexts, there are scenarios in which the channel can be unknown, which can be studies and showed in chapter four.

Chapter 4 considered the spectrum sensing cognitive radio among the secondary users in the conventional wireless network when the links are also assumed to are uncertain. In this work, two cases are studied where both the channel and the noise uncertain available. Two solutions were developed using the MAP estimation method based on the Bayesian theorem in null hypothesis while the convex optimization method is utilized when the channel uncertainty is unknown. The traditional solution following our proposed robust solution is shown in the past chapter. This formation problem is compared with another art-of-states detector with the presence of the channel and the noise uncertainty.

It can be seen from the previous chapters that we have presented a novel multiantenna detector for Gaussian distribution signals spatially (with different variances) where the noises and the channel uncertainty are excited. In particular, we intend to use the B-GLRT for detecting an observed data random variable in presence of the iid noises, Although this detector does not admit, in general, robust B-GLRT, we have shown those proposed methods .

Finally, we perceive that the derivation considered here assumes knowledge of the channel uncertainty. While this may be reasonable in some contexts, there are scenarios in which the channel can be unknown, which can be studied and showed in chapter four. The two solutions were analyzed through numerical simulation, and the result shows the validation and convergence of the clarifications for different network parameters. Also, the result shows that the proposed solution outperforms the sub-optimal one.

5.2 Future Work

We are intending to continue reaching towards the next directions:

5.2.1 Spectrum Sensing mmwave Backscattering Radio

The cognitive radio (CR) technique has been extensively studied in recent decades with the goal of improving spectrum efficiency. This will be accomplished, as we discussed in previous chapters, by allowing secondary users access to the designated radio spectrum. In a CR system, the secondary user is allowed to use the radio spectrum assigned to the PU in an opportunistic or spectrum sharing approach. In other words, the second receiver (SRx) can share the secondary transmission from the second transmitter (STx) to the same spectrum with the primary transmission from the primary transmitter (PTx) to the primary receiver (PRx). The process is carried out by detecting the interference level from STx to PRx in order to ensure that the caused interference level to the primary transmission is below a set of common threshold values. The development of spectrum efficiency may be limited by transmission between end-users. As a result, the typical transmitter architecture demands the use of energy active components, such as active radio.

Backscattering radio technology, such as ambient backscatter communications (AmBC), has recently been used to develop wireless transmitters that do not require active components, resulting in a significant reduction in signal power consumption. Furthermore, STx uses backscattering modulation to send its message across the ambient radio frequency (RF) waves. We can use the spectrum sensing technique to estimate the signal between the two independent backscattering processes in this new technology that works with enhanced wireless communication like 5G and 6G. We may also investigate network uncertainty, such as noise and channel uncertainty, using this method. We intend to use a variety of prior models based on Bayesian theorems, such as the Dirichlet-Multinomial model, which may be found in Appendix A.

5.2.2 Massive MIMO Cognitive Backscattering Communications for Future Wireless Networks, Symbiotic Radio:

In this technique, more than one Users can share the same spectrum sensing signal utilizing a special technique. In subse-quent work, both spectrum sensing models (in perfect CSI and non-perfect CSI) shall be considered and studied deeply. In a summary, a new approach called symbiotic radio (SR) is being investigated as a way to combine the benefits of CR and AmBC while addressing the drawbacks of both. SR, like CR, has two spectrum sharing systems: the primary system and the secondary system. SR, but at the other side, creates mutually beneficial spectrum sharing rather than blocking it in CR. In comparison to AmBC, the SR approach uses collaborative decoding to provide very dependable backscattering communications. As a result, SR is also known as cognitive backscattering communications, which improves wireless network spectrum and energy efficiency.

5.2.3 Non Orthogonal Multiple Access(NOMA)

In this work, we shall investigate a cooperative non-orthogonal multiple access (CNOMA) scheme based on spectrum sensing where perfect CSI and non-perfect CSI exists, not only the channel uncertainty can be addressed in this 5G technique, but also the noise uncertainty can be addressed, this problem can be formulated in multiple methods and solved in multiple approaches. the statistical approach can be involved in this problem.

Appendix A

Dirichlet-Multinomial

A.1 Dirichlet-Multinomial Model

$$p_1,\ p_K \sim \text{Dir}(\alpha_1,\ \ldots,\ \alpha_K)$$

$y_1, \ldots y_K$ ~Mult $(p_1,\ \ldots, p_K)$

This class is essentially just the exponential family of distributions.

Posterior.

$$f(\theta|D) \propto f(\theta,\ D)$$

$$= f(p_1,\ \ldots, p_K|\alpha_1,\ .\ \alpha_K) \prod_{y_i \in D} f(y_i|p_1,\ \ldots p_K)$$

$$\propto \prod_{j=1}^{K} p_j^{\alpha_j - 1} \prod_{y_i \in D} \prod_{j=1}^{K} p_j^{y_i^{(j)}}$$

$$= \prod_{j=1}^{K} p_j^{\alpha_j - 1 + \Sigma_{y_i \in D} y_i^{(j)}}$$

This density is exactly that of a Dirichlet distribution, except we have

$$\alpha'_j = \alpha_j + \sum_{y_i \in D} y_i^{(j)}$$

That is, $f(\theta|D) = \text{Dir}(\alpha'_1,\ \alpha'_K)$.

Posterior Predictive.

$$f(y|D) = \int f(y|\theta) f(\theta|D) d\theta$$

$$= \int f(y|p_1,\ p_K) f(p_1,\ p_K|D) dS_K$$

$$= \int \frac{\Gamma(n+1)}{\prod_{j=1}^K \Gamma(y^{(j)}+1)} \prod_{j=1}^K p_j^{y^{(j)}} \frac{\Gamma(\sum_{j=1}^K \alpha'_j)}{\prod_{j=1}^K \Gamma(\alpha'_j)} \prod_{j=1}^K p_j^{\alpha'_j - 1} dS_K$$

$$= \frac{\Gamma(n+1)}{\prod_{j=1}^K \Gamma(y^{(j)}+1)} \frac{\Gamma(\sum_{j=1}^K \alpha'_j)}{\prod_{j=1}^K \Gamma(\alpha'_j)} \int \prod_{j=1}^K p_j^{y^{(j)} + \alpha'_j - 1} dS_K$$

$$= \frac{\Gamma(n+1)}{\prod_{j=1}^K \Gamma(y^{(j)}+1)} \frac{\Gamma(\sum_{j=1}^K \alpha'_j)}{\prod_{j=1}^K \Gamma(\alpha'_j)} \frac{\prod_{j=1}^K \Gamma(y^{(j)} + \alpha'_j)}{\Gamma(n + \sum_{j=1}^K \alpha'_j)} \quad (1)$$

where dS_K denotes integrating $(p_1,\ p_K)$ with respect to the $(K-1)$ simplex.

Marignal. This derivation is almost the same as the posterior predictive.

$$f(D) = \int f(\theta|\alpha) \prod_{y_i \in D} f(y_i|\theta) d\theta$$

$$= \int f(p_1,\ .\ p_K|\alpha_1,\ \alpha_K) \prod_{y_i \in D} f(y_i|p_1,\ .\ p_K) dS_K$$

$$= \int \frac{\Gamma(\sum_{j=1}^K \alpha_j)}{\prod_{j=1}^K \Gamma(\alpha_j)} \prod_{j=1}^K p_j^{\alpha_j - 1} \prod_{y_i \in D} \frac{\Gamma(n+1)}{\prod_{j=1}^K \Gamma(y_i^{(j)}+1)} \prod_{j=1}^K p_j^{y_i^{(j)}} dS_K$$

$$= \frac{\Gamma(\sum_{j=1}^K \alpha_j)}{\prod_{j=1}^K \Gamma(\alpha_j)} [\prod_{y_i \in D} \frac{\Gamma(n+1)}{\prod_{j=1}^K \Gamma(y_i^{(j)}+1)}] \int \prod_{j=1}^K p_j^{\sum_{y_i \in D} y_i^{(j)} + \alpha_j - 1} dS_K$$

$$= \frac{\Gamma(\sum_{j=1}^{K} \alpha_j)}{\prod_{j=1}^{K} \Gamma(\alpha_j)} [\prod_{y_i \in D} \frac{\Gamma(n+1)}{\prod_{j=1}^{K} \Gamma(y_i^{(j)} + 1)}] \frac{\prod_{j=1}^{K} \Gamma(\sum_{y_i \in D} y_i^{(j)} + \alpha_j)}{\Gamma(|D| n + \sum_{j=1}^{K} \alpha_j)} \quad (2)$$

Dirichlet-Categorical

The derivations here are almost identical to before (with some minor syntatic differences).

Model.

$$p_1, \ p_K \sim \text{Dir}(\alpha_1, \ \alpha_K)$$

$$y \sim \text{Cat}(p_1, \ \ldots, p_K)$$

Posterior.

$$f(\theta|D) \propto f(\theta, \ D)$$

$$= f(p_1, \ \ldots, p_K|\alpha_1, \ . \alpha_K) \prod_{y_i \in D} f(y_i|p_1, \ \ldots p_K)$$

$$\propto \prod_{j=1}^{K} p_j^{\alpha_j - 1} \prod_{y_i \in D} \prod_{j=1}^{K} p_j^{1\{y_i = j\}}$$

$$= \prod_{j=1}^{K} p_j^{\alpha_j - 1 + \Sigma_{y_i \in D} 1\{y_i = j\}}$$

This density is exactly that of a Dirichlet distribution, except we have

$$\alpha_j' = \alpha_j + \sum_{y_i \in D} \text{I}\{y_i = j\}$$

That is, $f(\theta|D) = \text{Dir}(\alpha_1', \ \alpha_K')$.

Posterior Predictive.

$$f(y = x|D) = \int f(y = x|\theta) f(\theta|D) d\theta$$

$$= \int f(y = x|p_1, \ p_K) f(p_1, \ p_K|D) dS_K$$

$$= \int p_x \frac{\Gamma(\sum_{j=1}^{K} \alpha_j')}{\prod_{j=1}^{K} \Gamma(\alpha_j')} \prod_{j=1}^{K} p_j^{\alpha_j' - 1} dS_K$$

$$= \frac{\Gamma(\sum_{j=1}^{K} \alpha'_j)}{\prod_{j=1}^{K} \Gamma(\alpha'_j)} \int \prod_{j=1}^{K} p_j^{1\{x=j\}+\alpha'_j-1} dS_K$$

$$= \underline{\Gamma(\sum_{j=1}^{K} \alpha'_j)} \prod_{j=1}^{K} \Gamma(1\{x=j\} + \alpha'_j)$$

$$\overline{\prod_{j=1}^{K} \Gamma(\alpha'_j) \Gamma(1 + \sum_{j=1}^{K} \alpha'_j)}$$

$$= \frac{\alpha'_x}{\sum_{j=1}^{K} \alpha'_j} \quad (3)$$

where we used the fact that $\Gamma(n+1) = n\Gamma(n)$ to simplify the second to last line.

Marignal.

$$f(D) = \int f(\theta|\alpha) \prod_{y_i \in D} f(y_i|\theta) d\theta$$

$$= \int f(p_1, \ p_K|\alpha_1, \ \alpha_K) \prod_{y_i \in D} f(y_i|p_1, \ p_K) dS_K$$

$$= \int \frac{\Gamma(\sum_{j=1}^{K} \alpha_j)}{\prod_{j=1}^{K} \Gamma(\alpha_j)} \prod_{j=1}^{K} p_j^{\alpha_j - 1} \prod_{y_i \in D} \prod_{j=1}^{K} p_j^{1\{y_i=j\}} dS_K$$

$$= \frac{\Gamma(\sum_{j=1}^{K} \alpha_j)}{\prod_{j=1}^{K} \Gamma(\alpha_j)} \int \prod_{j=1}^{K} p_j^{\sum_{y_i \in D} 1\{y_i=j\}+\alpha_j-1} dS_K$$

$$= \frac{\Gamma(\sum_{j=1}^{K} \alpha_j)}{\prod_{j=1}^{K} \Gamma(\alpha_j)} \frac{\prod_{j=1}^{K} \Gamma(\sum_{y_i \in D} 1\{y_i = j\} + \alpha_j)}{\Gamma(|D| + \sum_{j=1}^{K} \alpha_j)} \quad (4)$$

Appendix B

First Paper Derivation

B.1 Section title

Here is some additional information which would have detracted from the point being made in the main article.

B.1.1 Subsection title

This section even has subtitles

Appendix 2

We have,

$$\hat{\Sigma}_{MAP0} = \underset{\Sigma_0}{\operatorname{argmax}} \left\{ \log \left(f\left(\mathbf{Y}|H_0, \Sigma_{MAP0}\right) f\left(\Sigma_{MAP0}\right) \right) \right\}$$

Taking log for both sides of eq. (4.20)

$$\begin{aligned}
&\log(f(\mathrm{Y}|H_0, \Sigma_0)) \\
&= -LM\log(\pi) + (-L)\log(\prod_{i=1}^{M}(\sigma_{0,i}^2)) - \sum_{i=1}^{M}\sum_{j=1}^{L}\frac{y_{ij}^* y_{ij}}{\sigma_{0,i}^2}
\end{aligned} \tag{B.1}$$

By taking log on both side of the both of eq. (4.16). We obtain

$$\begin{aligned}
&\log(f(\sigma_{0,1}^2, \cdots, \sigma_{0,M}^2)) = \\
&\log\left(\prod_{i=1}^{M}\frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)}(\sigma_{0,i}^2)^{-(\alpha_i+1)}\exp(\frac{-\beta_i}{\sigma_{0,i}^2})\right) \\
&= \log\left(\prod_{i=1}^{M}\frac{\beta_i^{\alpha}}{\Gamma(\alpha_i)}\right) - (\alpha_i+1)\sum_{i=1}^{M}\log\sigma_{0,i}^2 + \sum_{i=1}^{M}\frac{-\beta_i}{\sigma_{0,i}^2}
\end{aligned} \tag{B.2}$$

By adding eq. (B.1) and eq. (B.2), we obtain.

$$\begin{aligned}
\log(\sigma_{0,i}^2)_{MAP} = &-LM\log(\pi) + (-L)\log(\prod_{i=1}^{M}(\sigma_{0,i}^2)) \\
&- \sum_{i=1}^{M}\sum_{j=1}^{L}\frac{y_{ij}^* y_{ij}}{\sigma_{0,i}^2} + \log\left(\prod_{i=1}^{M}\frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)}\right) \\
&- (\alpha_i+1)\sum_{i=1}^{M}\log\sigma_{0,i}^2 + \sum_{i=1}^{M}\frac{-\beta_i}{\sigma_{0,i}^2}
\end{aligned} \tag{B.3}$$

Taking the derivative with respect $\sigma_{0,i}$ to eq. (B.3), we have

$$\frac{\partial \log \sigma_{0,i}^2}{\partial \sigma_{0,i}^2} = (\frac{L}{\sigma_{0,i}^2}) + (\frac{\sum_{j=1}^{L} y_{ij}^* y_{ij}}{(\sigma_{0,i}^2)^2}) - (\alpha_i + 1)(\frac{1}{\sigma_{0,i}^2}) + \frac{\beta_i}{(\sigma_{0,i}^2)^2} \tag{B.4}$$

setting the derivative to zero, we obtain the MAP of $\sigma_{0,i}^2$ as

$$\sigma_{0,i}^2 = \frac{\beta_i + \sum_{j=1}^{L} y_{ij}^* y_{ij}}{\alpha_i + 1 + L} \tag{B.5}$$

Appendix 2

Eq. (4.25) can be written equivalently as

$$\begin{aligned} \Sigma_{MAP_0} = & \underset{\substack{\Sigma_0:Digonal \\ \mathbf{M}:Digonal \\ \mathbf{X}\geq 0}}{\operatorname{argmax}} (L\,|\mathbf{X}| - tr\,\{\mathbf{RX}\} \\ & +(\alpha+1)\log|\mathbf{M}| - etr\,[\mathbf{BM}]) \end{aligned}$$

subject to

$$\Sigma_1^{-1} \leq \mathbf{X}$$
$$\Sigma_0^{-1} \leq \mathbf{M}$$

It also can be written as:

$$\Sigma_1^{-1} \leq \mathbf{X} \text{ as } (\Sigma_0 + \mathbf{H}\mathbf{H}^H)^{-1} \leq \mathbf{X}$$

By using Schur complement, the equation above can be written as:

$$(\Sigma_0 + \mathbf{H}\mathbf{H}^H)^{-1} \leq \mathbf{X} \Longleftrightarrow \begin{bmatrix} \mathbf{X} & \mathbf{I} \\ \mathbf{I} & \Sigma_0 + \mathbf{H}\mathbf{H}^H \end{bmatrix} \geq 0$$

Similarly,

$$\Sigma_0^{-1} \leq \mathbf{M} \Longleftrightarrow \begin{bmatrix} \mathbf{M} & \mathbf{I} \\ \mathbf{I} & \Sigma_0 \end{bmatrix} \geq 0$$

.

Then, P1 reduces to :

$$\begin{aligned} \Sigma_0 \quad = \quad & \underset{\substack{\Sigma_0 : Digonal \\ \mathbf{M} : Digonal \\ \mathbf{X} \geq 0}}{\operatorname{argmax}} \left(L \left| \mathbf{X} \right| - tr \left\{ \mathbf{RX} \right\} \right. \\ & \left. + (\alpha + 1) \log \left| \mathbf{M} \right| - tr \left[\mathbf{BM} \right] \right) \end{aligned}$$

P2: subject to

$$\begin{aligned} & \begin{bmatrix} \mathbf{M} & \mathbf{I} \\ \mathbf{I} & \Sigma_0 \end{bmatrix} \geq 0 \\ & \begin{bmatrix} \mathbf{X} & \mathbf{I} \\ \mathbf{I} & \Sigma_0 + \mathbf{H}\mathbf{H}^H \end{bmatrix} \geq 0 \\ & \mathbf{M} \geq \mathbf{X}. \end{aligned} \tag{B.6}$$

Appendix C

Second Paper Derivation

C.1 Section title

Here is some additional information which would have detracted from the point being made in the main article.

C.1.1 Subsection title

This section even has subtitles

Appendix 1

1. Eq. (4.25) can be written equivalently as

$$\begin{aligned}\mathbf{R}_{\eta} &= \underset{\substack{\mathbf{\Gamma}_0:Digonal \\ \mathbf{M}:Digonal \\ \mathbf{\Psi}\geq 0}}{\operatorname{argmax}} \left(L\left|\mathbf{\Psi}\right| - tr\left\{\mathbf{R\Psi}\right\}\right. \\ &\quad \left. +(\alpha+1)\log\left|\mathbf{M}\right| - etr\left[\mathbf{BM}\right]\right)\end{aligned}$$

subject to

$$\mathbf{\Gamma}_1^{-1} \leq \mathbf{\Psi}$$

$$\mathbf{\Gamma}_0^{-1} \leq \mathbf{M}$$

$$\mathbf{M} \geq \mathbf{\Psi}.$$

This problem can be solved by using Schur complement as shown in our previous work in [42] :

$$\begin{bmatrix} \mathbf{\Psi} & \mathbf{I} \\ \mathbf{I} & \mathbf{R}_\eta + \underbrace{(\widehat{\mathbf{R}}_h + \Delta\mathbf{R}_h)}_{\mathbf{R}_h} \end{bmatrix} \geq 0$$

Similarly,

$$\begin{bmatrix} \mathbf{M} & \mathbf{I} \\ \mathbf{I} & \mathbf{R}_\eta \end{bmatrix} \geq 0$$

2. In sub-optimal method (B-GLRT3), $\Delta\mathbf{R}_h$ is equal to $+\epsilon\mathbf{I}_{Nr}$ then problem reduces to

$$\begin{aligned} \mathbf{R}_\eta \quad = \quad & \underset{\substack{\mathbf{\Gamma}_0:Digonal \\ \mathbf{M}:Digonal \\ \mathbf{\Psi}\geq 0}}{\operatorname{argmax}} \left(L\left|\mathbf{\Psi}\right| - tr\left\{\mathbf{R}\mathbf{\Psi}\right\} \right. \\ & \left. +(\alpha+1)\log\left|\mathbf{M}\right| - etr\left[\mathbf{B}\mathbf{M}\right]\right) \end{aligned}$$

subject to

$$\begin{bmatrix} \mathbf{\Psi} & \mathbf{I} \\ \mathbf{I} & \mathbf{R}_\eta + \underbrace{(\widehat{\mathbf{R}}_h + \epsilon\mathbf{I}_{Nr})}_{\mathbf{R}_h} \end{bmatrix} \geq 0$$

$$\begin{bmatrix} \mathbf{M} & \mathbf{I} \\ \mathbf{I} & \mathbf{R}_\eta \end{bmatrix} \geq 0$$

$$\mathbf{M} \geq \mathbf{\Psi}$$

In sub-optimal method (B-GLRT3), $\mathbf{\Delta R}_h$ is equal to $+\epsilon \mathbf{I}_{Nr}$ then problem reduce to

$$\begin{aligned} \mathbf{R}_\eta \quad = \quad & \underset{\substack{\mathbf{\Gamma}_0 : Digonal \\ \mathbf{M} : Digonal \\ \mathbf{\Psi} \geq 0}}{\operatorname{argmax}} \left(L \left| \mathbf{\Psi} \right| - tr \left\{ \mathbf{R\Psi} \right\} \right. \\ & \left. + (\alpha + 1) \log |\mathbf{M}| - etr \left[\mathbf{BM} \right] \right) \end{aligned}$$

subject to

$$\begin{bmatrix} \mathbf{\Psi} & \mathbf{I} \\ \mathbf{I} & \mathbf{R}_\eta + \underbrace{(\widehat{\mathbf{R}}_h + \epsilon \mathbf{I}_{Nr})}_{\mathbf{R}_h} \end{bmatrix} \geq 0$$

$$\begin{bmatrix} \mathbf{M} & \mathbf{I} \\ \mathbf{I} & \mathbf{R}_\eta \end{bmatrix} \geq 0$$

$$\mathbf{M} \geq \mathbf{\Psi}$$

Appendix 2

1.

$$\begin{aligned} P1 \quad : \quad & \underset{\Gamma_0 : Diagonal}{\operatorname{argmax}} \left(L \log |\mathbf{R}_h + \mathbf{\Gamma}_0|^{-1} \right. \\ & -tr[\mathbf{R} \left(\mathbf{R}_h + \mathbf{\Gamma}_0 \right)_1^{-1}] \\ & \left. + (\rho + 1) \log \left| \mathbf{\Gamma}_0^{-1} \right| - tr[\mathbf{B\Gamma}_0^{-1}] \right) \end{aligned} \tag{C.1}$$

subject to

$$(\mathbf{R}_h + \mathbf{\Gamma}_0)^{-1} \leq \mathbf{\Psi}$$

$$\mathbf{\Gamma}_0^{-1} \leq \mathbf{M}$$

$$\mathbf{M} \geq \mathbf{\Psi}.$$

This problem can be solved according to the solution that is mentioned in appendix (1.1).

2.

$$\begin{aligned} P1 \quad : \quad & \underset{\Delta \mathbf{R}_h}{\operatorname{argmax}} (L \log |(\mathfrak{R} + \Delta \mathbf{R}_h)|^{-1} \\ & -tr[\mathbf{R}\,(\mathfrak{R} + \Delta \mathbf{R}_h)^{-1}] \\ & +(\rho + 1) \log \left|\mathbf{\Gamma}_0^{-1}\right| - tr[\mathbf{K}\mathbf{\Gamma}_0^{-1}]) \end{aligned} \tag{C.2}$$

subject to

$$(\mathfrak{R} + \Delta \mathbf{R}_h)^{-1} \leq \mathbf{\Psi}$$

$$\|\Delta \mathbf{R}_h\|_F \leq \epsilon.$$

This equation can also be solved using Schur complement, the problem will become:

$$\begin{bmatrix} \mathbf{\Psi} & \mathbf{I} \\ \mathbf{I} & (\mathfrak{R} + \Delta \mathbf{R}_h) \end{bmatrix} \geq 0$$

Bibliography

[1] Roy D Yates and David J Goodman. Probability and stochastic processes. *John Willey & Sons*, 1999.

[2] Ian F Akyildiz, Brandon F Lo, and Ravikumar Balakrishnan. Cooperative spectrum sensing in cognitive radio networks: A survey. *Physical communication*, 4(1):40–62, 2011.

[3] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008.

[4] Tevfik Yucek and Huseyin Arslan. A survey of spectrum sensing algorithms for cognitive radio applications. *IEEE communications surveys & tutorials*, 11(1):116–130, 2009.

[5] Beibei Wang and KJ Ray Liu. Advances in cognitive radio networks: A survey. *IEEE Journal of selected topics in signal processing*, 5(1):5–23, 2010.

[6] Jinlong Wang, Shuo Feng, Qihui Wu, Xueqiang Zheng, and Yuhua Xu. Hierarchical cognition cycle for cognitive radio networks. *China Communications*, 12(1):108–121, 2015.

[7] Zhengyu Zhang and Ling Qiu. Optimal progressive precoder design for arq packet retransmissions in nonregenerative mimo relay systems. In *2012 IEEE 75th Vehicular Technology Conference (VTC Spring)*, pages 1–5. IEEE, 2012.

[8] An-Zhi Chen, Zhi-Ping Shi, Hongxia Sun, Zhen-Qing He, Fan Bu, and Dandi Yang. A low-complexity spectrum sensing method for noncircular signal in cognitive radio networks with multiple receive antennas. *IEEE Communications Letters*, 2019.

[9] Daniel Fink. A compendium of conjugate priors. *See http://www. people. cornell. edu/pages/df36/CONJINTRnew% 20TEX. pdf*, 46, 1997.

[10] Srinu Sesham and Samrat L Sabat. Spectrum sensing for cognitive radio networks. In *White Space Communication*, pages 117–151. Springer, 2015.

[11] Naima Kaabouch. *Handbook of research on software-defined and cognitive radio technologies for dynamic spectrum management*. IGI global, 2014.

[12] Fatima Salahdine, Hassan El Ghazi, Naima Kaabouch, and Wassim Fassi Fihri. Matched filter detection with dynamic threshold for cognitive radio networks. In *2015 international conference on wireless networks and mobile communications (WINCOM)*, pages 1–6. IEEE, 2015.

[13] Hector Reyes, Sriram Subramaniam, Naima Kaabouch, and Wen Chen Hu. A spectrum sensing technique based on autocorrelation and euclidean distance and its comparison with energy detection for cognitive radio networks. *Computers & Electrical Engineering*, 52:319–327, 2016.

[14] N Armi, MZ Yusoff, and NM Saad. Decentralized cooperative user in opportunistic spectrum access system. In *2012 4th International Conference on Intelligent and Advanced Systems (ICIAS2012)*, volume 1, pages 179–183. IEEE, 2012.

[15] Amir Ghasemi and Elvino S Sousa. Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs. *IEEE Communications magazine*, 46(4):32–39, 2008.

[16] Joseph Mitola and Gerald Q Maguire. Cognitive radio: making software radios more personal. *IEEE personal communications*, 6(4):13–18, 1999.

[17] Subhajit Chatterjee, Jibendu Sekhar Roy, and Partha Pratim Bhattacharya. Spectrum sensing techniques for cognitive radio–a survey. *Int. J. of Appl. Engin. Res*, 10(7):16665–16684, 2015.

[18] Hui Shi, Tetsushi Abe, Takahiro Asai, and Hitoshi Yoshino. Relaying schemes using matrix triangularization for mimo wireless networks. *IEEE Transactions on Communications*, 55(9):1683–1688, 2007.

[19] Yu Zhang, Hanwen Luo, and Wen Chen. Efficient relay beamforming design with sic detection for dual-hop mimo relay networks. *IEEE transactions on vehicular technology*, 59(8):4192–4197, 2010.

[20] Youhua Fu, Luxi Yang, and Wei-Ping Zhu. A nearly optimal amplify-and-forward relaying scheme for two-hop mimo multi-relay networks. *IEEE Communications Letters*, 14(3):229–231, 2010.

[21] Young-Min Cho, Janghoon Yang, Jeongwook Seo, and Dong Ku Kim. Sum mse minimization for downlink multi-relay multi-user mimo network. *KSII Transactions on Internet & Information Systems*, 8(8), 2014.

[22] Zukang Shen, Runhua Chen, Jeffrey G Andrews, Robert W Heath, and Brian L Evans. Low complexity user selection algorithms for multiuser mimo systems with block diagonalization. *IEEE Transactions on Signal Processing*, 54(9):3658–3663, 2006.

[23] Yuan Yu and Yingbo Hua. Power allocation for a mimo relay system with multiple-antenna users. *IEEE Transactions on Signal Processing*, 58(5):2823–2835, 2010.

[24] Zhengyu Zhang, Ling Qiu, and Jie Xu. Progressive linear precoder optimization for arq packet retransmissions in nonregenerative mimo relay systems. *IEEE transactions on signal processing*, 61(1):68–81, 2012.

[25] Rahul Tandra and Anant Sahai. Snr walls for signal detection. *IEEE Journal of selected topics in Signal Processing*, 2(1):4–17, 2008.

[26] Richard M Feldman and Ciriaco Valdez-Flores. *Applied probability and stochastic processes*. Springer, 2010.

[27] Cheng Zhang, Yongming Huang, Yindi Jing, Shi Jin, and Luxi Yang. Sum-rate analysis for massive mimo downlink with joint statistical beamforming and user scheduling. *IEEE Transactions on Wireless Communications*, 16(4):2181–2194, 2017.

[28] Alexandros Manolakos, Yair Noam, and Andrea J Goldsmith. Interference due to null space mismatch in cooperative multipoint mimo cellular networks. In *2014 IEEE International Conference on Communications (ICC)*, pages 5148–5153. IEEE, 2014.

[29] Feifei Gao, Rui Zhang, Ying-Chang Liang, and Xiaodong Wang. Design of learning-based mimo cognitive radio systems. *IEEE Transactions on Vehicular Technology*, 59(4):1707–1720, 2010.

[30] Ahmed Al-Asadi, Muthana Al-Amidie, Athanasios C Micheas, Ronald G McGarvey, and Naz E Islam. Worst case fair beamforming for multiple multicast groups in multicell networks. *IET Communications*, 13(6):664–671, 2018.

[31] Jyotiprasad Medhi. *Stochastic processes*. New Age International, 1994.

[32] Meisam Razaviyayn. Successive convex approximation: Analysis and applications. 2014.

[33] Anthony Lin et al. Binary search algorithm. *WikiJournal of Science*, 2(1):1, 2019.

[34] Zdzislaw Brzezniak and Tomasz Zastawniak. *Basic stochastic processes: a course through exercises*. Springer Science & Business Media, 2000.

[35] Sheldon M Ross, John J Kelly, Roger J Sullivan, William James Perry, Donald Mercer, Ruth M Davis, Thomas Dell Washburn, Earl V Sager, Joseph B Boyce, and Vincent L Bristow. *Stochastic processes*, volume 2. Wiley New York, 1996.

[36] Andrea Goldsmith. *Wireless communications*. Cambridge university press, 2005.

[37] Sissi Xiaoxiao Wu, Wing-Kin Ma, and Anthony Man-Cho So. Physical-layer multicasting by stochastic transmit beamforming and alamouti space-time coding. *IEEE Transactions on Signal Processing*, 61(17):4230–4245, 2013.

[38] Sudharman K Jayaweera and Carlos Mosquera. A dynamic spectrum leasing (dsl) framework for spectrum sharing in cognitive radio networks. In *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, pages 1819–1823. IEEE, 2009.

[39] Alladi Ramakrishnan. Probability and stochastic processes. In *Prinzipien der Thermodynamik und Statistik/Principles of Thermodynamics and Statistics*, pages 524–651. Springer, 1959.

[40] Adam Bobrowski. *Functional analysis for probability and stochastic processes: an introduction*. Cambridge University Press, 2005.

[41] Mokhtar S Bazaraa, Hanif D Sherali, and Chitharanjan M Shetty. *Nonlinear programming: theory and algorithms*. John Wiley & Sons, 2013.

[42] Muthana Al-Amidie, Ahmed Al-Asadi, Athanasios C Micheas, and Naz E Islam. Spectrum sensing based on bayesian generalised likelihood ratio for cognitive radio systems with multiple antennas. *IET Communications*, 13(3):305–311, 2018.

[43] Ka Lung Law, Christos Masouros, and Marius Pesavento. Transmit beamforming for interference exploitation in the underlay cognitive radio z-channel. *arXiv preprint arXiv:1606.06504*, 2016.

[44] Saad Qaisar, Rana Muhammad Bilal, Wafa Iqbal, Muqaddas Naureen, and Sungyoung Lee. Compressive sensing: From theory to applications, a survey. *Journal of Communications and networks*, 15(5):443–456, 2013.

[45] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

[46] Yair Noam, Alexandros Manolakos, and Andrea J Goldsmith. Null space learning with interference feedback for spatial division multiple access. *IEEE Transactions on Wireless Communications*, 13(10):5699–5715, 2014.

[47] Michael Alan Howarth Dempster, Gautam Mitra, and Georg Pflug. *Quantitative fund management*. Chapman and Hall/CRC, 2008.

[48] Jonathan Borwein and Adrian S Lewis. *Convex analysis and nonlinear optimization: theory and examples*. Springer Science & Business Media, 2010.

[49] Dimitrios Christopoulos, Symeon Chatzinotas, and Björn Ottersten. Weighted fair multicast multigroup beamforming under per-antenna power constraints. *IEEE Transactions on Signal Processing*, 62(19):5132–5142, 2014.

[50] Erik Axell and Erik G Larsson. Optimal and near-optimal spectrum sensing of ofdm signals in awgn channels. In *Cognitive Information Processing (CIP), 2010 2nd International Workshop on*, pages 128–133. IEEE, 2010.

[51] Michael Grant, Stephen Boyd, and Yinyu Ye. Cvx: Matlab software for disciplined convex programming (2008). *Web page and software available at http://stanford. edu/ boyd/cvx*, 2015.

[52] Xin Kang, Rui Zhang, Ying-Chang Liang, and Hari Krishna Garg. Optimal power allocation strategies for fading cognitive radio channels with primary user outage constraint. *IEEE Journal on Selected Areas in Communications*, 29(2):374–383, 2011.

[53] Gerard J Foschini and Michael J Gans. On limits of wireless communications in a fading environment when using multiple antennas. *Wireless personal communications*, 6(3):311–335, 1998.

[54] Nasrullah Armi, Mohd Zuki Yusoff, and Naufal M Saad. Cooperative spectrum sensing in decentralized cognitive radio system. In *Eurocon 2013*, pages 113–118. IEEE, 2013.

[55] Shruti Dutta and Komal Arora. *Improvement in Spectrum Sensing Techniques for Cognitive Radio*. PhD thesis, Lovely Professional University, 2017.

[56] Neda Aboutorab, Wibowo Hardjawana, and Branka Vucetic. A new iterative doppler-assisted channel estimation joint with parallel ici cancellation for high-mobility mimo-ofdm systems. *IEEE Transactions on Vehicular Technology*, 61(4):1577–1589, 2012.

[57] Elrharras Abdessamad, Rachid Saadane, Mohamed El Aroussi, Mohammed Wahbi, and Abdellatif Hamdoun. Spectrum sensing with an improved energy detection. In *2014 International Conference on Multimedia Computing and Systems (ICMCS)*, pages 895–900. IEEE, 2014.

[58] P Avinash, R Gandhiraj, and KP Soman. Spectrum sensing using compressed sensing techniques for sparse multiband signals. *International Journal of Scientific & Engineering Research*, 3(5):1–5, 2012.

[59] Mohammad Reza Abedi, Nader Mokari, Mohammad Reza Javan, and Halim Yanikomeroglu. Secure communication in ofdma-based cognitive radio networks: An incentivized secondary network coexistence approach. *IEEE Transactions on Vehicular Technology*, 66(2):1171–1185, 2016.

[60] James W Demmel. *Applied numerical linear algebra*. SIAM, 1997.

[61] Mohd Zuki bin Yusoff and Mohamad Naufal bin Mohamad Saad. Decentralized cooperative user in opportunistic spectrum access system. 2012.

[62] Erik Axell and Erik G Larsson. Optimal and sub-optimal spectrum sensing of ofdm signals in known and unknown noise variance. *IEEE J. Sel. Areas Commun.*, 29(2):290–304, 2011.

[63] Christoph Stöckle, Jawad Munir, Amine Mezghani, and Josef A Nossek. Channel estimation in massive mimo systems using 1-bit quantization. In *2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–6. IEEE, 2016.

[64] Shimin Gong, Ping Wang, and Wei Liu. Spectrum sensing under distribution uncertainty in cognitive radio networks. In *2012 IEEE International Conference on Communications (ICC)*, pages 1512–1516. IEEE, 2012.

[65] Tengyi Zhang, Yuan Wu, Ke Lang, and Danny HK Tsang. Optimal scheduling of cooperative spectrum sensing in cognitive radio networks. *IEEE Systems Journal*, 4(4):535–549, 2010.

[66] Waleed Ejaz, Najam ul Hasan, Muhammad Awais Azam, and Hyung Seok Kim. Improved local spectrum sensing for cognitive radio networks. *EURASIP Journal on Advances in Signal Processing*, 2012(1):242, 2012.

[67] K Aditya Mohan and Chandra R Murthy. Cooperative sequential binary hypothesis testing using cyclostationary features. In *2010 IEEE 11th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–5. IEEE, 2010.

[68] Athar Ali Khan, Mubashir Husain Rehmani, and Martin Reisslein. Cognitive radio for smart grids: Survey of architectures, spectrum sensing mechanisms, and networking protocols. *IEEE Communications Surveys & Tutorials*, 18(1):860–898, 2015.

[69] Shridhar Mubaraq Mishra, Rahul Tandra, and Anant Sahai. The case for multiband sensing. In *Proc. of the Forty-fifth Annual Allerton Conference on Communication, Control, and Computing*, pages 1–10. Citeseer, 2007.

[70] F Weidling, D Datla, V Petty, P Krishnan, and GJ Minden. A framework for rf spectrum measurements and analysis. In *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005.*, pages 573–576. IEEE, 2005.

[71] Xiang Ling, Bin Wu, Hong Wen, Pin-Han Ho, Zhiqiang Bao, and Lili Pan. Adaptive threshold control for energy detection based spectrum sensing in cognitive radios. *IEEE Wireless Communications Letters*, 1(5):448–451, 2012.

[72] CH Lim. Adaptive energy detection for spectrum sensing in unknown white gaussian noise. *IET communications*, 6(13):1884–1889, 2012.

[73] Miguel López-Benítez and Fernando Casadevall. Improved energy detection spectrum sensing for cognitive radio. *IET communications*, 6(8):785–796, 2012.

[74] Johanna Vartiainen, Heli Sarvanko, Janne Lehtomaki, Markku Juntti, and Matti Latva-aho. Spectrum sensing with lad-based methods. In *2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 1–5. IEEE, 2007.

[75] Youness Arjoune and Naima Kaabouch. A comprehensive survey on spectrum sensing in cognitive radio networks: Recent advances, new challenges, and future research directions. *Sensors*, 19(1):126, 2019.

[76] Ezio Biglieri, Robert Calderbank, Anthony Constantinides, Andrea Goldsmith, Arogyaswami Paulraj, and H Vincent Poor. *MIMO wireless communications*. Cambridge university press, 2007.

[77] Wibowo Hardjawana, Rui Li, Branka Vucetic, Yonghui Li, and Xuezhi Yang. A new iterative channel estimation for high mobility mimo-ofdm systems. In *2010 IEEE 71st Vehicular Technology Conference*, pages 1–5. IEEE, 2010.

[78] Nina Wang, Guan Gui, Zhi Zhang, Tian Tang, and Jun Jiang. A novel sparse channel estimation method for multipath mimo-ofdm systems. In *2011 IEEE Vehicular Technology Conference (VTC Fall)*, pages 1–5. IEEE, 2011.

[79] Yang Liu and Serdar Sezginer. Iterative compensated mmse channel estimation in lte systems. In *2012 IEEE International Conference on Communications (ICC)*, pages 4862–4866. IEEE, 2012.

[80] Johanna Ketonen, Markku Juntti, Jari Ylioinas, and Joseph R Cavallaro. Decision-directed channel estimation implementation for spectral efficiency improvement in mobile mimo-ofdm. *Journal of Signal Processing Systems*, 79(3):233–245, 2015.

[81] Rui Zhang. On active learning and supervised transmission of spectrum sharing based cognitive radios by exploiting hidden primary radio feedback. *IEEE Transactions on Communications*, 58(10):2960–2970, 2010.

[82] Kun-Yu Wang, Neil Jacklin, Zhi Ding, and Chong-Yung Chi. Robust {MISO Transmit} optimization under outage-based qos constraints in two-tier heterogeneous networks. *IEEE Transactions on Wireless Communications*, 12(4):1883–1897, 2013.

[83] Shuai Ma and Dechun Sun. Chance constrained robust beamforming in cognitive radio networks. *IEEE Communications Letters*, 17(1):67–70, 2012.

[84] Daniel Pérez Palomar, John M Cioffi, and Miguel Angel Lagunas. Uniform power allocation in mimo channels: A game-theoretic approach. *IEEE Transactions on Information Theory*, 49(7):1707–1727, 2003.

[85] Huiyue Yi. Nullspace-based secondary joint transceiver scheme for cognitive radio mimo networks using second-order statistics. In *2010 IEEE International Conference on Communications*, pages 1–5. IEEE, 2010.

[86] Danijela Cabric. Addressing feasibility of cognitive radios. *IEEE Signal Process. Mag.*, 25(6), 2008.

[87] Emmanuel J Candès et al. Compressive sampling. In *Proceedings of the international congress of mathematicians*, volume 3, pages 1433–1452. Madrid, Spain, 2006.

[88] JB Hiriart-Urruty and C Lemaréchal. Fundamentals of convex 978 analysis, 2012.

[89] R Tyrrell Rockafellar. *Convex analysis*, volume 36. Princeton university press, 1970.

[90] David G Luenberger, Yinyu Ye, et al. *Linear and nonlinear programming*, volume 2. Springer, 1984.

[91] Haitham Hindi. A tutorial on convex optimization. In *Proceedings of the 2004 American Control Conference*, volume 4, pages 3252–3265. IEEE, 2004.

[92] Gesualdo Scutari and Ying Sun. Parallel and distributed successive convex approximation methods for big-data optimization. In *Multi-agent Optimization*, pages 141–308. Springer, 2018.

[93] Robert M Freund. The steepest descent algorithm for unconstrained optimization and a bisection line-search method. *Journal of Massachusetts Institute of Technology. United States of america*, 2004.

[94] Andrea Goldsmith, Syed Ali Jafar, Ivana Maric, and Sudhir Srinivasa. Breaking spectrum gridlock with cognitive radios: An information theoretic perspective. *Proceedings of the IEEE*, 97(5):894–914, 2009.

[95] Qing Zhao and Brian M Sadler. A survey of dynamic spectrum access. *IEEE Signal Process. Mag.*, 24(3):79–89, 2007.

[96] Zeng Yonghong and Liang Ying-Chang. Eigenvalue-based spectrum sensing algorithms for cognitive radio. *IEEE Trans. on Commun.*, 57(6):1784–1793, 2009.

[97] Rui Zhang and Ying-Chang Liang. Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks. *IEEE J. Sel. topics in Signal Process.*, 2(1):88–102, 2008.

[98] Pascal Bianchi, Merouane Debbah, Mylène Maïda, and Jamal Najim. Performance of statistical tests for single-source detection using random matrix theory. *IEEE Trans.Inf. theory*, 57(4):2400–2419, 2011.

[99] Yonghong Zeng, Ying-Chang Liang, and Rui Zhang. Blindly combined energy detection for spectrum sensing in cognitive radio. *IEEE Signal Process. Lett.*, 15:649–652, 2008.

[100] Simon Haykin. Cognitive radio: brain-empowered wireless communications. *IEEE J. Sel. areas in Commun.*, 23(2):201–220, 2005.

[101] Josep Font-Segura and Xiaodong Wang. Glrt-based spectrum sensing for cognitive radio with prior information. *IEEE Trans.Commun.*, 58(7):2137–2146, 2010.

[102] Huan Cong Nguyen, Elisabeth De Carvalho, and Ramjee Prasad. Spectrum sensing for cognitive radio based on multiple antennas. In *Vehicular Technology Conf. (VTC Spring), 2012 IEEE 75th*, pages 1–5. IEEE, 2012.

[103] Xi Yang, Shengliang Peng, Pengcheng Zhu, Hongyang Chen, and Xiuying Cao. Effect of correlations on the performance of glrt detector in cognitive radios. *IEICE Trans. Commun.*, 94(4):1089–1093, 2011.

[104] Abbas Taherpour, Masoumeh Nasiri-Kenari, and Saeed Gazor. Multiple antenna spectrum sensing in cognitive radios. *IEEE Trans. on Wirel. Commun.*, 9(2):814–823, 2010.

[105] David Ramírez, Gonzalo Vazquez-Vilar, Roberto López-Valcarce, Javier Vía, and Ignacio Santamaría. Detection of rank-p signals in cognitive radio networks with uncalibrated multiple antennas. *IEEE Trans. on Signal Process.*, 59(8):3764–3774, 2011.

[106] Ashish Pandharipande and J-PMG Linnartz. Performance analysis of primary user detection in a multiple antenna cognitive radio. In *Commun., 2007. ICC'07. IEEE International Conf.*, pages 6482–6486. IEEE, 2007.

[107] Erik G Larsson and Petre Stoica. *Space-time block coding for wireless communications*. Cambridge university press, 2008.

[108] Vijaykumar Kuppusamy and Rajarshi Mahapatra. Primary user detection in ofdm based mimo cognitive radio. In *2008 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom 2008)*, pages 1–5. IEEE, 2008.

[109] Steven M Kay. Fund. stat. signal processing. detec. theory, volume ii. 1998.

[110] Arnold Zellner. On assessing prior distributions and bayesian regression analysis with g-prior distributions. *Bayesian inference and decision techniques*, 1986.

[111] Saeid Sedighi, Abbas Taherpour, Josep Sala-Alvarez, and Tamer Khattab. On the performance of hadamard ratio detector-based spectrum sensing for cognitive radios. *IEEE Transactions on Signal Processing*, 63(14):3809–3824, 2015.

[112] Saeid Sedighi, Abbas Taherpour, and Shaghayegh SM Monfared. Bayesian generalised likelihood ratio test-based multiple antenna spectrum sensing for cognitive radios. *IET Commun.*, 7(18):2151–2165, 2013.

[113] Maria J Bayarri, James O Berger, Anabel Forte, G García-Donato, et al. Criteria for bayesian model choice with application to variable selection. *Ann. stat.*, 40(3):1550–1577, 2012.

[114] J-L Gauvain and Chin-Hui Lee. Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains. *IEEE Trans. speech and audio process.*, 2(2):291–298, 1994.

[115] Mark A McHenry, Peter A Tenhula, Dan McCloskey, Dennis A Roberson, and Cynthia S Hood. Chicago spectrum occupancy measurements & analysis and a long-term studies proposal. In *Proceedings of the first international workshop on Technology and policy for accessing spectrum*, page 1. ACM, 2006.

[116] Gürkan Gür and Fatih Alagöz. Green wireless communications via cognitive dimension: an overview. *IEEE network*, 25(2):50–56, 2011.

[117] Hongjian Sun, Arumugam Nallanathan, Cheng-Xiang Wang, and Yunfei Chen. Wideband spectrum sensing for cognitive radio networks: a survey. *IEEE Wireless Communications*, 20(2):74–81, 2013.

[118] Jun Ma, Guodong Zhao, and Ye Li. Soft combination and detection for cooperative spectrum sensing in cognitive radio networks. *IEEE Transactions on Wireless Communications*, 7(11):4502–4507, 2008.

[119] Qing-hua Shen, Xuan-li Wu, Xue-jun Sha, and Nai-tong Zhang. Cooperative spectrum sensing under imperfect channel estimation. In *Communications and Mobile Computing (CMC), 2010 International Conference on*, volume 2, pages 174–178. IEEE, 2010.

[120] S Force. Spectrum policy task force report. *Federal Communications Commission ET Docket 02, vol. 135*, 2002.

[121] Anant Sahai, Niels Hoven, and Rahul Tandra. Some fundamental limits on cognitive radio. In *Allerton Conference on Communication, Control, and Computing*, pages 1662–1671. Monticello, Illinois, 2004.

[122] Alexander Sonnenschein and Philip M Fishman. Radiometric detection of spread-spectrum signals in noise of uncertain power. *IEEE Transactions on Aerospace and Electronic Systems*, 28(3):654–660, 1992.

[123] Hao He, Geoffrey Ye Li, and Shaoqian Li. Adaptive spectrum sensing for time-varying channels in cognitive radios. *IEEE Wireless Communications Letters*, 2(2):1–4, 2013.

[124] Abbas Taherpour, Masoumeh Nasiri-Kenari, and Saeed Gazor. Multiple antenna spectrum sensing in cognitive radios. *IEEE transactions on wireless communications*, 9(2):814–823, 2010.

[125] Erik Axell, Geert Leus, Erik G Larsson, and H Vincent Poor. Spectrum sensing for cognitive radio: State-of-the-art and recent advances. *IEEE Signal processing magazine*, 29(3):101–116, 2012.

[126] Roberto López-Valcarce, Gonzalo Vazquez-Vilar, and Josep Sala. Multiantenna spectrum sensing for cognitive radio: overcoming noise uncertainty. In *Cognitive Information Processing (CIP), 2010 2nd International Workshop on*, pages 310–315. IEEE, 2010.

[127] Lei Shen, Haiquan Wang, Wei Zhang, and Zhijin Zhao. Multiple antennas assisted blind spectrum sensing in cognitive radio channels. *IEEE Commun. Letters*, 16(1):92–94, 2012.

[128] Mark Kliger and Joseph M Francos. Map model order selection rule for 2-d sinusoids in white noise. *IEEE Transactions on Signal Processing*, 53(7):2563–2575, 2005.

[129] Petre Stoica and Prabhu Babu. On the exponentially embedded family (eef) rule for model order selection. *IEEE Signal Processing Letters*, 19(9):551–554, 2012.

[130] Gang Liang, D Mitchell Wilkes, and James A Cadzow. Arma model order estimation based on the eigenvalues of the covariance matrix. *IEEE transactions on signal processing*, 41(10):3003–3009, 1993.

[131] Sreejith Kallummil and Sheetal Kalyani. High snr consistent linear model order selection and subset selection. *IEEE Transactions on Signal Processing*, 64(16):4307–4322, 2016.

[132] Andrea Goldsmith, Syed Ali Jafar, Nihar Jindal, and Sriram Vishwanath. Capacity limits of mimo channels. *IEEE Journal on selected areas in Communications*, 21(5):684–702, 2003.

[133] Sriram Vishwanath, Nihar Jindal, and Andrea Goldsmith. Duality, achievable rates, and sum-rate capacity of gaussian mimo broadcast channels. *IEEE Transactions on Information Theory*, 49(10):2658–2668, 2003.

[134] Arogyaswami J Paulraj, Dhananjay A Gore, Rohit U Nabar, and Helmut Bolcskei. An overview of mimo communications-a key to gigabit wireless. *Proceedings of the IEEE*, 92(2):198–218, 2004.

[135] Chinmay S Vaze and Mahesh K Varanasi. On the achievable rate of the fading dirty paper channel with imperfect csit. In *2009 43rd Annual Conference on Information Sciences and Systems*, pages 346–351. IEEE, 2009.

[136] Chinmay S Vaze and Mahesh K Varanasi. Dirty paper coding for fading channels with partial transmitter side information. In *2008 42nd Asilomar Conference on Signals, Systems and Computers*, pages 341–345. IEEE, 2008.

[137] Chinmay S Vaze and Mahesh K Varanasi. Dirty paper coding for the mimo cognitive radio channel with imperfect csit. In *2009 IEEE International Symposium on Information Theory*, pages 2532–2536. IEEE, 2009.

[138] Le-Nam Tran, Markku Juntti, Mats Bengtsson, and Bjorn Ottersten. Weighted sum rate maximization for mimo broadcast channels using dirty paper coding and zero-forcing methods. *IEEE Transactions on Communications*, 61(6):2362–2373, 2013.

[139] Adarsh Patel and Aditya K Jagannatham. Non-antipodal signaling based robust detection for cooperative spectrum sensing in mimo cognitive radio networks. *IEEE Signal Processing Letters*, 20(7):661–664, 2013.

[140] Adarsh Patel, Bhishm Tripathi, and Aditya K Jagannatham. Robust estimator-correlator for spectrum sensing in mimo cr networks with csi uncertainty. *IEEE Wireless Communications Letters*, 3(3):253–256, 2014.

[141] Steven M Kay. *Fundamentals of statistical signal processing: Practical algorithm development*, volume 3. Pearson Education, 2013.

[142] Samuel Stanley Wilks. Mathematical statistics. 1947.

[143] George V Moustakides. Finite sample size optimality of glr tests. *arXiv preprint arXiv:0903.3795*, 2009.

[144] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.

[145] Mohannad Al-Ali and K.C. Ho. Transmit precoding in underlay mimo cognitive radio with unavailable or imperfect knowledge of primary interference channel. *IEEE Transactions on Wireless Communications*, pages 1–1, 01 2016.

[146] Shree Krishna Sharma, Symeon Chatzinotas, and Bjorn Ottersten. Snr estimation for multi-dimensional cognitive receiver under correlated channel/noise. *IEEE Trans. on Wirel. Commun.*, 12(12):6392–6405, 2013.

[147] Fuhui Zhou and Norman C Beaulieu. An improved and more accurate expression for a pdf related to eigenvalue-based spectrum sensing. *IEEE Systems Journal*, 13(2):1320–1323, 2018.

VITA

Muthana Al-Amidie was born on Nov. 11^{th} 1982 in Babylon, Iraq. He received the B.Sc. and M.Sc. degrees in Electrical Engineering from the University of Babylon, Babylon, University of Technology, Iraq, in 2004 and 2011, respectively. He joined the Electrical and Computer Engineering department at the University of Missouri-Columbia in January 2015 and since then he has begun working with Prof. Nas Islam on his research. He received a master of engineering degree in electrical engineering in Dec. 2020 and He expects to receive the Doctor of Philosophy degree on Dec. 2021.

Al-Amidie is married to Dr. Sura Muhsin with three kids: Zain, Ali, Sara. After completing the Ph.D. requirements, he will return back to his country and serve as a lecturer in Electrical Engineering- University of Babylon, Iraq.