Robust optimization and its applications in cognitive radio networks.

Gong, Shimin.

2013

Gong, S. (2013). Robust optimization and its applications in cognitive radio networks.
Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/55063>

<https://doi.org/10.32657/10356/55063>

NANYANG TECHNOLOGICAL UNIVERSITY

Robust Optimization and Its Applications in Cognitive Radio Networks

A dissertation submitted to the
Nanyang Technological University

by

Shimin Gong

in fulfillment of the requirement
for the degree of Doctor of Philosophy
of Computer Engineering

27 November 2013

Abstract

An optimal design of the cognitive radio may not operate properly due to unexpected fluctuations of system parameters in non-ideal working environment. Among various uncertain factors, the most notorious factor is the fluctuation of channel gain due to the stochastic nature of wireless channels. In our research, we try to propose a new framework that brings uncertainty into the analysis and design of a practical system. As a demonstration, we apply this framework to study two major aspects of the cognitive radio networks, i.e., spectrum identification and exploitation.

Our first challenge is the characterization of channel uncertainty, which usually results in unreliable sensing performance. Though a lot of existing works focus on improving the detection probability, few works consider the reliability or robustness of the detector. The design objective of a robust detector is to provide a sensing performance that is not changing significantly with respect to the fluctuations of wireless channel. In this work, we propose a new model to describe the channel uncertainty, namely, the distribution uncertainty, that combines two well-known approaches, i.e., the stochastic and the worst-case robust approaches. This model allows the uncertain parameter to draw from a distribution function, but the type of distribution is not deterministic and can be arbitrarily selected from a pool of distributions, i.e., uncertainty set. To the best of our knowledge, we are the first to introduce distribution uncertainty into robust sensing of cognitive radio networks.

Based on the characterization of channel uncertainty, we study the robust performance of spectrum sensing, and aim at providing analytical results for the lower bound of detection probability, which gives the secondary users (SUs) a guaranteed performance even under worst-case channel fluctuation. Specifically, we propose two types of uncertainties. For the moment-based uncertainty, the lower bound of detection probability can be found in a convex semi-definite program. While for the other reference-based uncertainty, though we cannot find closed-form result, we propose an iterative procedure that will converge to the lower detection bound.

Another challenge is to design the system parameters and improve the robust detection performance. We consider the robust design of decision thresholds in multi-user cooperative spectrum sensing. It is shown to be a non-convex problem, however,

we can approximate the design problem by a series of tractable semi-definite programs and propose an iterative algorithm to search the optimal decision threshold for each SU while maintaining the desired false alarm probability. We also demonstrate, through simulations, that the robust design can provide more stable sensing performance no matter how the channel gain fluctuates.

Besides, we also face a challenge in designing efficient spectrum utilization mechanism. Considering a power control problem, if SUs' estimations of the interference power at primary user (PU) side are subject to channel uncertainty, their transmissions will easily violate PUs' interference power constraints thus degrade PUs' throughput performance. Based on our work in robust spectrum sensing, we present a robust formulation of the power control problem for SUs, and propose several iterative algorithms to determine the robust transmit power of SUs. Simulation results show that our robust design provides better PU protection than the existing works which fail to take account of channel uncertainty.

In summary, our works investigate two important aspects in cognitive radio networks, i.e., spectrum identification by spectrum sensing, and efficient exploitation by power control. We find that channel uncertainty is a common challenge for both of these two aspects. Our main contribution lies in the introduction of distribution uncertainty into channel modeling, based on which we propose a novel framework to design the robust versions of spectrum sensing and power control algorithms.

Acknowledgments

The journey of a PhD likes a solo traveler on the way of searching for the most beautiful scenes. My professor inspires my desire to exploit the ultimate aesthetics, providing me drinking water when I feel thirsty, a compass when I get lost, and the skills to fight with ferocious storms and beasts. I would express my deep gratitude to my professor, Dr. Ping Wang, who has led me into the wonderland of research.

Over the past four years, she has taught me a lot on how to find topics and how to research. She inspires my enthusiasm of thinking independently, but also shares her insights and experiences on research. She values independent thinking, but also encourages cross-border collaborations. She is not only a perfect supervisor for my research, but also a good colleague and friend in my future career path.

I also want to thank Professor Jianwei Huang, and Professor Weihua Zhuang for accepting me as a visiting scholar. The discussions with them turned out to be very helpful for my research and related to Chapter 4 and Chapter 5, respectively.

Finally, I devote the most special gratitude to my family for their unconditional love and support, my fiancée for her understanding and endurance.

This page is intentionally left blank

Contents

Abstract	i
Acknowledgments	iii
List of Figures	viii
List of Tables	ix
List of Nomenclature	x
1 Introduction	1
1.1 Background and Motivations	2
1.2 Research Scope and Problems	4
1.2.1 System Robustness	4
1.2.2 Spectrum Identification	4
1.2.3 Spectrum Exploitation	5
1.3 Research Methodology	6
1.4 Contributions and Organization	8
2 Literature Review	11
2.1 Optimization of Sensing Parameters	11
2.1.1 Sensing Decision Threshold	12
2.1.2 Sensing Time and Tradeoffs	14
2.2 Multi-user Cooperative Sensing	15
2.3 Joint Sensing and Power Control	17
2.4 Robust Cognitive Power Control	19
3 Robust Spectrum Sensing for Single-User	21
3.1 Performance Metrics and Uncertainty Models	22
3.2 Analysis of the Moment-based Uncertainty	26
3.2.1 Lower Bound of Detection Probability	27
3.2.2 Equivalent Semi-definite Program	27
3.2.3 Bound Achieving Distribution	29
3.3 Analysis of the Reference-based Uncertainty	30

3.3.1	Reference-based Uncertainty Model	31
3.3.2	Determine a Robust Decision Threshold	32
3.3.3	Lower Bound of Detection Probability	35
3.4	Numerical Results	36
3.4.1	False Alarm Probability under Noise Uncertainties	37
3.4.2	Detection Bounds with Moment Uncertainties	39
3.4.3	The Importance of a Reference Distribution	40
3.5	Summary	42
3.6	Appendix	43
3.6.1	Proof for Proposition 3.1	43
3.6.2	Proof for Proposition 3.2	43
3.6.3	Proof for Proposition 3.3	45
4	Robust Spectrum Sensing for Multi-User	46
4.1	Multi-user Spectrum Sensing Model	47
4.1.1	Sensing Structure	47
4.1.2	GDUS Model for Multi-User	48
4.1.3	Robust Sensing Design	50
4.2	Algorithm for the Robust Design	51
4.2.1	Initial Setting	52
4.2.2	Iterative Algorithm	53
4.3	Gradient Aided Decision Threshold Update	54
4.3.1	Atomic Movement on Boundary	55
4.3.2	Ascent Search Direction	56
4.3.3	Optimal Step Size	57
4.4	Numerical Analysis	59
4.4.1	Comparison between Robust and Nominal Design	59
4.4.2	Algorithm Convergence	61
4.4.3	Reverse Ordering between λ and μ	62
4.5	Summary	64
4.6	Appendix	65
4.6.1	Proof for Theorem 4.1	65
4.6.2	Proof for Proposition 4.1	66
4.6.3	Proof for Theorem 4.2	66
4.6.4	Proof for Proposition 4.2	68
4.6.5	The Relation Between Δ_i and Δ_j in Table 4.1	69
4.6.6	Proof for Proposition 4.4	70

5	Robust Power Control with Distribution Uncertainty	71
5.1	System Model	72
5.1.1	Time-varying Channel Model	73
5.1.2	Stochastic PU Protection and QoS Provisioning	73
5.1.3	Reference-based Distribution Uncertainty	75
5.2	Case I: SAPs Transmit with the Same Power	76
5.2.1	Protection for Primary Network	76
5.2.2	Worst-case Interference Power	78
5.2.3	Controlling Robust Transmit Power	79
5.3	Case II: SAPs Transmit with Distinct Power	79
5.3.1	Primal-Dual Decomposition	80
5.3.2	Distributed Power Control Algorithm	81
5.3.3	Algorithm Implementation	83
5.4	Performance Evaluation	85
5.4.1	Channel Uncertainty	86
5.4.2	PU Protection	88
5.4.3	SU Performance	91
5.5	Summary	93
5.6	Appendix: Proof for Proposition 5.2	96
6	Conclusions and Future Works	98
6.1	Conclusions	98
6.2	Future Works	99
6.2.1	Cooperative Sensing with Reference-based Uncertainty	99
6.2.2	Joint Design of Sensing Parameters and Power Control	100
6.2.3	Asymmetry in Distribution Uncertainty	100
6.2.4	Heterogeneous Uncertainty Set	101
6.2.5	Distributed Implementation	102
	List of Publications	103
	References	105

List of Figures

2.1	Spectrum sensing as a binary hypothesis testing problem.	12
2.2	A typical sensing structure.	14
2.3	Network structure of cooperative spectrum sensing.	16
3.1	Quadratic curves for lower and upper bound problems.	29
3.2	False alarm probability with different decision threshold.	38
3.3	Actual false alarm probability under noise fluctuations.	39
3.4	Lower detection bounds with different moment uncertainties.	40
3.5	False alarm probability with different distance limits.	41
3.6	Lower detection bounds with different uncertainties.	42
4.1	Performance comparison between the nominal and robust designs. . . .	61
4.2	The values of projection on gradient direction decrease in each iteration.	62
4.3	Detection probabilities converge to the same level irrespective of different IDT settings.	63
4.4	Each SU's decision threshold converges to a common value under different IDT settings (and different SUs have different values).	63
4.5	Decision thresholds vary with their received signal strength.	64
5.1	Primary and secondary networks.	73
5.2	Non-stationary channel model with two states.	74
5.3	System structure for robust power control.	84
5.4	Empirical setting for distance limit.	87
5.5	Concavity of $u_k(\mathbf{p}_{-k}, p_k)$ and $I(\mathbf{p}_{-k}, p_k)$ w.r.t. p_k	89
5.6	PU protection with robust and nominal power control.	90
5.7	Robust power control vs. nominal power control.	92
5.8	The comparison between robust UTP and DTP methods.	94
5.9	Performance comparison as SAP 3 moves away from PBS.	95

List of Tables

1.1	Worldwide spectrum occupancy measurements.	3
4.1	Strategies given $G_i(t) < G_j(t) < 0$	56

Chapter 1

Introduction

Over nearly 100 years, the spectrum usage are managed in a centralized manner by government agencies, and this manner worked pretty well to avoid power races and control interference over the past years. As these government agencies can have better understanding of the spectrum profile, they can make the best use of different spectrum bands by allocating them for different purposes. For example, the low frequencies radio signals can travel long distance, thus these low spectrum bands are mostly used by national defence and broadcasting. Until recently, this centralized regulatory policy has allocated up to 95% of the spectrum to fixed license holders for exclusive usage [1].

However, nowadays the expansion of wireless technologies, e.g., cellular communications, wireless local area network (WLAN), and the rapid growth of wireless services (e.g., online video, game) evoke dramatic increase in the demand for radio spectrum, overwhelming the traditional fixed spectrum allocation strategy. As most of the spectrum has been assigned to license holders, the available spectrum has been increasingly depleting and becoming a costly resource, which raises an obstruction to the further development of new wireless technologies. Facing this situation, government agencies over different countries are in urgent need of new wireless technologies that allow spectrum sharing among different applications. In the following of this chapter, we first review some spectrum measurement reports from different countries and the potential technique that is proposed to improve spectrum efficiency. Within this background, we then identify our interests of research and methodologies. Our research contributions are presented in the following chapters.

1.1 Background and Motivations

In order to have a better understanding of how wireless system uses the allocated spectrum, the policy makers in different countries conducted independent spectrum surveys within their borders. Though most part of the spectrum has been allocated to license holders, worldwide spectrum measurement campaign [2–7] shows that the overall spectrum utilization is highly inefficient as the licensed channels remain idle for most of the time. Besides, the utilization is also severely unbalanced in different spectrum bands and geographical locations. Thus, a feasible solution to resolve this paradox does not lie in the discovery of new allocable spectrum resource, but rather a new paradigm that exploits the vacancies of such licensed spectrum.

The National Telecommunications and Information Administration (NTIA) of the United States has performed the first broadband spectrum surveys ranging from 108 MHz to 19.7 GHz over several metropolitan areas [2]. It is shown that the spectrum occupancy varies very much over geographic locations. Generally, spectrum occupancy is higher in costal cities than in midwestern cities on a band-by-band basis due to the large density of population in costal cities. Later in 2005, Mark A. McHenry *et.al.* [3] proposed long-term spectrum occupancy measurements ranging from 30 MHz to 3 GHz in Chicago. These measurements were taken in two days and used to identify which spectrum bands are under low utilization, thus can be exploited to develop new spectrum access technologies. The measurements report showed an averaged spectrum occupancy of 17.4%, while the utilization over different spectrum bands ranged from 70.9% to nearly empty in the 1240 MHz to 1850 MHz frequency band. Geographical variations of the spectrum utilization were also revealed in this report and a comparison showed that Chicago’s average spectrum utilization exceeded New York’s by roughly 1/3. Besides, these measurements also showed the existence of “white space” over time domain. For example, the usage in Industrial, Scientific and Medical (ISM) band was declined between roughly midnight and 6 am.

Similar spectrum measurements [4] were performed in four different locations in Southern China. This measurement report did not present numerical numbers regarding the occupancy in different spectrum bands. Instead, the spectrum occupancy

Table 1.1: Worldwide spectrum occupancy measurements.

Country	Measure period	Spectrum range(MHz)	Average usage	Candidate bands for spectrum sharing
USA [3]	2 days	30-3000	17.4%	1240-1850 MHz
Spain [8]	2 days	75-3000	22.57%	GSM 900 and UMTS
Singapore [5]	12 days	80-5850	4.54%	TV Band V
New Zealand [6]	12 weeks	806-2750	6.21% (outdoor) 5.72% (indoor)	Directional fixed link
Vietnam [7]	4 month	20-3000	11.19-13.74%	Analog TV band (470-806MHz)

was studied by two probability distribution functions obtained through a least-square regression analysis of the measured data. The first distribution described the length of time of consecutive spectrum occupancy, while the other distribution referred to the length of time interval between two consecutive “white spaces” in time domain. Both of them were shown to follow exponential-like distributions. This report also studied the temporal, spectral, and spatial correlations of spectrum occupancy, based on which a pattern mining algorithm was proposed to predict the spectrum occupancy with high accuracy. To have an overall understanding, we summarize worldwide spectrum occupancy measurements as in Table 1.1. It is a common phenomenon that the spectrum occupancy is quite low in a very large spectrum range. However, the same spectrum band may have different usage levels in different countries. In the last column of Table 1.1, we list the most appropriate candidate bands for developing new spectrum sharing technologies in different countries.

These strong evidences of inefficient spectrum utilization urge the industry and research communities to open the licensed bands for opportunistic usage by unlicensed users. Thus, the concept of dynamic spectrum access [9] is proposed as a solution to this inefficiency. A promising technology enabling dynamic spectrum access is the cognitive radio, which was first proposed by Dr. Joseph Mitola III [10] in 1999. By the definition of Simon Haykin [11], “*cognitive radio is an intelligent wireless communication system that can learn from its surrounding environment and adapt to it by changing its operating parameters, such as transmit power, carrier frequency, and modulation type.*” Specifically, in face of spectrum inefficiency, the cognitive radios

(i.e., secondary users, SUs) are allowed to access licensed bands as long as they do not generate harmful interference to the licensed users (i.e., primary users, PUs).

1.2 Research Scope and Problems

To realize the vision of cognitive radio, spectrum identification and exploitation are two main parts of the system design. Opportunity identification realizes the function of accurately identifying and tracking of channel status in different spectral bands, time and space. After identifying the channel status, a coordination mechanism is required to regulate how SUs behave such that the explored spectrum opportunities can be efficiently exploited by multiple competing SUs. Thus, opportunity exploitation refers to a channel access policy, e.g., whether to access or not, which channel to transmit, and how to set operating parameters.

1.2.1 System Robustness

Our research works span the algorithm design for spectrum identification and exploitation with a focus on the system robustness or reliability when it is implemented under non-ideal practical environment. System robustness is of great interest to the industrial community, and is used to describe how a decision variable affects the system performance when some design parameters are subject to small perturbation. An optimal solution to a nominal problem (original design problem that overlooks the parameter perturbation) may easily bring the actual performance down to an unacceptable level in practice. However, if we can model such perturbation properly and present a robust design, we will obtain much stable performance, which is more preferable in practice.

1.2.2 Spectrum Identification

Spectrum identification can be acquired either through the information exchange between PUs and SUs, or via SUs' spectrum sensing and prediction. The first method requires the cooperation between PUs and SUs, which introduces extra signaling overhead and sometimes may be infeasible. The second method does not rely on the information provided by PUs, but demands accurate sensing algorithms and sometimes

collaboration among SUs. In our research, we mainly focus on the robust design of spectrum sensing methods. Spectrum sensing is usually formulated as a hypothesis testing problem, where an SU decides whether the PU is present by comparing a function of the measured channel samples with a pre-designed decision threshold.

The analysis of sensing performance, system throughput, and interference level of cognitive radio networks in the literature often assumes that the signal distributions are known [12–14]. For example, the authors in [15–17] derived analytical expressions for the detection probability under different channel models. Unfortunately, knowing precise information regarding the primary signals’ probability distribution function is a very strong assumption that often does not hold in practice. For example, the mobility of wireless nodes may significantly change the signal distributions. Therefore, our work in this part aims to study the sensing performance with imperfect knowledge of signal distributions, and design the robust sensing structure in terms of the worst-case performance guarantee.

1.2.3 Spectrum Exploitation

Efficient spectrum exploitation requires the unlicensed SUs to limit their interference to the licensed PUs. Limiting interference to PUs on one hand is dictated by spectrum regulators to protect the license holder. On the other hand, it is also desired by SUs to maximize their overall performance as excessive interference usually causes transmission failures and requires additional spectrum opportunities for retransmission. Therefore, effective interference management is required at the SU side, and in general contains two aspects, i.e., interference awareness and interference mitigation. Interference awareness refers to the detection of potential active PUs and the estimation of interference at PU receivers. So far, most of the spectrum sensing algorithms focus on detecting PU transmitters with limited considerations in estimating the interference at PU receivers. Interference mitigation uses a coordination mechanism among SUs as well as the PUs in a centralized or decentralized manner, so as to mitigate mutual interference and improve SUs’ performance. It can be achieved in different ways. In a single channel case, power control can be employed to maximize

SUs' overall performance subject to an interference constraint at the PU side. While in a multi-channel case, individual SU can first choose an operating channel then a power level to avoid severe interference with other SUs and active PUs on the same channel.

In this work, we narrow down the spectrum exploitation as a power control problem of SUs, which first requires precise channel information to be aware of the interference at PU receivers. However, in general, there is no regular information exchange between PUs and SUs, which implies that SUs are unable to obtain up-to-date channel information at the PU side. Besides, the small-scale fading, in addition to shadowing, brings great uncertainty in the estimation of channel gain. Therefore, it is technically challenging for SUs to accurately control the interference at PU receivers through power control.

1.3 Research Methodology

The optimal design of spectrum sensing or power control algorithms requires the optimizations of various decision variables. We basically rely on the techniques of convex optimization in our study. As we have explained in last section, an optimal design may be not always desirable since it causes severe performance degradation in face of parameter perturbations [18, 19]. Thus, we resort to robust optimization techniques to enhance system robustness [18–22]. The basic procedure to study a robust optimization problem contains the following steps.

- (i) Identify the source of uncertainties of the system parameter. In wireless communications, the parameter uncertainty may be due to the stochastic nature of wireless channels. For example, the channel gain is time-varying and can be described by a random variable. Another inevitable source of uncertainty is caused by the channel estimation errors due to limited sensing capacity in channel measurements. For example, the estimation of channel gain requires sampling of the received signals, which is subject to the finite precision or quantization errors [20].

- (ii) Model the range of uncertain parameter. In order to study the effect of uncertain parameter on system performance, we need to quantify the range of uncertain parameter by some mathematical form, namely, the uncertainty set containing all possible instances of the random parameter. Note that the estimation errors are usually bounded, therefore the simplest uncertainty set can be defined by a box constraint, in which each term of the uncertain parameter has an upper and lower bounds. More general uncertainty set can be defined in terms of a vector norm or in an ellipsoid [19, 21].
- (iii) Formulate robust counterpart of the original design problem. When the uncertainty set of system parameters is explicitly considered and added as a constraint in the original design problem, we obtain the robust counterpart of a nominal optimization problem. Considering the worst-case performance, the robust counterpart aims to re-design the decision variables such that system performance is maximized even if the uncertain parameter takes a worst-case instance in the uncertainty set.
- (iv) Solve the robust counterpart by the methods for convex optimization. Generally, the robust counterpart will be harder than the nominal optimization problem due to the presence of extra constraint. Usually, proper transformation is required to shape the robust counterpart into a convex optimization [19]. For more general case, convex transformation is not available, thus specific approximate or heuristic algorithms should be proposed on a case-by-case basis [22].

For both of the research problems proposed in Section 1.2, we identify the source of uncertainty as the fluctuations of channel gain between SUs and PUs, which possesses a probability distribution function (PDF) but unknown to SUs. Therefore, we have to define an uncertainty set that contains all possible instances of the PDF in order to characterize the channel fluctuation in worse-case. The challenge is that, to secure a PDF, we need to assign values to infinite many variables, each of which corresponds to a point on the curve of PDF.

In our research, we propose some new methods to define the uncertainty set for signal's PDF, namely, distribution uncertainty. Based on this uncertainty model, we present the robust design of spectrum sensing as a max-min problem. The minimization is over the distribution uncertainty set and the maximization is over the decision threshold. Jointly, they achieve a robust design that provides optimal sensing performance under worst-case channel fluctuation. For the power control problem, we consider the same uncertainty model in SUs' channel estimation, and apply it to the robust design of power control problem, which provides guaranteed protection for PUs in a time-varying channel environment.

1.4 Contributions and Organization

Our contribution first lies in the introduction of distribution uncertainty into the design of spectrum sensing and power control algorithms. Most of the existing works deal with channel uncertainty by modeling the channel gain as a combination of deterministic and uncertain components. In general, these works fall into two approaches, i.e., stochastic and the worst-case approaches. A stochastic approach assumes the uncertain component to follow a known distribution function [12–14], which is often unavailable or very pricey to obtain in practice. While a worst-case approach restricts the fluctuations of the uncertain component to be within a bounded and convex set [23, 24]. Compared with the stochastic approach and other non-robust methods that assume full knowledge of the channel gain, the worst-case approach can provide the highest PU protection level, however, with the price of conservative performance for SU transmissions. Our proposed distribution uncertainty combines these two methods, providing a trade-off study between conservativeness and practicability.

Secondly, we characterize the distribution uncertainty by two different models, i.e., moment-based uncertainty and reference-based uncertainty. All these models define convex uncertainty sets for the unknown distribution for the preference of an analytical study. In the moment-based uncertainty, we characterize the distribution uncertainty by a series of inequality constraints on the moment statistics, e.g., sample

mean and variance estimates of the received signals. In the reference-based uncertainty, we exploit the distribution information embedded in the historical data, from which we extract a reference distribution by goodness-of-fit test. Then we propose a new uncertainty model in which the actual signal distribution is allowed to be different from the reference distribution, and we qualify the difference in terms of a probabilistic distance measure. Our research shows that this reference-based uncertainty model is more flexible in characterizing signal's fluctuations than the moment-based uncertainty model.

Thirdly, we present analytical studies on the sensing performance based on these distribution uncertainty models. For single-user case, we study the sensing performance in terms of lower and upper detection bounds. When moment-based uncertainty is considered, we show that the detection bounds can be obtained from equivalent semi-definite programs (SDPs) that are solvable by interior point algorithm. For the reference-based uncertainty, we propose two iterative methods to study the false alarm probability and detection bounds, respectively. For multi-user case, we focus on the moment-based uncertainty model, formulate the robust design as a max-min problem, and propose an iterative algorithm that decomposes the problem into a series of semi-definite programs, which can be solved easily by conic optimization tools.

Finally, we apply the reference-based uncertainty model to study the robust power control problem, and develop two iterative algorithms to search for the optimal transmit power for different network settings. For the case where all SUs transmit with the same power, we update SUs' transmit power by a bisection method until the interference requirement is met at PU receivers. For the second case where each SU may choose different transmit power, this problem becomes non-convex and we develop a heuristic algorithm that allows each SU to iteratively maximize its own utility. Simulation results show that the heuristic algorithm for the second case provides better quality of service (QoS) for SUs than the bisection method in the first case, however, at the cost of higher computational complexity.

The following chapters are organized as follows. Chapter 2 presents a literature review on related topics. Chapter 3 introduces the moment-based and reference-based distribution uncertainty models, and studies the robust sensing performance

for single-user case under these uncertainty models, respectively. Chapter 4 extends the study of sensing performance into multi-user case where the primary signal is subject to a moment-based distribution uncertainty model. Chapter 5 employs the reference-based distribution uncertainty model to study SUs' power control problem where the channel information is partially observed by overhearing the feedbacks from PU receivers. Finally we summarize our research and propose some future research directions in Chapter 6.

Chapter 2

Literature Review

Spectrum sensing and power control are generally physical layer issues. Antenna set mounted in transceiver samples the received signal for some time when it is tuned to certain spectrum band, and takes the sampling results as input to a decision module where detection and decision theory are applied. Based on the sensing results, SUs tune their transmit power levels so as to maximize SUs' aggregate throughput without degrading PUs' performance too much. To achieve this objective, every steps in the sensing and power control process should be properly designed to make the whole process working in best condition. In this chapter, we present an overview of the main research results in optimizing the sensing parameters, for example, channel sampling rate, sensing time, and decision threshold. Noting the connections between spectrum sensing and power control, we also review some existing works presenting jointly optimizations between sensing parameters and power control. In the final, we analyze the sources of channel uncertainties which cause the power control problem a rather challenging issue, and present an overview about the robust power control methods in literature.

2.1 Optimization of Sensing Parameters

There are a lot of sensing techniques (e.g., energy-based [25, 26] or feature-based [27–30] detectors), but they all face a similar problem of making a decision between two hypotheses: absence or presence of the PU signals. Considering energy-based

detector, the sensing problem is modeled as a binary hypothesis testing as follows:

$$\begin{aligned} H_0 : \quad \xi(k) &= n(k); \\ H_1 : \quad \xi(k) &= s(k) + n(k); \end{aligned}$$

where $n(k)$ and $s(k)$ denote the noise and PU signals, respectively. $\xi(k)$ is the received signal at sensing instance k . In the optimization of the decision threshold λ for the energy detector, there are two conflicting goals: maximizing the detection probability P_d and minimizing the false alarm probability P_f (i.e., the shaped areas in Figure 2.1). For a single user case, if $\xi(k)$ exceeds the decision threshold λ , the SU's detector will return 1 indicating the presence of PU on the sensing channel (i.e., hypothesis H_1), otherwise return 0 indicating an idle channel (i.e., hypothesis H_0).

Figure 2.1: Spectrum sensing as a binary hypothesis testing problem.

2.1.1 Sensing Decision Threshold

Based on the hypothesis testing model, a critical design parameter is the decision threshold λ . If the signal distributions of $s(k)$ and $n(k)$ are Gaussian and known in advance, two major performance metrics P_d and P_f can be represented in a closed-form [31, 32] as follows:

$$P_f(\lambda, t_s) = Q\left(\left(\frac{\lambda}{\sigma_0^2} - 1\right) \sqrt{f_s t_s}\right) \quad (2.1)$$

$$P_d(\lambda, t_s) = Q\left(\left(\frac{\lambda}{\sigma_0^2} - \gamma - 1\right) \sqrt{\frac{f_s t_s}{2\gamma + 1}}\right) \quad (2.2)$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{t^2}{2}\right) dt$. σ_0^2 denotes the variance of noise signal, γ represents the received signal to noise ratio (SNR), t_s is the sensing time, and f_s is

the channel sampling frequency. Given a maximum false alarm probability α , we can obtain the lower bound of decision threshold λ_{min} from (2.1). Similarly, if there is a minimum target detection probability, (2.2) defines the upper bound of decision threshold λ_{max} . Assuming $\lambda_{min} \leq \lambda_{max}$, a feasible choice of decision threshold lies in the set $[\lambda_{min}, \lambda_{max}]$. For example, Moghimi *et al.* [33] choose the decision threshold that minimizes the false alarm probability in a Non-Gaussian noise channel while the detection probability is maintained at a target level. Some other works in literature [34, 35] aim to minimize the total error probability, i.e., false alarm probability P_f and the probability of missed detection $P_m = 1 - P_d$. For example, Joshi *et al.* [35] minimize a weighted combination of P_m and P_f through the adaptation of decision threshold λ , considering the variance σ_0^2 noise signal changing over time. The authors formulate a constrained optimization problem and solve it using the Lagrange method. The adaptation of decision threshold follows a gradient descent direction to minimize the total error probability.

The choice of decision threshold also affects the average transmission rate of secondary network. Smaller P_f implies larger transmission opportunities, however it also leads to higher miss detection probability which increases the chance of collision with PUs. A collision may incur retransmissions of both PUs and SUs, which actually decreases the average transmission rate. Therefore, the authors in [36] consider maximizing SU's average transmission rate subject to a constraint on PUs' interference level. Note that the received signal to interference plus noise ratio (SINR) at SU may change over time, the authors propose an adaptive threshold control based on an intuitive assumption that the decision threshold is a linear function of the SINR. Note that SUs having different distance with PU may introduce different level of interference, the authors in [37] define a new performance metric, namely, the probability of interference with PUs. The optimal design of decision threshold aims to maximize the detection probability subject to an interference probability constraint. Simulations show that this new metric can potentially improve SU's spectrum utilization.

2.1.2 Sensing Time and Tradeoffs

We also note that P_f and P_d are functions of the sensing time. Combining (2.1) and (2.2), we obtain the sensing time tradeoff between false alarm probability P_f and detection probability P_d as follows:

$$P_f = Q\left(\sqrt{2\gamma+1}Q^{-1}(\bar{P}_d) + \gamma\sqrt{f_s t_s}\right) \quad (2.3)$$

$$P_d = Q\left(\frac{1}{\sqrt{2\gamma+1}}\left(Q^{-1}(\bar{P}_f) - \gamma\sqrt{f_s t_s}\right)\right) \quad (2.4)$$

A typical sensing structure is given in Figure 2.2. The channel is divided into time frames with equal time length T . At the beginning of each frame, a time slot t_s is allocated for spectrum sensing and the left part $T - t_s$ is for SU' data transmission. If SU takes more time on spectrum sensing, it will have better sensing results as indicated by (2.1) and (2.2), however, less transmission time in the other hand.

Figure 2.2: A typical sensing structure.

This tradeoff between sensing time and throughput performance has been broadly studied in literature [31, 38–44]. Liang *et al.* [31] design the optimal sensing time to maximize SUs' achievable throughput subject to a target detection probability constraint for the protection of PUs. The sensing-throughput tradeoff problem is approximated as a convex optimization and shown to possess one optimal sensing time which yields the highest throughput for the secondary network. Sensing time optimization is further brought to multi-user case [38] and a joint study with transmit power control to maximize SUs' throughput while keeping the interference power at PU receivers under certain threshold [39, 40]. Consider the stochastic nature of wireless channels, another framework has been studied in [41, 42] where the sensing tradeoff optimization relates to the outage probability under different fading channels. The authors in [41] give a closed-form expression for the outage probability of SU's

transmission and aim to minimize it with a target detection probability constraint over Rayleigh fading channel. Similarly, the authors in [42] optimize the sensing time subject to an outage constraint on the miss detection probability over Nakagami-fading channel. The results show that different fading parameter of the Nakagami channel leads to different sensing time. To overcome the sensing throughput tradeoff, Stotas *et al.* [43, 44] introduce a novel receiver and frame structure that allow the spectrum sensing and data transmission are performed at the same time. In this structure, sensing time is not a critical parameter as SU can sense the channel during the whole frame length. Alternatively, the authors propose an algorithm that obtains the optimal target detection probability and transmit power.

2.2 Multi-user Cooperative Sensing

Due to the channel fading effects, multi-user cooperative spectrum sensing is proposed to improve the detection performance [15, 45, 46]. A basic network structure of cooperative spectrum sensing is shown in Figure 2.3. Information exchange between cooperative SUs can help each other better understand the environment. Usually, there is a center for decision or data fusion among the cooperating SUs. The optimal design of decision fusion (or hard combination) rules such as “OR”, “AND”, and “K-out-of-N” is discussed in [47–49]. The authors in [47] consider minimizing the total error rate by choosing the optimal combining number K in the “K-out-of-N” rule. A fast sensing algorithm is also proposed to determine the minimum number of cooperative users N to achieve the target error probability bound. The authors in [48] propose an efficient technique to implement the “OR” rule in the fusion center. This technique reduces the time overhead in reporting channel by allowing all SUs to send one-bit local decisions simultaneously to the fusion center, which makes a final decision depending on whether the aggregated signal in the reporting channel exceeds a predetermined decision threshold.

The optimization of data fusion (or soft combination) schemes have been studied in [50–52], where an optimal combining weight is multiplied with each SU’s preprocessed sensing data before summing up at the fusion center. Based on a simple intu-

Figure 2.3: Network structure of cooperative spectrum sensing.

ition that a larger weight should be assigned to a higher signal strength, the authors in [50, 52] propose an optimal linear combination scheme to minimize the interference to PUs. The authors formulate a nonconvex optimization problem and propose a heuristic algorithm to determine the optimal combining weight for each SU. This linear combination scheme has assumed that all SUs' channel sensing information are synchronized. However, as different SUs may have different sensing schedules, this may incur performance loss for the synchronized linear combination scheme. Therefore, the authors in [51] propose a probability based combination scheme, which minimizes the total Bayesian risk in sensing and accounts the time offsets in SUs' sensing observations into the calculation of combining weights.

Multi-bit data combining is proposed in [49], which bridges the gap between hard and soft combination schemes. By optimizing the quantization level and local decision thresholds, the sensing performance of a multi-bit combining rule is significantly better than that of the one-bit hard combination rules, and converges fast than that of the soft combination rules as the quantization level increases. Some other works [53–55] present intensive studies on the performance comparison between decision and data

fusion. Signal correlation between different SUs is also a critical factor that affects the sensing performance. Empirical study suggests an exponential-type correction function for spatial shadowing [15, 56]. To counteract the side-effect of correlation in data fusion, a user selection algorithm is proposed in [57] to adaptively choose the uncorrelated SUs in cooperation based on a spatial correlation coefficient. The impact of correlation on decision fusion (i.e., “K-out-of-N” rule) is studied in [58] where the authors show that, for each correlation coefficient, there is an optimal value for K minimizing the total sensing errors and obtained in a genetic algorithm.

2.3 Joint Sensing and Power Control

Spectrum sensing and power control of cognitive radios are inherently connected. In one aspect, the sensing information can help to determine the optimal power allocation in different spectrum bands or time slots. In another aspect, SUs’ power budget poses constraints on the power allocation in sensing and transmission periods, therefore helps to optimize the sensing parameters. For example, the authors in [59] characterize the impacts of SUs’ transmit power control on the occurrence of spectrum opportunities and the reliability of spectrum sensing.

Lots of works in literature are devoted to jointly optimize the sensing parameters and the power control mechanism. The authors in [60] propose to joint optimize the inter-sensing duration and power control for SUs. According to PUs’ channel utilization statistics and current sensing result, SUs optimize the inter-sensing duration to capture the recurrence of spectrum opportunity (if current sensing indicates presence of PUs) or avoid undue collisions (if current sensing indicates absence of PUs). SUs’ power control is also considered in the inter-sensing duration optimization such that the interference power at PU receivers is below a threshold value in every sensing-transmission period. Within a sensing-transmission period, there is also a tradeoff between sensing time and transmission time. Therefore, the authors in [61] propose a joint optimization of the sensing-time and power control algorithm, which is an extension of previous work with fixed transmit power [38, 39]. There are also lots of works in literature optimizing jointly the sensing channel selection and power control

algorithm, alternatively, the power allocation problem over multiple primary channels or sub-carriers. Hoang *et al.* [62] consider multiple secondary base stations serving fixed-location subscribers with known channel information. The objective is to maximize the total number of active subscribers by the channel assignment and power control among multiple secondary base stations. While in [63], the authors consider time-varying channel conditions and develop an online algorithm that dynamically learns the channel statistics, based on which the joint channel selection and power control is performed under the constraints of average power budget and maximum collision probability requirement.

As an independent part, power control has also been extensively studied and there are some algorithms working very well under existing wireless networks, e.g., cellular network [64]. However, due to highly dynamic spectrum environment and strict interference constraint to PUs, it is very hard to transplant these algorithms into cognitive radio networks. Therefore, the research community needs to propose new power control methods specific for cognitive radio networks. One power control criterion is to maximize SUs' performance while satisfying a required level of protection for PUs. For instance, some research works embody the protection for PUs by the constraint on interference power at PU receivers. In [65, 66], maximizing SUs' throughput is formulated in a dynamic programming problem with constraints on SUs' total energy budget and PUs' interference power. Moreover, in [67, 68], SUs' SINR and buffer length are considered as the performance metric in the power control problem, respectively. More practical power control algorithms are proposed by [69, 70], in which the SUs can assess their interference by listening to PUs' feedback channel, and adjust their transmit power to keep the interference power at PU receivers below predefined threshold value. The authors in [71] consider a similar approach as in [69, 70] to maximize SUs aggregate throughput, while the protection for PUs is to guarantee PUs' queue stability.

2.4 Robust Cognitive Power Control

Most of the studies in literature have assumed that complete channel information is available to SUs (e.g., channel gain between SU transceivers and PU transceivers). However, it is unlikely for the SUs to have precise channel information when PUs are not obliged to join a full cooperation with SUs and provide PU side channel information to SUs. Though [69–71] propose the power control method by listening to PUs’ feedback channel, it may be unreliable for real-time power control due to limited observations of feedback from PU receivers. Specifically, PU receivers may send feedback (e.g., ACK packets) sporadically after receiving a bulk of data streaming, thus channel estimation in this way may not capture the changes of channel characteristics timely. If power control is based on out-of-date channel estimates, SUs cannot know the actual interference at the PU receivers, easily leading to violations of PUs’ interference constraints, especially when channel gain changes over time [72, 73].

Without information exchange between SUs and PUs, the SUs have to solely rely on their spectrum sensing capability. However, due to hardware constraint and sensing overhead, spectrum sensing is also unreliable and sensing errors become inevitable [74, 75]. Konrad *et al.* demonstrate in [76] that channel characteristics exhibit time-varying effects in a long time period, which is caused by the change of *physical channel conditions*, e.g., a light-of-sight (LOS) path between transceivers may exist for some time and disappear when the path is blocked temporarily. To make accurate estimation, the physical channel conditions should remain stable for a sufficient time to provide enough channel samples. Let *static period* denote a small time duration in which SUs remain stationary and physical channel conditions do not change. However, such a static period may be very short due to user mobility, resulting in very limited data samples for channel estimation. Moreover, SUs are generally unable to detect when a static period starts or how long it lasts, therefore, the channel samples may stem from different static periods, which gives inaccurate channel information in different static periods.

To avoid excessive interference caused by channel uncertainty, some power control methods model the fluctuations of channel gain as a combination of deterministic

and uncertain components. In general, these methods fall into two approaches, i.e., stochastic and the worst-case robust approaches. A *stochastic approach* assumes the uncertain component to follow a known distribution function, which usually leads to chance constraints in the power control problem. For a practical implementation, the chance constraints are further transformed into convex forms. For example, Zheng *et al.* in [77] consider the uncertain component as Gaussian noise and convert interference outage probability into a generalized Marcum's Q function [78]. In a similar way, Dall'Anese *et al.* in [79] approximate both PUs' aggregate interference power (AIP) and SUs' SINRs by log-normal distributed random variables, and then solve the resulting problem via sequential geometric programming. The challenge of a stochastic approach lies in the assumption of a known distribution function, which is often unavailable, or is very pricey to obtain in practice. Hence, some other works employ a *worst-case robust approach* that merely restricts the uncertain component to be fluctuating within a bounded and convex set, e.g., the ellipsoid set in [23, 80]. Zheng *et al.* in [23] aim to maximize SUs' SINRs subject to a power budget and interference power constraint at PU receivers. The power control problem is approximated into a semi-definite program (SDP) based on rank relaxation which can be solved by two randomized algorithms proposed in [24]. While Sun *et al.* in [80] minimize the SUs' total power consumption under constraints on SUs' QoS and PUs' interference power. The power control problem is transformed into a second order cone programming and solved by the Lagrangian dual method in a distributed way. In [72, 73], the authors describe the channel uncertainty by general norm and polyhedron model, which help to transform the interference power constraints into convex forms. Compared with the stochastic approach and other non-robust methods that assume full knowledge of the channel gain, the worst-case robust approach can provide the highest PU protection level, however, with the price of conservative performance for SU transmissions.

Chapter 3

Robust Spectrum Sensing for Single-User

Spectrum sensing aims to determine the presence of PUs based on channel characteristics. It is usually formulated as a hypothesis testing problem, where an SU decides whether the PU is present by comparing a function of the measured channel samples with a pre-designed decision threshold. The analysis of sensing performance, system throughput, and interference level of cognitive radio networks in the literature often assumes that the signal distributions are known [12–17]. The challenging issue is that it is very hard and pricey to know information regarding the primary signals' probability distribution in practice.

In this chapter, we consider both the uncertainties of noise and PUs' signals in single-user spectrum sensing. Considering different channel conditions, we mainly define two types of uncertainty models, i.e., moment-based and reference-based uncertainty models. In the first part, we consider the moment-based uncertainty model that is based on moment statistics, e.g., sample mean and variance estimates of the received signals. In the second part, we exploit the distribution information embedded in the historical data, from which we extract a reference distribution by goodness of fit test. Then we propose a new uncertainty model in which the actual signal distribution is allowed to fluctuate around such reference distribution, and qualify their discrepancy in terms of a probabilistic distance measure.

To study the detection performance in terms of lower and upper detection bounds, we formulate a robust optimization problem with above defined uncertainty models,

respectively. For the moment-based uncertainty model, we show that the detection bounds can be obtained from equivalent SDP that are solvable by interior point algorithm, and we derive an analytical formula, which can help SUs to quick assess the detection bounds instead of solving a SDP. For the reference-based uncertainty model, we propose two iterative methods to obtain the false alarm probability and detection bounds, respectively.

The rest of this chapter 3 is organized as follows. Section 3.1 describes the spectrum sensing model and different signal uncertainties. In Section 3.2, we study the detection performance with moment-based uncertainties, while in Section 3.3, we extend the study into reference-based uncertainty model. Section 3.4 gives some numerical results and Section 3.5 summarizes the main contributions of this chapter.

3.1 Performance Metrics and Uncertainty Models

We consider multiple SUs and PUs sharing the same spectrum in a cognitive radio network. Assuming no information exchange between PUs and SUs, each SU needs to perform spectrum sensing independently. The objective of spectrum sensing is to detect the presence of primary signals on the primary channel, and we model this problem as a binary hypothesis testing. Specifically, if the received signal ξ at an SU's receiver exceeds a decision threshold λ , the SU's detector will return 1 indicating the presence of PU on the sensing channel (i.e., hypothesis H_1), otherwise return 0 indicating an idle channel (i.e., hypothesis H_0). Thus the decision results can be viewed as drawn from a decision function $h(\xi, \lambda) = \mathbf{1}(\xi > \lambda)$ where indicator $\mathbf{1}(A)$ equals 1 if event A is true and 0 otherwise. Given this decision function, we will study the tradeoff between detection probability Q_d and false alarm probability Q_f . Let $f_0(\xi)$ be the distribution function of the received signal at an SU's receiver when PUs are absent, then the false alarm probability is given by $Q_f = \mathbb{E}_{f_0}[h(\xi, \lambda)] = \int_{\xi \in S} h(\xi, \lambda) f_0(\xi) d\xi$, where S denotes the set of all possible values of the received signal strength. When PUs are present, the received signal ξ will exhibit a different distribution function $f_1(\xi)$, and we express the detection probability as $Q_d = \mathbb{E}_{f_1}[h(\xi, \lambda)] = \int_{\xi \in S} h(\xi, \lambda) f_1(\xi) d\xi$.

Then our task is to maximize the detection probability Q_d subject to a false alarm probability requirement, i.e., $Q_f \leq \alpha$. From the above expressions, we note that the knowledge about two distribution functions $f_0(\xi)$ and $f_1(\xi)$ are crucial for an analytical study of the detection performance. However, due to the stochastic nature of wireless channels, it is usually impossible to obtain exact information about these distributions. That's because, distribution $f_0(\xi)$ depends on environment noise and interference from other active users, which is usually time-varying. Besides, multiple PUs may have different transmission techniques and traffic patterns, therefore, the signal distribution $f_1(\xi)$ may also be highly dynamic in terms of the statistical information. All these factors make it very complicated to model these distributions in closed-forms without significant simplifications, and if possible, it is also computational demanding to analyze system performance such as throughput and sensing overhead. Besides, a deterministic model of the primary signals' distribution is usually vulnerable to mismatch with the real channel condition. For example, the received signals exhibit different distributions depending on whether there is a line-of-sight between transmitter and receiver.

Researchers have been trying to design robust sensing strategies that do not rely on parametric distributions, and are tolerable to signal fluctuations. The authors in [81] proposed a nonparametric correlation detector, which does not require closed-form distribution functions for noise and PUs' signals. However, it still requires extra knowledge about the cyclic frequencies of PUs' signals. Nonparametric detection methods were also proposed based on goodness of fit test, which evaluates whether a sequence of signal samples is drawn from some known distribution. The authors in [82, 83] constructed the test statistic by empirical distribution of the signal samples, assuming a known noise distribution in the goodness of fit test. Higher order signal statistics, such as skewness and kurtosis, were also employed in the goodness of fit test [84, 85] to improve the robustness against environment variations. Though nonparametric goodness of fit test assumes no pre-knowledge regarding PUs' signal distribution, it needs to know the noise distribution in advance, and is usually difficult to study the detection performance analytically.

To analyze the impact of channel uncertainty on the sensing performance, we first need a quantitative way to characterize the channel uncertainty. In the following, we present some moment-based uncertainty models that are defined by signal's moment statistics, e.g., sample mean and variance estimates of the received signals.

- (i) Known Distribution with Uncertain Statistics (KDUS) model: In some cases, the environment noise is relatively stable and the channel condition is well informed, or we are targeting at specific primary signals. Therefore, we are able to know or assume some parametric distribution function either for noise or PUs' signals in advance [16]. However, we still need to estimate the distribution parameters (e.g., mean and variance) before we can fully characterize the signal distribution. Usually, parameter estimation is based on signal samples which bear limited information about the real signal distribution, thus estimation errors are inevitable, and we define the signal uncertainty in terms of these estimation errors:

$$\mathcal{U}_d(\mu_0, \sigma_0, \gamma_\mu, \gamma_\sigma) = \left\{ f(\xi) \in \mathcal{G} \mid \begin{array}{l} (\mathbb{E}_f[\xi] - \mu_0)^2 \leq \gamma_\mu \sigma_0^2 \\ |\mathbb{E}_f[(\xi - \mu_0)^2] - \sigma_0^2| \leq \gamma_\sigma \sigma_0^2 \end{array} \right\}, \quad (3.1)$$

where $\mathcal{G}$ is a set of distribution functions that belong to the same family, e.g., Gaussian distributions. Due to the fluctuations of wireless channel, the sample mean μ_0 and variance σ_0^2 are slightly different from the real distribution mean $\mu = \mathbb{E}_f[\xi]$ and variance $\sigma^2 = \mathbb{E}_f[(\xi - \mu)^2]$ ¹. These two inequalities in (3.1) describe how likely the distribution mean $\mathbb{E}[\xi]$ and the signal samples ξ are close to the sample estimate μ_0 , respectively. Parameters γ_μ and γ_σ regulate the uncertainty size and provide a way to evaluate the confidence in sample estimates μ_0 and σ_0^2 , respectively. A theoretical work in [86] demonstrated how to choose proper values for γ_μ and γ_σ based on the sample size.

- (ii) General Distribution with Exact Statistics (GDES) model: If pre-knowledge about PUs is not available in practice, it is often difficult for the SU to know precisely the primary signal's distribution function $f_1(\xi)$, as it is often time

¹Here we replace μ by its estimation μ_0 in the expression of distribution variance.

varying and experiences attenuation, shadowing, and multi-path fading before reaching the SU's receiver. Instead, estimates of the signal statistics may provide a practical way to study the signal's properties. If the signal detector is capable of sampling the primary channel for a long time, these sample estimates will be close to the real statistics with high confidence. In this chapter, we consider PUs' signal up to its second-order moment statistic and define the distribution uncertainty as follows:

$$\mathcal{U}_e(\mu_0, \sigma_0) = \left\{ f(\xi) \in \mathcal{M} \mid \begin{array}{l} \mathbb{E}_f[\xi] = \mu_0 \\ \mathbb{E}_f[(\xi - \mu_0)^2] = \sigma_0^2 \end{array} \right\}, \quad (3.2)$$

where $\mathcal{M}$ denotes the set of all possible distribution functions, which only requires $\int_{\xi \in S} f(\xi) d\xi = 1$ and $f(\xi) \geq 0$ for all $\xi \in S$. In this model, we relax the distribution function $f(\xi)$ to be any form of distributions, while restrict the signal statistics by a series of equality constraints.

- (iii) General Distribution with Uncertain Statistics (GDUS) model: In a more general setting, the signal distribution is uncertain and the signal statistics are also subject to estimation errors due to limited observations and measurement noise. For example, the spectrum environment may have been changed while the detector still maintains outdated statistical information. Even though we may update the signal statistics by the most recent channel measurements, we still lack the confidence to entirely rely on these estimates. Instead, it is reasonable and safe to assume that these sample estimates are within small ranges of their real values, and we can express the distribution uncertainty as follows:

$$\mathcal{U}_g(\mu_0, \sigma_0, \gamma_\mu, \gamma_\sigma) = \left\{ f(\xi) \in \mathcal{M} \mid \begin{array}{l} (\mathbb{E}_f[\xi] - \mu_0)^2 \leq \gamma_\mu \sigma_0^2 \\ |\mathbb{E}_f[(\xi - \mu_0)^2] - \sigma_0^2| \leq \gamma_\sigma \sigma_0^2 \end{array} \right\}. \quad (3.3)$$

Uncertainty (3.3) has the same inequality constraints as in set (3.1), their difference lies in the distribution sets $\mathcal{G}$ and $\mathcal{M}$. Note that $\mathcal{G}$ represents a specific distribution family, while set $\mathcal{M}$ does not pose any restrictions on its type. We also note that uncertainty (3.2) is a special case of set (3.3) when parameters γ_μ and γ_σ are set to zeros.

3.2 Analysis of the Moment-based Uncertainty

For the hypothesis testing problem, we first determine a decision threshold λ according to the false alarm probability requirement, i.e., $Q_f \leq \alpha$ where α is the maximum tolerable false alarm probability. Since Gaussian distribution is a good approximation for noise signal in both theory and practice, we can model its fluctuations by set (3.1). Mathematically, the choice of decision threshold λ should satisfy

$$Q_f^U(\lambda) \triangleq \max_{f_0(\xi)} \mathbb{E}_{f_0}[h(\xi, \lambda)] = \alpha, \quad (3.4)$$

where $f_0(\xi)$ is subject to set $\mathbf{U}_d(\mu_{n,0}, \sigma_{n,0}, \gamma_{n,\mu}, \gamma_{n,\sigma})$. Parameters $\mu_{n,0}$ and $\sigma_{n,0}^2$ denote the sample estimates of noise signal's distribution mean μ_n and variance σ_n^2 , respectively, and their estimation errors are regulated by $\gamma_{n,\mu}$ and $\gamma_{n,\sigma}$, respectively. Function $Q_f^U(\lambda)$ denotes the worst-case false alarm probability with fixed decision threshold λ . When $f_0(\xi)$ follows a Gaussian distribution, we have $Q_f(\lambda) = Q\left(\frac{\lambda - \mu_n}{\sigma_n}\right)$ where $Q(\cdot)$ is the complementary cumulative distribution function of standard Gaussian distribution, and the worst-case false alarm probability is given by

$$Q_f^U(\lambda) = \max_{\mu_n, \sigma_n} Q\left(\frac{\lambda - \mu_n}{\sigma_n}\right) = Q\left(\frac{\lambda - (\mu_{n,0} + \sigma_{n,0}\sqrt{\gamma_{n,\mu}})}{\sigma_{n,0}\sqrt{1 + \gamma_{n,\sigma}}}\right).$$

The second equation is due to the monotonic decreasing of $Q(\cdot)$, i.e., $\max_{\mu_n, \sigma_n} Q\left(\frac{\lambda - \mu_n}{\sigma_n}\right) = Q\left(\frac{\min(\lambda - \mu_n)}{\max \sigma_n}\right)$. We note that the worst-case fluctuation occurs when the noise signal takes the largest mean $\mu_{n,0} + \sigma_{n,0}\sqrt{\gamma_{n,\mu}}$ and variance $\sigma_{n,0}^2(1 + \gamma_{n,\sigma})$ in its uncertainty set. Therefore, the robust decision threshold should be set as

$$\lambda = \mu_{n,0} + \sigma_{n,0}\sqrt{\gamma_{n,\mu}} + \sigma_{n,0}\sqrt{1 + \gamma_{n,\sigma}}Q^{-1}(\alpha). \quad (3.5)$$

Given this decision threshold, we aim to analyze the sensing performance in terms of detection probability. However, deterministic study usually fails when distribution $f_1(\xi)$ is subject to uncertainties. Instead, we can study its lower and upper bounds, which can be used as important reference in real-world decision making. As we focus on the worst-case scenario, the following study intends to find a lower bound of detection probability. Similar approaches are also applicable for studying the upper bound of detection probability.

3.2.1 Lower Bound of Detection Probability

Considering the uncertainty in (3.1) for primary signal, the lower detection bound can be processed in the same way as for false alarm probability. For example, if $f_1(\xi)$ is a Gaussian distribution, the lower bound $Q_d^L(\lambda)$ is given by

$$Q_d^L(\lambda) = Q\left(\frac{\lambda - (\mu_{p,0} - \sigma_{p,0}\sqrt{\gamma_{p,\mu}})}{\sigma_{p,0}\sqrt{1 + \gamma_{p,\sigma}}}\right) \quad (3.6)$$

and it is achieved when $\mu_p = \mu_{p,0} - \sigma_{p,0}\sqrt{\gamma_{p,\mu}}$ and $\sigma_p = \sigma_{p,0}\sqrt{1 + \gamma_{p,\sigma}}$ where $\mu_{p,0}$ and $\sigma_{p,0}^2$ denotes the sample estimates of distribution mean μ_p and variance σ_p^2 of the received primary signal, respectively. As for uncertainty sets (3.2) and (3.3), we have the following optimization problem:

$$\min_{f_1(\xi)} \quad \mathbb{E}_{f_1}[h(\xi, \lambda)] \quad (3.7a)$$

$$s.t. \quad f_1(\xi) \in \mathbf{U}_g(\mu_{p,0}, \sigma_{p,0}, \gamma_{p,\mu}, \gamma_{p,\sigma}). \quad (3.7b)$$

Since (3.2) is just a special case of set (3.3), we only consider set $\mathbf{U}_g(\mu_{p,0}, \sigma_{p,0}, \gamma_{p,\mu}, \gamma_{p,\sigma})$ in problem (3.7a)-(3.7b). It intends to maximize the cumulative density lies within set $[\lambda, \infty)$, and the decision variable $f_1(\xi)$ is actually the density allocation on each point ξ in its feasible set subject to a series of linear inequality constraints defined in (3.3). Therefore, problem (3.7a)-(3.7b) can be viewed as a linear program with infinite number of decision variables (i.e., densities on each point $\xi \in S$) and finite number of constraints. A straightforward way to solve this problem is to approximate the feasible set S by a finite set of sample points. Then we only consider allocating probability mass at each of these discrete points. Since simplex method for linear program is very efficient in practice, we can draw a large samples such that the granularity is small enough to ensure desired accuracy. With extreme large samples, this results from linear approximation will be trustworthy and considered as a benchmark for other solution methods.

3.2.2 Equivalent Semi-definite Program

Besides the finite linear approximation, we try to solve the above problem by an accurate as well as efficient method². Note that problem (3.7a)-(3.7b) is a semi-

²We try to transform the problem into a convex form that can be solved efficiently by existing optimization toolbox.

infinite program as the functional variable $f_1(\xi)$ is actually a vector with infinite length. The authors in [86] proposed a method that can transform such semi-infinite program into a convex semi-definite program, which help us determine the detection performance by existing optimization tools.

For the lower detection bound, the worst-case distribution is expected to have a larger variance, thus we consider $\mathbb{E}[(\xi - \mu_{p,0})^2] \leq (1 + \gamma_{p,\sigma})\sigma_{p,0}^2$ in constraint (3.7b). By the Lagrangian method, we have its Lagrangian function as follows

$$\begin{aligned} \Lambda(f_1, \eta, Z, \nu) = & \max_Z \mathbb{E}[h(\xi, \lambda) + \eta - 2(\omega + \mu_{p,0}\nu)\xi + \nu\xi^2] + 2\mu_{p,0}\omega \\ & + [\mu_{p,0}^2 - (1 + \gamma_{p,\sigma})\sigma_{p,0}^2] \nu - \eta - \gamma_{p,\mu}s - \sigma_{p,0}^2\kappa, \end{aligned} \quad (3.8)$$

where η , Z , and ν are the Lagrangian multipliers associated with different moment constraints. Details are given in the Appendix 3.6. In a Lagrangian method, we first need to minimize $\Lambda(f_1, \eta, Z, \nu)$ in terms of $f_1(\xi)$ and then maximize it over the Lagrangian multipliers. However, we note that the minimization of $\Lambda(f_1, \eta, Z, \nu)$ in terms of $f_1(\xi)$ easily becomes negative infinite if there exist some $\xi_e \in S$ such that $h(\xi_e, \lambda) + \eta - 2(\omega + \mu_{p,0}\nu)\xi_e + \nu\xi_e^2 < 0$, i.e., we can set $f_1(\xi_e) = +\infty$ and 0 elsewhere. To make the minimization of $\Lambda(f_1, \eta, Z, \nu)$ bounded, we therefore have the following condition for any $\xi \in S$ when solving the primal problem

$$\nu\xi^2 - 2(\omega + \mu_{p,0}\nu)\xi + \eta \geq -h(\xi, \lambda). \quad (3.9)$$

The LHS of (3.9) is a polynomial of degree two, which defines a quadratic curve $c(\nu, \omega, \eta) = \nu\xi^2 - 2(\omega + \mu_{p,0}\nu)\xi + \eta$ and it should dominate a step function $-h(\xi, \lambda)$, i.e., i.e., $c(\nu, \omega, \eta) \geq -h(\xi, \lambda)$ for any $\xi \in S$. Figure 3.1 plots the relation between the quadratic curve and the step function. We may have different curves (e.g., curves c_1 and c_2) when the coefficients (ν, ω, η) take different values. Therefore, condition (3.9) actually looks for a set of coefficients (ν, ω, η) for the quadratic curve such that it always dominates $-h(\xi, \lambda)$. For comparison, we also plot the curve (i.e., c_3) for the upper bound of detection probability, which requires $c(\nu, \omega, \eta)$ to dominate $h(\xi, \lambda)$. From this geometric interpretation, we pay attention to two points, i.e., the quadratic curve should be greater than 0 at $\xi = \lambda$ and greater than 1 at $\xi = \mu_{p,0} + \frac{\omega}{\nu}$. Then, we can transform problem (3.7a)-(3.7b) into a simpler form.

Figure 3.1: Quadratic curves for lower and upper bound problems.

Proposition 3.1 *Finding the lower detection bound subject to a distribution uncertainty $\mathbf{U}_g(\mu_{p,0}, \sigma_{p,0}, \gamma_{p,\mu}, \gamma_{p,\sigma})$ is equivalent to solving a semi-definite program as follows:*

$$\max_{\eta, \nu, \omega} \quad [\mu_{p,0}^2 - (1 + \gamma_{p,\sigma})\sigma_{p,0}^2]\nu + 2\mu_{p,0}\omega - \eta - 2\sigma_{p,0}|\omega|\sqrt{\gamma_{p,\mu}} \quad (3.10a)$$

$$s.t. \quad \nu\lambda^2 - 2(\nu\mu_{p,0} + \omega)\lambda + \eta \geq 0, \quad (3.10b)$$

$$(\mu_{p,0} - \lambda)\nu + \omega \geq 0, \quad (3.10c)$$

$$\begin{bmatrix} \eta + 1 & \mu_{p,0}\nu + \omega \\ \mu_{p,0}\nu + \omega & \nu \end{bmatrix} \succeq 0. \quad (3.10d)$$

The proof for Proposition 3.1 is given in Appendix 3.6. The first two constraints (3.10b) and (3.10c) are linear with respect to variables η , ν , and ω , while the last constraint (3.10d) defines the linear matrix inequality. For such a semi-definite program, we have the interior-point algorithm [87] and existing optimization toolbox such as SeDuMi [88] to solve it efficiently.

3.2.3 Bound Achieving Distribution

An optimal distribution $f_e(\xi) \in \mathbf{U}_g$ that achieves the detection bound in problem (3.10a)-(3.10d) is defined as an *extremal distribution*. We find that the extremal distribution for problem (3.10a)-(3.10d) only assigns positive mass on two points where the quadratic curve $c(\nu, \omega, \eta)$ is tangent with $-h(\xi, \lambda)$. Specifically, the tangential

points locate at $\{\lambda, \mu_{p,0} + \frac{\omega}{\nu}\}$ as indicated by two hollow points in Figure 3.1, thus we can construct an extremal distribution $f_e(\xi)$ that achieves the lower bound as follows:

$$f_e(\xi) = \begin{cases} 1 - Q_d^L(\lambda), & \text{if } \xi = \lambda, \\ Q_d^L(\lambda), & \text{if } \xi = \mu_{p,0} + \frac{\omega}{\nu}, \\ 0, & \text{otherwise,} \end{cases}$$

where $Q_d^L(\lambda)$ denotes the lower bound of detection probability and is given by the optimum of problem (3.10a)-(3.10d). Note that the extremal distribution takes a discrete form. To study its properties, we have the following proposition, which moreover gives a closed-form expression for the lower detection bound.

Proposition 3.2 *The extremal distribution achieving the lower bound is with mean and variance as follows:*

$$\begin{aligned} \mu_e &= \mu_{p,0} - \sigma_{p,0} \sqrt{\gamma_{p,\mu}}, \\ \sigma_e^2 &= (1 + \gamma_{p,\sigma} - \gamma_{p,\mu}) \sigma_{p,0}^2. \end{aligned}$$

Moreover, given the decision threshold $\lambda \leq \mu_e$, the lower bound can be expressed in an analytical form:

$$Q_d^L(\lambda) = [1 + \sigma_e^2 / (\mu_e - \lambda)^2]^{-1}. \quad (3.11)$$

The proof for Proposition 3.1 is given in Appendix 3.6. The formula in (3.11) provides an easy way to measure the lower bound, instead of solving the optimization problem in (3.10a)-(3.10d).

3.3 Analysis of the Reference-based Uncertainty

Up to this point, we present an equivalent problem and a closed-form formula to study the lower detection bound, considering moment uncertainties. Moreover, we show that the lower detection bound is achievable by a discrete distribution. Since the signal samples from channel sensing hardly follow a discrete distribution, the resulting detection bound may be rather conservative in practice. In this part, we intend to obtain less conservative bounds by extracting a reference distribution, rather than just signal statistics, from historical data.

3.3.1 Reference-based Uncertainty Model

Though signal distribution is fluctuating over time and hard to describe in a closed-form expression, we may consider empirical distribution as a useful reference and allow the actual signal distribution to shift around it. For example, we can assume that the noise signal $f_0(\xi)$ is more or less close to a known Gaussian distribution $f_n^0(\xi)$, which can be obtained based on long-term field measurement. Similarly, primary signal's reference distribution $f_p^0(\xi)$ can take different closed-form expressions as in [16, 17] when channel conditions fluctuate.

The difference between $f_0(\xi)$ (or $f_1(\xi)$, respectively) and its reference $f_n^0(\xi)$ (or $f_p^0(\xi)$, respectively) can be described by a probabilistic distance measure. For example, the Kullback-Leibler (KL) divergence [89] is a non-symmetric measure of the difference between two probability distributions, i.e., $f(\xi)$ and $f^0(\xi)$. Generally, one of the distribution $f(\xi)$ represents the signal's real distribution through precise modeling. The reference $f^0(\xi)$ is a closed-form approximation based on theoretic assumptions and simplifications. The definition of KL divergence between two continuous distributions is given as follows:

$$D_{KL}(f(\xi), f^0(\xi)) = \int_{\xi \in S} [\ln f(\xi) - \ln f^0(\xi)] f(\xi) d\xi. \quad (3.12)$$

When distributions $f(\xi)$ and $f^0(\xi)$ are close to each other, the distance measure is close to zero. Based on the KL divergence, we define the distribution uncertainty as follows:

$$\mathcal{U}_r(f^0(\xi), D_0) = \{f(\xi) \mid \mathbb{E}_f [\ln f(\xi) - \ln f^0(\xi)] \leq D_0\}, \quad (3.13)$$

where $D_0 > 0$ represents a distance limit and is obtained from empirical data or real-time measurement. It indicates signal's fluctuation level. If the signal is highly volatile, we have less confidence on the reference distribution and thus set a larger distance limit. In practice, we can set $f_n^0(\xi)$ and $f_p^0(\xi)$ as those distributions commonly employed in literature [16, 17, 90, 91] and maintain the distance limit during channel measurements. Each time we fit the sensed signal samples into a closed-form distribution and calculate the KL divergence with respect to $f_n^0(\xi)$ (or $f_p^0(\xi)$, respectively)

if PU is absent (or present, respectively), then we overwrite the stored distance limit by the new KL divergence if it becomes larger. The reference distributions can also be updated online if real-time measurements indicate large deviations, in terms of KL divergence, from their reference distributions.

3.3.2 Determine a Robust Decision Threshold

Consider the noise signal $f_0(\xi)$ with reference distribution $f_n^0(\xi)$ and distance limit D_n , we have the following constraints for noise distribution $f_0(\xi)$:

$$\mathbb{E}_{f_0} [\ln f_0(\xi) - \ln f_n^0(\xi)] \leq D_n, \quad (3.14a)$$

$$\mathbb{E}_{f_0}[1] = 1. \quad (3.14b)$$

By the Lagrangian method, we have the worst-case false alarm probability as follows:

$$Q_f^U(\lambda) = \min_{\tau, \eta} \max_{f_0(\xi)} \mathbb{E}_{f_0} \left[h(\xi, \lambda) - \eta - \tau \ln \frac{f_0(\xi)}{f_n^0(\xi)} \right] + \tau D_n + \eta,$$

where $\tau \geq 0$ and η are Lagrangian multipliers associated with constraints (3.14a) and (3.14b), respectively. Let $\mathcal{P}(\lambda, f_0, \tau, \eta) = \mathbb{E}_{f_0} \left[h(\xi, \lambda) - \eta - \tau \ln \frac{f_0(\xi)}{f_n^0(\xi)} \right]$, then the derivative of $\mathcal{P}(\lambda, f_0, \tau, \eta)$ with respect to f_0 is given by

$$\partial \mathcal{P} / \partial f_0 = \int_{\xi \in S} \left(h(\xi, \lambda) - \tau \ln \frac{f_0(\xi)}{f_n^0(\xi)} - \tau - \eta \right) d\xi.$$

By the Karush-Kuhn-Tucker (KKT) optimality conditions, we thus have:

$$h(\xi, \lambda) - \tau \ln \frac{f_0(\xi)}{f_n^0(\xi)} - \tau - \eta = 0, \quad (3.15a)$$

$$\int_{\xi \in S} f_0(\xi) d\xi = 1, \quad (3.15b)$$

$$D_n - \mathbb{E} \left[\ln \frac{f_0(\xi)}{f_n^0(\xi)} \right] \geq 0, \quad (3.15c)$$

$$\tau \left(D_n - \mathbb{E} \left[\ln \frac{f_0(\xi)}{f_n^0(\xi)} \right] \right) = 0. \quad (3.15d)$$

From (3.15a), the optimal distribution function is as follows:

$$f_0^*(\xi) = f_n^0(\xi) \exp \left(\frac{h(\xi, \lambda) - \eta}{\tau} - 1 \right). \quad (3.16)$$

The dual variables (τ, η) in (3.16) should be chosen properly such that conditions (3.15b)-(3.15d) are satisfied. Specifically, we have the following results:

Proposition 3.3 *The choice of (τ, η) is a solution of the following nonlinear equations.*

$$H_1(\tau, \eta) \triangleq R(\lambda)e^{-\eta/\tau} + S(\lambda)e^{(1-\eta)/\tau} - 1 = 0, \quad (3.17a)$$

$$H_2(\tau, \eta) \triangleq S(\lambda)e^{(1-\eta)/\tau} - \eta - \tau(1 + D_n) = 0, \quad (3.17b)$$

where $S(\lambda) = (1 - F_n^0(\lambda)) \exp(-1)$, $R(\lambda) = F_n^0(\lambda) \exp(-1)$, and $F_n^0(\lambda) = \int_{\xi \geq \lambda} f_n^0(\xi) d\xi$ denotes the complementary cumulative distribution function of reference distribution $f_n^0(\xi)$.

The proof for Proposition 3.3 is given in Appendix 3.6. However, it is still very hard to obtain an explicit solution from (3.17a) and (3.17b), thus we propose the Newton iterations as detailed in Algorithm 1. First, we take linear approximations for the nonlinear equations $H_1(\tau, \eta)$ and $H_2(\tau, \eta)$:

$$\begin{bmatrix} H_1(\tau_k + \Delta\tau_k, \eta_k + \Delta\eta_k) \\ H_2(\tau_k + \Delta\tau_k, \eta_k + \Delta\eta_k) \end{bmatrix} \approx \begin{bmatrix} H_1(\tau_k, \eta_k) \\ H_2(\tau_k, \eta_k) \end{bmatrix} + \mathbf{J}_k \begin{bmatrix} \Delta\tau_k \\ \Delta\eta_k \end{bmatrix},$$

where $\mathbf{J}_k$ denotes the Jacobian matrix evaluated at (τ_k, η_k) . If (τ_k, η_k) is not feasible, then we find an update vector $[\Delta\tau_k, \Delta\eta_k]^T$ such that

$$\mathbf{J}_k \begin{bmatrix} \Delta\tau_k \\ \Delta\eta_k \end{bmatrix} = - \begin{bmatrix} H_1(\tau_k, \eta_k) \\ H_2(\tau_k, \eta_k) \end{bmatrix}.$$

Note that τ_k should be projected on non-negative real number during iterations. For simplicity, we use $[x]^+ = \max\{x, 0\}$ to denote the projection on non-negative real number.

Once we determine the solution for (3.17a) and (3.17b) in Proposition 3.3, we obtain the worst-case false alarm probability from (3.15a) and (3.15d) as follows:

$$Q_f^U(\lambda) = \mathbb{E}_{f_0^*}[h(\xi, \lambda)] = (1 + D_n)\tau + \eta. \quad (3.18)$$

Then our task is to find the decision threshold λ such that $Q_f^U(\lambda) = \alpha$, which involves the calculation of inverse function of $Q_f^U(\lambda)$ and it is not directly obtainable from (3.18). However, we have the following property regarding function $Q_f^U(\lambda)$ that may help us design a search method.

Algorithm 1 Search for robust decision threshold

Input: Reference distribution $f_n^0(\xi)$, distance limit D_n , search radius l , and tolerance ϵ

Output: Robust decision threshold such that $Q_f^U(\lambda^*) = \alpha$

```

1: initialize  $\lambda_- = 0$ ,  $\lambda^- = l\mu_{p,0}$ , and set  $\mathbf{H}(\tau, \eta) = [H_1(\tau, \eta), H_2(\tau, \eta)]^T$ 
2: while  $|\lambda_- - \lambda^-| > \epsilon$ 
3:   set  $\bar{\lambda} = \frac{\lambda_- + \lambda^-}{2}$  and initiate the time iteration  $k = 1$ 
4:   while  $|\mathbf{H}(\tau_k, \eta_k)| > \epsilon$ 
5:     evaluate  $\mathbf{H}(\tau_k, \eta_k)$  and Jacobian matrix  $\mathbf{J}(\tau_k, \eta_k)$ 
6:     solve  $\mathbf{J}(\tau_k, \eta_k)\Delta\mathbf{x}_k = -\mathbf{H}(\tau_k, \eta_k)$ 
7:     update  $\tau_{k+1} = [\tau_k + \Delta\tau_k]^+$ ,  $\eta_{k+1} = \eta_k + \Delta\eta_k$ 
8:     update  $Q_f^U(\bar{\lambda}) = (1 + D_n)\tau_{k+1} + \eta_{k+1}$ 
9:     set  $k = k + 1$ 
10:  end while
11:  if  $(Q_f^U(\bar{\lambda}) - \alpha)(Q_f^U(\lambda_-) - \alpha) < 0$ 
12:    set  $\lambda^- = \bar{\lambda}$ 
13:  else
14:    set  $\lambda_- = \bar{\lambda}$ 
15:  end if
16:  if  $|Q_f^U(\bar{\lambda}) - \alpha| < \epsilon$ 
17:    break
18:  end if
19: end while
20: set  $\lambda^* = \bar{\lambda}$ 
    
```

Proposition 3.4 *Worst-case false alarm probability $Q_f^U(\lambda)$ is non-increasing with respect to the decision threshold λ .*

The conclusion in Proposition 3.4 is obvious since we have

$$dQ_f^U(\lambda)/d\lambda = d\mathbb{E}_{f_0^*}[h(\xi, \lambda)]/d\lambda = -f_0^*(\lambda) \leq 0.$$

Though direct solution to $Q_f^U(\lambda) = \alpha$ is not available, the monotonicity of $Q_f^U(\lambda)$ enlightens us a bisection method to search the solution for $Q_f^U(\lambda) = \alpha$. We propose the search procedure in Algorithm 1. It performs the search within an interval set $[0, l\mu_{p,0}]$, where l is an empirical constant such that $Q_f^U(l\mu_{p,0}) < \alpha$. From lines 4 to 10

of Algorithm 1, we use the Newton iterations to solve the equations in Proposition 3.3 and obtain the worst-case false alarm with fixed decision threshold³. Then we compare the false alarm probabilities at $\bar{\lambda}$ and λ_- with the required false alarm probability α , respectively. The comparing result helps to shrink the search region as in lines 11 – 15.

3.3.3 Lower Bound of Detection Probability

Given the decision threshold in Algorithm 1, we proceed to analyze the detection bound with distance limit D_p and reference distribution $f_p^0(\xi)$, which can be a Rayleigh or Rician distribution depending on different channel models. We have assumed that the real distribution $f_1(\xi)$ is not very different from the reference distribution, however, when we actually detect a very large deviation from the reference distribution, we can update $f_p^0(\xi)$ as the most suitable one from a set of candidate models by goodness of fit test. With this assumption, we consider the reference model as the intersection of moment uncertainty and a reference distribution, i.e., $\mathbf{U}_r(f_p^0(\xi), D_p) \cap \mathbf{U}_g(\mu_{p,0}, \sigma_{p,0}, \gamma_{p,\mu}, \gamma_{p,\sigma})$. Then, we obtain the KKT conditions for the lower bound problem by the same way as in (3.15a)-(3.15d):

$$\tilde{h}(\xi, \lambda) + \eta + \tau \ln \frac{f_1(\xi)}{f_p^0(\xi)} + \tau = 0, \quad (3.19a)$$

$$\int_{\xi \in S} f_1(\xi) d\xi = 1, \quad (3.19b)$$

$$\tau \left(\mathbb{E} \left[\ln \frac{f_1(\xi)}{f_p^0(\xi)} \right] - D_p \right) = 0, \quad (3.19c)$$

$$\nu (\mathbb{E}[\xi^2 - 2\mu_{p,0}\xi] + \mu_{p,0}^2 - (1 + \gamma_{p,\sigma})\sigma_{p,0}^2) = 0, \quad (3.19d)$$

$$\omega (\mathbb{E}[-2\xi] + 2\mu_{p,0} - 2\sigma_{p,0}\text{Sign}(\omega)\sqrt{\gamma_{p,\mu}}) = 0, \quad (3.19e)$$

where $\tilde{h}(\xi, \lambda) \triangleq h(\xi, \lambda) - 2(\omega + \mu_{p,0}\nu)\xi + \nu\xi^2$ and $(\eta, \nu, \omega, \tau)$ are the Lagrangian multipliers. Then we obtain the optimal distribution function as follows:

$$f_1^*(\xi) = f_p^0(\xi) \exp \left(-\frac{1}{\tau} [\tilde{h}(\xi, \lambda) + \eta] - 1 \right). \quad (3.20)$$

³In Algorithm 1, we use $[x]^+ = \max\{x, 0\}$ to denote the projection on non-negative real number.

If $(\eta, \nu, \omega, \tau)$ satisfy the equations (3.19a)-(3.19e), we can obtain the lower bound of detection probability $Q_d^L(\lambda) = \mathbb{E}_{f_1^*}[h(\xi, \lambda)]$ as follows:

$$\begin{aligned} Q_d^L(\lambda) &= [\mu_{p,0}^2 - (1 + \gamma_{p,\sigma})\sigma_{p,0}^2]\nu + 2\mu_{p,0}\omega \\ &\quad - 2|\omega|\sigma_{p,0}\sqrt{\gamma_{p,\mu}} - (1 + D_p)\tau - \eta. \end{aligned} \quad (3.21)$$

However, it is very complicated to transform the KKT conditions (3.19a)-(3.19e) into a compact form similar to that in Proposition 3.3, and we are unable to solve $(\eta, \nu, \omega, \tau)$ by Newton iterations as in Algorithm 1. Here, we try to solve it in another way: At some iteration k , we fix $(\eta_k, \nu_k, \omega_k, \tau_k)$ and find the optimal distribution function in (3.20). Then we determine the numerical approximations for signal moments $\mathbb{E}[1]$, $\mathbb{E}[\xi]$, $\mathbb{E}[\xi^2]$, and $\mathbb{E}\left[\ln \frac{f_1^*(\xi)}{f_p^0(\xi)}\right]$, respectively. These values are used to check whether (3.19b)-(3.19e) are satisfied. If not satisfied, we will update $\mathbf{x}_k = (\eta_k, \nu_k, \omega_k, \tau_k)$ in a direction $\Delta_k = [\Delta_\eta, \Delta_\nu, \Delta_\omega, \Delta_\tau]$ given as follows:

$$\begin{aligned} \Delta_\eta &= \int_{\xi \in S} f_1^*(\xi) d\xi - 1, \\ \Delta_\nu &= [\mathbb{E}[\xi^2] - 2\mu_{p,0}\mathbb{E}[\xi] + \mu_{p,0}^2 - (1 + \gamma_{p,\sigma})\sigma_{p,0}^2]^+, \\ \Delta_\omega &= -2\mathbb{E}[\xi] + 2\mu_{p,0} - 2\sigma_{p,0}\text{Sign}(\omega)\sqrt{\gamma_{p,\mu}}, \\ \Delta_\tau &= \left[\mathbb{E}\left[\ln \frac{f_1^*(\xi)}{f_p^0(\xi)}\right] - D_p \right]^+. \end{aligned}$$

Then we have $\mathbf{x}_{k+1} = T(\mathbf{x}_k) \triangleq \mathbf{x}_k - \text{diag}(\beta(k))\Delta_k$, where vector $\beta(k)$ denotes a sufficiently small step-size at k -th iteration for each term in $\mathbf{x}_k$. Given this new variable $\mathbf{x}_{k+1} = (\eta_{k+1}, \nu_{k+1}, \omega_{k+1}, \tau_{k+1})$, we return to update the distribution function as in (3.20). With proper choice of the step size $\beta(k)$, the update $T(\mathbf{x}_k)$ forms a contracting mapping and the update procedure will converge to a fixed point $\mathbf{x}_\infty$. The choice of step size $\beta(k)$ and the convergence properties were studied in [92]. Finally, the lower detection bound is obtained by substituting $\mathbf{x}_\infty$ into (3.21).

3.4 Numerical Results

In the numerical results, we first obtain the robust decision thresholds under different uncertainties, and verify their robustness by simulated noise signals. Specifically, we

simulate noise fluctuations using random variables that are drawn from different distributions. Then, we count the number of events when signal sample is greater than the decision threshold, and use it to estimate the actual false alarm probability. We set $\alpha = 0.1$ as the required false alarm probability, and define $\gamma_{n,\mu} = \gamma_{n,\sigma} = 0.1$ to regulate the size of noise uncertainty. The validation of detection bounds contains two parts. For the moment uncertainties, we set $\mu_{p,0} = 3$ and $\sigma_{p,0}^2 = 1$ as the sample mean and variance of the received primary signal, respectively, and $\gamma_{p,\mu} = \gamma_{p,\sigma} = 0.1$ to regulate their fluctuating ranges, respectively. Then we compare the lower detection bounds obtained through finite linear approximations, equivalent semi-definite program, and closed-form formula in (3.11), respectively. As a benchmark method, the finite linear approximation will help us identify the correctness of the semi-definite equivalence and formula in (3.11). For the reference uncertainty model, we assume a known reference distribution, which can be channel-dependent and updated online when spectrum environment changes. Without loss of generality, we assume that the reference distributions of noise and primary signals both follow Gaussian distributions, then we numerically study the false alarm probability and detection probability by our proposed iterative procedures. Their convergent values are also compared with that of the moment uncertainties.

3.4.1 False Alarm Probability under Noise Uncertainties

When we consider the KDUS model in (3.1), we obtain a larger detection threshold in (3.5) which we denote as λ_{KDUS} , compared with a nominal design⁴ $\lambda_{DSM} = \mu_{n,0} + \sigma_{n,0}Q^{-1}(\alpha)$. Figure 3.2 plots the maps from decision thresholds to false alarm probabilities under deterministic signal model (DSM) and the KDUS model, respectively. Though we did not analyze the false alarm probability with GDES and GDUS models, the formulation in (3.7a)-(3.7b) can be slightly modified to study the worst-case false alarm probability. We also plot their results in Figure 3.2 for comparison. Note that $\lambda_{DSM} < \lambda_{KDUS} < \lambda_{GDES} < \lambda_{GDUS}$, which implies that a larger decision

⁴It means that we assume a deterministic signal model, and denote the resulting decision threshold as λ_{DSM} .

Figure 3.2: False alarm probability with different decision threshold.

threshold is required to ensure the same false alarm probability if signal's distribution information is unavailable.

To investigate the actual false alarm probability with these decision thresholds, we emulate noise signal $f_0(\xi)$ by a stream of random samples that are generated from different Gaussian distributions. The noise signal's mean and variance are confined by parameters $\gamma_{n,\mu}$ and $\gamma_{n,\sigma}$, respectively. Therefore, the actual noise samples are randomly drawn from any Gaussian distribution with $\mu_{n,0} - \sqrt{\gamma_{n,\mu}}\sigma_{n,0} \leq \mu_n \leq \mu_{n,0} + \sqrt{\gamma_{n,\mu}}\sigma_{n,0}$ and $\sigma_{n,0}^2(1 - \gamma_{n,\sigma}) \leq \sigma_n^2 \leq \sigma_{n,0}^2(1 + \gamma_{n,\sigma})$. Then we apply the emulated noise signal to different decision thresholds, and check the actual false alarm probabilities. Given different required false alarm probabilities (x -axis in Figure 3.3), we plot the actual false alarm probabilities with different decision thresholds in Figure 3.3. In case I, the noise mean μ_n and variance σ_n^2 are uniformly drawn from their fluctuating ranges, respectively. We can see that, the required false alarm probability can be guaranteed by decision threshold λ_{DSM} on average, however, it is violated in case II where the noise signal tends to take the largest mean $\mu_{n,0} + \sqrt{\gamma_{n,\mu}}$ and variance $\sigma_{n,0}^2(1 + \gamma_{n,\sigma})$. Since the KDUS model considers worst-case signal fluctuation in designing the decision threshold, it still guarantees the actual false alarm probability to be less than the required level in case II.

Figure 3.3: Actual false alarm probability under noise fluctuations.

3.4.2 Detection Bounds with Moment Uncertainties

To study the detection probability, we fix the decision threshold as λ_{KDUS} . In Figure 3.4, we first plot the lower detection bound with KDUS model, which is explicitly given in (3.6). Regarding the GDUS model in problem (3.7a)-(3.7b), we have discussed two solution methods. The first method considers a finite linear approximation of the original semi-infinite linear program by dividing the feasible set into discrete points with a grid size Δ . While the second method finds the exact lower bound by solving an equivalent semi-definite program. To compare these two methods, we take different grid sizes for the finite linear approximation, i.e., $\Delta_1 = 0.1$ and $\Delta_2 = 0.01$, so as to check the approximation accuracy with different grid size. Figure 3.4 shows that the linear approximation and the semi-definite program (3.10a)-(3.10d) provide very close results when the grid size becomes smaller. When the grid size is small enough, the linear approximation actually achieves the same bound with that of problem (3.10a)-(3.10d), which also implies that the dual problem (3.10a)-(3.10d) preserves strong duality with the original problem (3.7a)-(3.7b). Besides, Figure 3.4 shows that the analytical bound in (3.11) is consistent with the lower bound obtained in (3.10a)-(3.10d). This observation enables SUs to quick assess the lower bound of detection probability without solving the linear approximation or semi-definite program.

Figure 3.4: Lower detection bounds with different moment uncertainties.

3.4.3 The Importance of a Reference Distribution

We assume that the reference distribution of noise signal is a standard Gaussian distribution $f_n^0(\xi)$. Considering the same fluctuation range for noise signal, i.e., $\gamma_{n,\mu} = \gamma_{n,\sigma} = 0.1$, we find that the KL divergence between $f_n^0(\xi)$ and any distribution $f_0(\xi)$ in $\mathcal{U}_d(0, 1, 0.1, 0.1)$ is always less than 0.076. Therefore, we set the distance limit D_n in an equivalent magnitude, e.g., $D_n = 0.1$. For different choices of decision threshold λ , we solve a pair of variables (τ, η) that satisfies equations (3.17a) and (3.17b). Then we obtain the worst-case false alarm probability $Q_f^U(\lambda) = (1 + D_n)\tau + \eta$ for three different distance limits D_n , which are plotted in Figure 3.5, respectively. We also plot false alarm probabilities with different moment uncertainties in the same figure. It is obvious that decision threshold with the reference model increases with the increase of the distance limit. This observation is intuitive since larger distance limit allows the noise signal to fluctuate more intensively. We also note that the false alarm probability is a decreasing function of the decision threshold. Thus, we use the bisection method in Algorithm 1 to search the decision threshold that satisfies the false alarm probability requirement. Comparing with moment uncertainties, the reference model produces larger decision thresholds than that of the KDUS model and less than that of the GDES and GDUS models.

Figure 3.5: False alarm probability with different distance limits.

A comparison in Figure 3.6 also shows a similar trend that the detection bound of a reference model is looser than that of the KDUS model, while tighter than that of the GDES and GDUS models. This is because, the KDUS model requires the signal to follow exactly a known distribution, while the reference model considers a more general case. It allows discrepancy between actual distribution and its reference, however, the discrepancy is limited and confined by a probabilistic distance measure, which poses a stronger assumption on the shape of distribution functions than that of the GDES and GDUS models. Simply put, the reference model allows the actual signal to follow a different distribution function, which should not be too abnormal based on historical data or empirical knowledge. In the simulation, we emulate primary signal by two streams of random samples. One is drawn from a Gaussian distribution (denoted as Gaussian in Figure 3.6) whose mean and variance are uniformly picked from sets $[\mu_{p,0} - \sqrt{\gamma_{p,\mu}}\sigma_{p,0}, \mu_{p,0} + \sqrt{\gamma_{p,\mu}}\sigma_{p,0}]$ and $[(1 - \gamma_{p,\sigma})\sigma_{p,0}^2, (1 + \gamma_{p,\sigma})\sigma_{p,0}^2]$, respectively. The other one follows a log-normal distribution (denoted as Lognormal in Figure 3.6) with mean and variance bounded in the same uncertainty set. Since the KDUS model assumes Gaussian distribution when calculating the detection bound, it well bounds the Gaussian signal as in Figure 3.6. However, it fails to bound it when the actual signal follows a log-normal distribution. Fortunately, our reference model provides tighter bound in this case since it is tolerable to the mismatch of distribution functions.

Figure 3.6: Lower detection bounds with different uncertainties.

3.5 Summary

In this chapter, we study the sensing performance with fluctuating signals, which are characterized by different uncertainty models. First, we find a robust decision threshold that guarantees the required false alarm probability under noise fluctuation. Then we consider the distribution uncertainty of PUs' signals and study the sensing performance in terms of the lower detection bound. For the moment-based uncertainty model, we show that the lower detection bound is equivalent to a semi-definite program, and present a closed-form formula to calculate the detection bound. Furthermore, we introduce a reference distribution into the signal uncertainties. This new model allows signal distribution to be different at each observation, however, not to be inconsistent with respect to the past observations or empirical knowledge. Numerical results show that the reference-based uncertainty model is flexible to characterize signal's fluctuation, and could improve the detection performance compared with moment-based uncertainties in general case.

3.6 Appendix

3.6.1 Proof for Proposition 3.1

For ease of analysis, we first rewrite the second constraint in (3.3) as

$$\mathbb{E} \begin{bmatrix} \sigma_{p,0}^2 & \xi - \mu_{p,0} \\ \xi - \mu_{p,0} & \gamma_{p,\mu} \end{bmatrix} \succeq 0,$$

then we assign different dual variables η , $Z \succeq 0$, and $\nu > 0$ to these constraints, respectively. Let $Z = \begin{bmatrix} \kappa & \omega \\ \omega & s \end{bmatrix}$, we have the Lagrangian function $\Gamma(f_1, \eta, Z, \nu)$ as in (3.8). By the geometric interpretation in Figure 3.1, we then have the dual problem

$$\max_{Z, \nu, \eta} \quad [\mu_{p,0}^2 - (1 + \gamma_{p,\sigma})\sigma_{p,0}^2] \nu + 2\mu_{p,0}\omega - \eta - \gamma_{p,\mu}s - \sigma_{p,0}^2\kappa \quad (3.23a)$$

$$s.t. \quad (\mu_{p,0}\nu + \omega)^2 - \nu(\eta + 1) \leq 0, \quad (3.23b)$$

$$\nu\xi^2 - 2(\omega + \mu_{p,0}\nu)\xi + \eta \geq 0, \quad \forall \xi \leq \lambda \quad (3.23c)$$

$$Z \succeq 0, \quad \nu \geq 0, \quad (3.23d)$$

where $Z \succeq 0$ implies $\kappa s - \omega^2 \geq 0$, $\kappa \geq 0$ and $s \geq 0$. If $s = 0$, we have $\kappa = \omega = 0$, then (3.23a) is reduced to $[\mu_{p,0}^2 - (1 + \gamma_{p,\sigma})\sigma_{p,0}^2] \nu - \eta$. If $s > 0$, then we have

$$\begin{aligned} & [\mu_{p,0}^2 - (1 + \gamma_{p,\sigma})\sigma_{p,0}^2] \nu + 2\mu_{p,0}\omega - \eta - \gamma_{p,\mu}s - \sigma_{p,0}^2\kappa \\ & \leq [\mu_{p,0}^2 - (1 + \gamma_{p,\sigma})\sigma_{p,0}^2] \nu + 2\mu_{p,0}\omega - \eta - 2\sqrt{\gamma_{p,\mu}\sigma}|\omega|. \end{aligned}$$

Let $Q_d^L(\lambda) = [\mu_{p,0}^2 - (1 + \gamma_{p,\sigma})\sigma_{p,0}^2] \nu + 2\mu_{p,0}\omega - \eta - 2\sqrt{\gamma_{p,\mu}\sigma}|\omega|$. If $\mu_{p,0} + \frac{\omega}{\nu} \leq \lambda$, then (3.23c) is reduced to $(\mu_{p,0}\nu + \omega)^2 - \nu\eta \leq 0$ and we have $Q_d^L(\lambda) \leq -\gamma_{p,\sigma}\sigma^2\nu - 2|\omega|\sqrt{\gamma_{p,\mu}\sigma} - \frac{\omega^2}{\nu} < 0$. Since a negative probability in this case is not feasible in practice, we consider the alternate case which leads to the equivalence in Proposition 3.1.

3.6.2 Proof for Proposition 3.2

Since an extremal distribution only assigns nonnegative probabilities on the points $\{\lambda, \mu + \frac{\omega^*}{\nu^*}\}$ that satisfies the equation $h(\xi, \lambda) = \nu^*\xi^2 - 2(\omega^* + \mu_{p,0}\nu^*)\xi + \eta^*$, and the inequality constraints (3.10b), (3.10c) hold with equalities at these points. Therefore we have $\mu_{p,0} + \frac{\omega}{\nu} - \lambda = \frac{1}{\sqrt{\nu}}$ and $\eta = \nu(\mu_{p,0} + \frac{\omega}{\nu})^2 - 1^5$. We substitute these two equations

⁵We remove the star mark (*) for simplicity.

into the objective function (3.10a), the lower bound is rewritten as in (3.24).

$$\begin{aligned}
 Q_d^L(\lambda) &= [\mu_{p,0}^2 - (1 + \gamma_{p,\sigma})\sigma_{p,0}^2] \nu + 2\mu_{p,0}\omega - \eta - 2\sqrt{\gamma_{p,\mu}}\sigma_{p,0}|\omega| \\
 &= \nu \left(\mu_{p,0} + \frac{\omega}{\nu} \right)^2 - \eta - \frac{\omega^2}{\nu} - 2\sqrt{\gamma_{p,\mu}}\sigma_{p,0}|\omega| - (1 + \gamma_{p,\sigma})\sigma_{p,0}^2\nu \\
 &= 1 - \nu \left(\sigma_{p,0}\sqrt{\gamma_{p,\mu}} + \frac{|\omega|}{\nu} \right)^2 - \nu(1 + \gamma_{p,\sigma})\sigma_{p,0}^2 \\
 &= -2(\sigma_{p,0}\sqrt{\gamma_{p,\mu}} + \lambda - \mu_{p,0})\sqrt{\nu}
 \end{aligned} \tag{3.24}$$

$$- [(\sigma_{p,0}\sqrt{\gamma_{p,\mu}} + \lambda - \mu_{p,0})^2 + (1 + \gamma_{p,\sigma} - \gamma_{p,\mu})\sigma_{p,0}^2] \nu. \tag{3.25}$$

Considering $\omega > 0$, it is further reduced to (3.25). Therefore, the optimal lower bound is given by

$$Q_d^L(\lambda) = \left[1 + \frac{(1 + \gamma_{p,\sigma} - \gamma_{p,\mu})\sigma_{p,0}^2}{(\sigma_{p,0}\sqrt{\gamma_{p,\mu}} + \lambda - \mu_{p,0})^2} \right]^{-1} \tag{3.26}$$

and it is achieved when

$$\sqrt{\nu^*} = \frac{\mu_{p,0} - \sigma_{p,0}\sqrt{\gamma_{p,\mu}} - \lambda}{(\sigma_{p,0}\sqrt{\gamma_{p,\mu}} + \lambda - \mu_{p,0})^2 + (1 + \gamma_{p,\sigma} - \gamma_{p,\mu})\sigma_{p,0}^2}. \tag{3.27}$$

Now we are to determine the mean and variance of the discrete extremal distribution function, i.e., $f_e(\xi_1 = \lambda) = 1 - Q_d^L(\lambda)$ and $f_e(\xi_2 = \mu_{p,0} + \frac{\omega}{\nu}) = Q_d^L(\lambda)$. Let μ_e and σ_e^2 denote the equivalent mean and variance of the extremal distribution $f_e(\xi)$, respectively.

$$\begin{aligned}
 \mu_e &= (1 - Q_d^L(\lambda))\lambda + Q_d^L(\lambda) \left(\mu_{p,0} + \frac{\omega}{\nu} \right) \\
 &= \lambda + \frac{Q_d^L(\lambda)}{\sqrt{\nu}} = \mu_{p,0} - \sigma_{p,0}\sqrt{\gamma_{p,\mu}},
 \end{aligned} \tag{3.28}$$

$$\begin{aligned}
 \sigma_e^2 &= (1 - Q_d^L(\lambda))(\mu_{p,0} - \sigma_{p,0}\sqrt{\gamma_{p,\mu}} - \lambda)^2 + Q_d^L(\lambda) \left(\frac{\omega}{\nu} + \sigma_{p,0}\sqrt{\gamma_{p,\mu}} \right)^2 \\
 &= (1 - Q_d^L(\lambda))(\mu_{p,0} - \sigma_{p,0}\sqrt{\gamma_{p,\mu}} - \lambda)^2 + Q_d^L(\lambda) \left(\frac{1}{\sqrt{\nu}} - \mu_{p,0} + \sigma_{p,0}\sqrt{\gamma_{p,\mu}} + \lambda \right)^2 \\
 &= (1 + \gamma_{p,\sigma} - \gamma_{p,\mu})\sigma_{p,0}^2.
 \end{aligned} \tag{3.29}$$

From the expressions in (3.28) and (3.29), the equivalent mean μ_e and variance σ_e^2 both have very simple forms. Specifically, distribution $f_e(\xi)$ will show the largest mean value $\mu_{p,0} + \sqrt{\gamma_{p,\mu}}\sigma_{p,0}$, while the variance σ_e^2 is less than the maximum $(1 + \gamma_{p,\sigma})\sigma_{p,0}^2$. Moreover, we can rewrite the lower bound in terms of μ_e and σ_e^2 as $Q_d^L(\lambda) = \left[1 + \frac{\sigma_e^2}{(\mu_e - \lambda)^2} \right]^{-1}$.

3.6.3 Proof for Proposition 3.3

The proof is straightforward by substituting distribution $f_0^*(\xi)$ in (3.16) back to conditions (3.15b) and (3.15d). Note that (3.15c) holds when we have (3.15d) and $\tau \geq 0$.

From (3.15b), we have

$$\begin{aligned} \int_{\xi \in S} f_0^*(\xi) d\xi &= \int_{\xi \in S} f_n^0(\xi) \exp\left(\frac{h(\xi, \lambda) - \eta}{\tau} - 1\right) d\xi \\ &= \int_{\xi < \lambda} f_n^0(\xi) \exp\left(-\frac{\eta}{\tau} - 1\right) d\xi + \int_{\xi \geq \lambda} f_n^0(\xi) \exp\left(\frac{1 - \eta}{\tau} - 1\right) d\xi \\ &= R(\lambda)e^{-\eta/\tau} + S(\lambda)e^{(1-\eta)/\tau} = 1. \end{aligned}$$

Thus we obtain the condition $H_1(\tau, \eta) = 0$ as in (3.17a). Similarly for (3.15d), we

have $\mathbb{E}_{f_0^*} \left[\ln \frac{f_0(\xi)}{f_n^0(\xi)} \right] = \frac{\mathbb{E}[h(\xi, \lambda)] - \eta}{\tau} - 1$, and it is equivalent to

$$\begin{aligned} \tau D_n &= \mathbb{E}[h(\xi, \lambda)] - \eta - \tau = \int_{\xi \geq \lambda} f_n^0(\xi) \exp\left(\frac{1 - \eta}{\tau} - 1\right) d\xi - \eta - \tau \\ &= S(\lambda)e^{(1-\eta)/\tau} - \eta - \tau, \end{aligned}$$

which is exactly the equation in (3.17b).

Chapter 4

Robust Spectrum Sensing for Multi-User

In Chapter 3, we propose several channel uncertainty models and present analytical study on the sensing performance of a single SU. In this chapter, we extend our study to the multi-user case and study the signal uncertainty in a GDUS model, i.e., the statistical features (i.e., mean and variance) of the primary signals are contaminated by estimation errors and the actual probability density function does not match what is often assumed in literature.

Though we have some analytical results for single user, there are some new challenges to study the robust detection performance for multiple users. In the single user case, local information (i.e., the noise uncertainty and false alarm probability requirement) is only accounted for the optimization of local decision threshold. While in cooperative sensing, except local information, each SU also needs to understand other SUs' channel conditions to jointly optimize their decision thresholds. In this Chapter, we show that the robust threshold design at each SU is a non-convex problem. Then we approximate it by a series of analytically tractable semi-definite programs, and propose an iterative algorithm to search the optimal decision threshold while maintaining the desirable false alarm probability.

The rest of the chapter is organized as follows. Section 4.1 describes the network model of cooperative spectrum sensing and our robust design problem. Then we propose the iterative algorithm in Section 4.2, and detail two main steps of the Algorithm in Section 4.3. Section 4.4 shows some numerical results and Section 4.5 summarizes this chapter.

4.1 Multi-user Spectrum Sensing Model

We consider a cognitive radio network with a set $\mathcal{N} = \{1, \dots, N\}$ of SUs and multiple PUs operating under the same spectrum channel in an underlay manner as illustrated in Figure 2.1. SUs may transmit simultaneously with the PUs, however, SUs should control their transmit power to restrict the interference at PU receivers below tolerable levels. We assume that SUs have no pre-knowledge about the PUs' transmission characteristics and application types. Due to the change of PUs' activities on that channel, the received signals at SUs' receivers would be highly dynamic in terms of the statistic information. There is also no information exchange between PUs and SUs, thus SUs need to acquire the spectrum information by themselves through spectrum sensing. SUs are interested in cooperating with each other to improve the sensing performance, since single user sensing is easily subject to errors due to channel fading and shadowing [16, 47, 52]. Without loss of generality, we assume that all N SUs join the cooperation and each SU has a local energy detector.

4.1.1 Sensing Structure

We assume that the local detector employs a threshold-based decision scheme, which is commonly employed in hypothesis testing problem [17, 47]. Our problem here is to choose the optimal decision thresholds for all SUs such that the sensing performance is maximized. Here sensing performance can be defined properly according to SUs' preferences. The whole process can be split into three phases as follows:

- Phase I (Sensing): All local detectors report the raw sensing results (i.e., raw data of signal samples) to the fusion center [46], which performs a *centralized optimization* over the decision thresholds and broadcasts the computed optimal decision thresholds to all SUs.
- Phase II (Detecting): Once SUs adopt the optimal decision thresholds, each SU samples the primary channel and makes local detection independently, by comparing signal samples and its own decision threshold.

- Phase III (Decision fusion): SUs report their local detection decisions to a fusion center. Then the fusion center makes a final decision on the existence of the primary signals based on a decision function as in (4.1).

Phase I can be viewed as an initialization process. After that, Phases II and III repeat in the following time slots until there is a need for re-initialization due to the environment change. We consider one-bit binary decision at local detectors, i.e., if the signal sample at one SU is larger (smaller, respectively) than its decision threshold, the channel is considered as busy (idle, respectively). We also consider a similar threshold structure at the fusion center, and define the multi-user decision function as follows:

$$h(\boldsymbol{\xi}, \boldsymbol{\lambda}) = \mathbf{1} \left(\sum_{i=1}^N \mathbf{1}(\xi_i \geq \lambda_i) \right), \quad (4.1)$$

where $\boldsymbol{\xi} = \{\xi_i\}_{i=1}^N$ and $\boldsymbol{\lambda} = \{\lambda_i\}_{i=1}^N$ denote the received signal strength and decision threshold at each user, respectively. Indicator $\mathbf{1}(A)$ equals 1 if A is true (or $A > 0$) and 0 otherwise. When $h(\boldsymbol{\xi}, \boldsymbol{\lambda}) = 1$, the fusion center reports the presence of a PU (hypothesis H_1), otherwise it reports an idle channel (hypothesis H_0). Then the detection probability Q_d and false alarm probability Q_f are given as follows:

$$\begin{aligned} Q_f &= \int_{\boldsymbol{\xi} \in S} h(\boldsymbol{\xi}, \boldsymbol{\lambda}) f_0(\boldsymbol{\xi}) d\boldsymbol{\xi}, \\ Q_d &= \int_{\boldsymbol{\xi} \in S} h(\boldsymbol{\xi}, \boldsymbol{\lambda}) f_1(\boldsymbol{\xi}) d\boldsymbol{\xi}, \end{aligned}$$

where S denotes the set of all possible sensing results in N -dimensional real space $\mathbb{R}^N$. Function $f_0(\boldsymbol{\xi})$ and $f_1(\boldsymbol{\xi})$ denote the signal distributions when PUs are absent and present, respectively. From the above expressions, we note that the knowledge about these two distribution functions $f_0(\boldsymbol{\xi})$ and $f_1(\boldsymbol{\xi})$ are crucial for an analytical study of the detection performance.

4.1.2 GDUS Model for Multi-User

In practice, it is difficult for an SU to know PUs' signal statistics in advance, as the signal is often time-varying and experiences attenuation, shadowing, and multi-path fading before reaching the SU's receiver. Instead, empirical estimates of the mean $\boldsymbol{\mu}$

and covariance matrix Σ may provide a practical way to study the signals' properties. Nevertheless, we still lack the confidence to entirely rely on these estimates of the signal statistics. These empirical estimates are based on signal samples which bear limited information about the original distribution, thus there may have discrepancies between the estimates and the real mean and variance of the distribution. However, it may be reasonable to assume that the sample estimates, i.e., mean $\boldsymbol{\mu}$ and covariance Σ , are fluctuating within small ranges of the true distribution statistics. Then we define the set of all possible distribution functions, namely, distribution uncertainty, as parameterized by γ_μ and γ_σ :

$$\mathcal{U} = \left\{ f_1(\boldsymbol{\xi}) \left| \begin{array}{l} \mathcal{P}(\boldsymbol{\xi} \in S) = 1 \\ (E[\boldsymbol{\xi}] - \boldsymbol{\mu})^T \Sigma^{-1} (E[\boldsymbol{\xi}] - \boldsymbol{\mu}) \leq \gamma_\mu \\ E[(\boldsymbol{\xi} - \boldsymbol{\mu})(\boldsymbol{\xi} - \boldsymbol{\mu})^T] \preceq (1 + \gamma_\sigma) \Sigma \end{array} \right. \right\}, \quad (4.2)$$

where $\mathcal{P}(\cdot)$ denotes the probability of some event. The first equality constraint requires $f_1(\boldsymbol{\xi})$ to be a valid distribution function. The following two inequalities assume that the distribution mean $E[\boldsymbol{\xi}]$ should lie in a small ellipsoid centered at its estimate $\boldsymbol{\mu}$, and the covariance matrix $E[(\boldsymbol{\xi} - \boldsymbol{\mu})(\boldsymbol{\xi} - \boldsymbol{\mu})^T]$ should lie in a positive semi-definite cone by matrix inequality, respectively. Actually, these two inequalities describe how likely the distribution mean $E[\boldsymbol{\xi}]$ and signal samples $\boldsymbol{\xi}$ are close to the sample estimate $\boldsymbol{\mu}$. The theoretical work in [86] has demonstrated how to choose proper values for γ_μ and γ_σ based on the sample size. In order to use this model in practice and track the change of PUs' signals, we can update the sample estimates $\boldsymbol{\mu}$ and Σ periodically. When the context changes and a local detector detects new samples deviating largely from previous sample statistics, this detector will report a context-change signaling to the fusion center. Then the fusion center starts over the sensing process from Phase I, in which the decision thresholds are optimized based on the new sample statistics.

When there are no transmissions from PUs, the background noise is also time-varying [81, 93], however, the uncertainty of noise signal $f_0(\boldsymbol{\xi})$ only affects the false alarm probability $Q_f = \int_{\boldsymbol{\xi} \in S} h(\boldsymbol{\xi}, \boldsymbol{\lambda}) f_0(\boldsymbol{\xi}) d\boldsymbol{\xi}$. This implies that we can study the distribution uncertainty of noise signal separately. Approximately, we can add a safety margin to the prescribed false alarm probability, thus the new limit is $(1 - \varepsilon)\alpha$,

to counteract the uncertainty of the noise signal. In this chapter, we focus on the choice of decision thresholds and their impacts on the detection probability. Thus we will assume that noise signal at each SU's receiver is deterministic and follows an independent and identical Gaussian distribution, i.e., $f_0(\boldsymbol{\xi}) = (f_0(\xi_1), \dots, f_0(\xi_N))$. The distribution parameters of $f_0(\boldsymbol{\xi})$ can be obtained through field measurements or empirical data.

4.1.3 Robust Sensing Design

As we often observe the case in practice, a nominal design based on deterministic assumptions of the system parameters usually leads to very bad performance when system parameter fluctuates slightly [18]. This shows the practical importance of robust design, which helps to rule out the possibility of severe performance degradation due to inaccurate parameter estimates. With the introduction of uncertainty set for signals' distribution, we intend to find a robust decision threshold vector that provides more stable detection performance against signals' fluctuations. This problem is equivalent to maximizing the worst-case detection performance as follows:

$$\max_{\boldsymbol{\lambda}} \min_{f_1(\boldsymbol{\xi})} \int_{\boldsymbol{\xi} \in S} h(\boldsymbol{\xi}, \boldsymbol{\lambda}) f_1(\boldsymbol{\xi}) d\boldsymbol{\xi} \quad (4.3a)$$

$$s.t. \quad f_1(\boldsymbol{\xi}) \in \mathcal{U}, \quad (4.3b)$$

$$\int_{\boldsymbol{\xi} \in S} h(\boldsymbol{\xi}, \boldsymbol{\lambda}) f_0(\boldsymbol{\xi}) d\boldsymbol{\xi} \leq \alpha, \quad (4.3c)$$

where the noise distribution $f_0(\boldsymbol{\xi})$ is assumed to be known in advance. The above problem (4.3a)-(4.3c) is actually the robust counter-part of a decentralized detection problem and is therefore NP-hard [94].

However the robust problem involves a solvable sub-structure (4.3a)-(4.3b). When we are given a decision threshold $\boldsymbol{\lambda}$ in the non-convex feasible set defined by the false alarm probability constraint (4.3c), the inner minimization problem (4.3a)-(4.3b) is equivalent to finding the lower bound of the detection probability with moment constraints. Some related algorithms were proposed in [95–97] to solve such kind of moment constrained problems. The most related work was presented in [86], where the authors proposed a semi-definite transformation to the problem (4.3a)-(4.3b) by

making use of the duality theory. However, this method does not apply directly to our problem due to the existence of the extra non-convex constraint (4.3c).

4.2 Algorithm for the Robust Design

Nevertheless, the duality transformation to a semi-definite program in [86] provides an insightful way to explore the problem structure. Our intuition is to make full use of the solvable structure through the duality transformation, and design an algorithm to update the decision threshold in its feasible set (4.3c). For the inner minimization problem, we first present an equivalent transformation that eliminates the function integrations associated with the distribution uncertainty, and then turn the max-min problem into a maximization problem as follows:

Theorem 4.1 *The problem (4.3a)-(4.3c) is equivalent to (4.4a)-(4.4d) as follows:*

$$\max_{\boldsymbol{\lambda}, Z, Q, r} \quad (\boldsymbol{\mu}\boldsymbol{\mu}^T - (1 + \gamma_\sigma)\Sigma) \otimes Q - \Sigma \otimes P + 2\boldsymbol{\mu}^T p - \gamma_\mu s - r \quad (4.4a)$$

$$s.t. \quad \boldsymbol{\xi}^T Q \boldsymbol{\xi} - 2\boldsymbol{\xi}^T (p + Q\boldsymbol{\mu}) + r \geq 0, \quad \forall \boldsymbol{\xi} \preceq \boldsymbol{\lambda}, \quad (4.4b)$$

$$\begin{bmatrix} 1 + r & (p + Q\boldsymbol{\mu})^T \\ p + Q\boldsymbol{\mu} & Q \end{bmatrix} \succeq 0, \quad (4.4c)$$

$$\int_{\boldsymbol{\xi} \in S} h(\boldsymbol{\xi}, \boldsymbol{\lambda}) f_0(\boldsymbol{\xi}) d\boldsymbol{\xi} \leq \alpha, \quad (4.4d)$$

where $Z = \begin{bmatrix} P & p \\ p^T & s \end{bmatrix}$ and Q are symmetric positive semi-definite matrices.

The proof of Theorem 4.1 is given in Appendix 4.6. The difficulty imposed by the false alarm probability constraint (4.4d) still exists in the equivalence (4.4a)-(4.4d). This motivates us to separate this non-convex constraint from the other parts of the problem. In other words, we first pick a feasible point $\boldsymbol{\lambda}$ in the non-convex set defined by (4.4d). When we manage to solve the sub-problem (4.4a)-(4.4c), we go back to this non-convex constraint (4.4d), and design specific search methods to improve the objective function. To recap, we solve the sub-problem (4.4a)-(4.4c) with a *fixed* decision threshold vector $\boldsymbol{\lambda}$ that satisfies (4.4d). Then we will iteratively improve the objective function value in (4.4a) by adjusting the decision threshold $\boldsymbol{\lambda}$ in the feasible set $\mathcal{D} = \{\boldsymbol{\lambda} \mid \int_{\boldsymbol{\xi} \in S} h(\boldsymbol{\xi}, \boldsymbol{\lambda}) f_0(\boldsymbol{\xi}) d\boldsymbol{\xi} \leq \alpha\}$. The algorithm continues until a

stable point is reached. Although we cannot prove the global optimality of such a stable point, simulation results show that such result leads to a much more robust performance comparing with the algorithms where the uncertainty is not explicitly taken into consideration.

4.2.1 Initial Setting

In order to solve problem (4.4a)-(4.4c), we first need to set an initial decision threshold (IDT) in the feasible set $\mathcal{D}$ defined by constraint (4.4d). We define the boundary of set $\mathcal{D}$ as $\partial\mathcal{D} = \{\boldsymbol{\lambda} \mid \int_{\boldsymbol{\xi} \leq \boldsymbol{\lambda}} f_0(\boldsymbol{\xi}) d\boldsymbol{\xi} = 1 - \alpha\}$. Theoretically, any feasible point can be the initial point. However we can prove that the optimal decision threshold lies in the boundary set $\partial\mathcal{D}$, which provides a better choice of the initial point.

Proposition 4.1 *The optimal decision threshold $\boldsymbol{\lambda}^*$ to the problem (4.3a)-(4.3c) is obtained in the boundary set $\partial\mathcal{D}$, i.e.,*

$$\int_{\boldsymbol{\xi} \leq \boldsymbol{\lambda}^*} f_0(\boldsymbol{\xi}) d\boldsymbol{\xi} = 1 - \alpha. \quad (4.5)$$

The proof is given in Appendix 4.6. Since there is no need to search the interior of the feasible set $\mathcal{D}$, Proposition 4.1 implies that we can start the algorithm by an IDT that satisfies (4.5). Moreover, we can restrict our search for new decision thresholds on the boundary set $\partial\mathcal{D}$. Note that the boundary set $\partial\mathcal{D}$ is not empty, thus we can always find a decision threshold that satisfies equation (4.5). To quick initialize a decision threshold, we can simply assume an equal IDT that assigns the same value for all SUs' decision thresholds, i.e., $\lambda_1 = \dots = \lambda_N = \lambda_0$. Note that the IDT can also be set in different ways. However, an equal IDT is easy to compute and our simulation results in Section 4.4 show that the convergence of the algorithm is not sensitive to the initial choice. Since the noise signal at each SU's receiver is known as independently and identically distributed, equation (4.5) becomes $\int_{\xi \leq \lambda_0} f_0(\xi) d\xi = (1 - \alpha)^{1/N}$, then the equal IDT leads to the same decision threshold $\lambda_0 = F_0^{-1}((1 - \alpha)^{1/N})$ for all SUs, where $F_0(\lambda_0) = \int_{\xi \leq \lambda_0} f_0(\xi) d\xi$ is the cumulative density function of the noise signal.

4.2.2 Iterative Algorithm

Given a fixed decision thresholds in the boundary set $\partial\mathcal{D}$, we now focus on solving the sub-problem (4.4a)-(4.4c). According to the equivalence between separation and optimization [96,98], an optimization problem is polynomially solvable if we can check whether a given point is feasible, or find a cutting hyperplane that separates an infeasible point from the feasible set, both in polynomial time. Note that the objective (4.4a) is linear and (4.4c) defines a convex positive semi-definite cone, thus it is easy to check their feasibility and find the separating hyperplanes. Therefore, the tractability of sub-problem (4.4a)-(4.4c) only depends on (4.4b), and consequently the answers to the following two questions: (1) For any feasible (Q, p, r) to the problem of (4.4a) subject to (4.4c), can we check the feasibility of constraint (4.4b) in polynomial time? (2) If (Q, p, r) is infeasible for (4.4b), can we find a hyperplane that separates (Q, p, r) from the feasible set of (4.4b)?

To answer these questions, let $\Phi(\boldsymbol{\xi}|Q, p, r) = \boldsymbol{\xi}^T Q \boldsymbol{\xi} - 2\boldsymbol{\xi}^T(p + Q\boldsymbol{\mu}) + r$ be a function of $\boldsymbol{\xi}$ when (Q, p, r) is fixed. Then the first question is equivalent to see whether we can check $\Phi(\boldsymbol{\xi}^*|Q, p, r) = \min_{\boldsymbol{\xi} \preceq \boldsymbol{\lambda}} \Phi(\boldsymbol{\xi}|Q, p, r) \geq 0$ in polynomial time. Since Q is positive semi-definite, the minimization of $\Phi(\boldsymbol{\xi}|Q, p, r)$ over a convex set $\boldsymbol{\xi} \preceq \boldsymbol{\lambda}$ will return a $\boldsymbol{\xi}^*$ and the optimum $\Phi(\boldsymbol{\xi}^*|Q, p, r)$ in polynomial time. When (Q, p, r) is infeasible for (4.4b), we have $\Phi(\boldsymbol{\xi}^*|Q, p, r) < 0$ and can generate a separating plane by $\boldsymbol{\xi}^*(\boldsymbol{\xi}^* - 2\boldsymbol{\mu})^T \otimes Q - 2(\boldsymbol{\xi}^*)^T p + r > 0$. Therefore, the sub-problem (4.4a)-(4.4c) can be solved by the ellipsoid method [98] in polynomial time.

However, the ellipsoid method actually suffers from numerical instability and poor performance in practice. In the following part, we are trying to exploit the problem structure and use it to guide the search for a new decision threshold. We first present a theorem that further simplifies the sub-problem (4.4a)-(4.4c) as follows:

Theorem 4.2 *At the optimum, sub-problem (4.4a)-(4.4c) has the following form:*

$$\max_{\lambda, Q, p, q, r, z} \quad z \quad (4.6a)$$

$$s.t. \quad \lambda(\lambda - 2\mu)^T \otimes Q - 2\lambda^T p + r \geq 0, \quad (4.6b)$$

$$(\mu\mu^T - (1 + \gamma_\sigma)\Sigma) \otimes Q + 2\mu^T p - z - r - q \geq 0, \quad (4.6c)$$

$$q - 2\sqrt{\gamma_\mu} \|Lp\|_2 \geq 0, \quad (4.6d)$$

$$\begin{bmatrix} r+1 & (p+Q\mu)^T \\ (p+Q\mu) & Q \end{bmatrix} \succeq 0. \quad (4.6e)$$

A detailed proof is given in Appendix 4.6. Problem (4.6a)-(4.6e) represents a conic optimization problem in its standard dual form, and can be easily solved by some optimization tool package such as SeDuMi [88]. Given the optimal solution Q , p , q , and r in problem (4.6a)-(4.6e), the following proposition provides an ascend search direction to update the decision threshold.

Proposition 4.2 *If $\lambda(t+1) = \lambda(t) + \Delta_t$ and $\Delta_t^T G(t) > 0$ where $G(t) = Q \cdot \lambda(t) - Q \cdot \mu - p$, then $\lambda(t+1)$ leads to an improved detection probability in problem (4.6a)-(4.6e).*

The proof of Proposition 4.2 is given in Appendix 4.6. Besides the search direction, the magnitude (namely, step size) of Δ_t is also critical in the update of decision threshold. We will come back to this issue in the next section. Based on Theorem 4.2 and Proposition 4.2, we sketch the process of our robust design in Algorithm 2, which terminates when we can no longer find a better update to improve the detection probability.

4.3 Gradient Aided Decision Threshold Update

To complete the algorithm, we are going to address two major steps unsolved in Algorithm 2. In the Line 3 of Algorithm 2, we first need to check whether there exists a feasible ascent search direction such that the objective value can be improved. In Line 6 of Algorithm 2, the second step is to choose the optimal one from all feasible search directions. Optimal search direction means the largest improvement on the objective function, and thus implies the largest projection on gradient direction. For

Algorithm 2 Gradient Aided Threshold Update Algorithm

Input: Sample estimates $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and range parameters γ_μ , γ_σ
Output: Robust threshold $\boldsymbol{\lambda}$ and detection probability z

```

1: initialization:
2:   set initial decision threshold  $\boldsymbol{\lambda}_0$  according to (4.5)
3: while there exists  $\Delta_t$  such that  $\Delta_t^T G(t) > 0$ 
4:   solve problem (4.6a)-(4.6e) with the fixed decision threshold  $\boldsymbol{\lambda}(t)$ 
5:   update detection probability  $z$  and  $G(t+1)$  according to Proposition 4.2
6:   find the optimal ascent search direction  $\Delta_{t+1}^* = \arg \max_{\Delta} \Delta^T G(t+1)$ 
7:   update decision threshold  $\boldsymbol{\lambda}(t+1) = \boldsymbol{\lambda}(t) + \Delta_{t+1}^*$ 
8:   set  $t = t + 1$  for next iteration
9: end while
    
```

the ease of study, we decompose the update of decision threshold into multiple atomic movements that involve only two SUs' updates. We then check all possible update strategies and determine the conditions to ensure the feasibility of those strategies. Focusing on the atomic movement also simplifies the process to find the optimal search direction, which essentially involves an easy-to-solve optimization over the step size given a search direction.

4.3.1 Atomic Movement on Boundary

From (4.5), we note that if the search for a new decision threshold is restricted in the boundary set $\partial\mathcal{D}$, the false alarm probability is always maintained at the same level. In other words, such an update is α -*Preserving* and there are always at least two SUs changing their decision thresholds: one increasing and one decreasing. When more than one SUs increase their decision thresholds, we can decompose the α -Preserving update into multiple atomic movements such that only one SU increases its decision threshold in an atomic movement.

With this decomposition in mind, we only need to study the atomic movement, i.e., we change only two SUs' decision thresholds at each iteration (one increasing and one decreasing). Suppose that SU i increases its decision threshold by Δ_i , i.e., $\lambda_i' = \lambda_i + \Delta_i$. To keep the IDT in the boundary set $\partial\mathcal{D}$, we need to change the decision threshold of

some other SU j in an opposite direction, i.e., $\lambda'_j = \lambda_j - \Delta_j$. Since other SUs' decision thresholds are not altered, α -Preserving implies $F_0(\lambda_i)F_0(\lambda_j) = F_0(\lambda_i + \Delta_i)F_0(\lambda_j - \Delta_j)$. Therefore we have the relation between Δ_j and Δ_i as follows:

$$\Delta_j = \lambda_j - F_0^{-1}(F_0(\lambda_j)F_0(\lambda_i)/F_0(\lambda_i + \Delta_i)). \quad (4.7)$$

When applying this adjustment $(\Delta_i, -\Delta_j)^1$ to the current decision threshold λ , we can ensure that the new decision threshold is still in the boundary set.

4.3.2 Ascent Search Direction

The search for decision threshold in the boundary set should not only preserve the same false alarm probability but also be able to improve the objective function, which requires a *feasible ascent search direction* Δ to possess positive projection on the gradient, i.e., $\Delta^T G(t) > 0$. There are few strategies that can be used to search for a new decision threshold. For any two SUs i and j in an atomic movement, we list all possible strategies and their outcomes, respectively, in the second and third columns of Table 4.1². The third column also shows the relations between Δ_j and Δ_i based on the analysis in Appendix 4.6.5. Then we can identify those cases that lead to

Table 4.1: Strategies given $G_i(t) < G_j(t) < 0$.

Thresholds	Adjustment strategies	Results		Cases
$\lambda_i \geq \lambda_j$	$\lambda'_i = \lambda_i + \Delta_i,$ $\lambda'_j = \lambda_j - \Delta_j$	$\Delta_i > \Delta_j$		case 1
	$\lambda'_i = \lambda_i - \Delta_i,$ $\lambda'_j = \lambda_j + \Delta_j$	$\lambda'_i > \lambda_j$	$\Delta_i > \Delta_j$	case 2
		$\lambda'_i \leq \lambda_j$	$\Delta_i < \Delta_j$	case 3
	$\lambda'_i = \lambda_i - \Delta_i,$ $\lambda'_j = \lambda_j + \Delta_j$	$\Delta_i < \Delta_j$		case 4
$\lambda_i < \lambda_j$	$\lambda'_i = \lambda_i + \Delta_i,$ $\lambda'_j = \lambda_j - \Delta_j$	$\lambda'_j > \lambda_i$	$\Delta_j > \Delta_i$	case 5
		$\lambda'_j < \lambda_i$	$\Delta_j < \Delta_i$	case 6

feasible ascent search directions. Note that there are only two non-zero terms in Δ , e.g., adjustments Δ_i and Δ_j coupled in (4.7), we define the projection on gradient as a function of Δ_i to simplify the analysis. For example, if we have $G_i(t) < G_j(t) < 0$

¹The tuple $(\Delta_i, -\Delta_j)$ represents an N -dimensional zero vector Δ with two non-zero terms, Δ_i and $-\Delta_j$, at the i -th and j -th positions, respectively. Then the new decision threshold will be given by $\lambda + \Delta$.

²We can make a similar table when $0 > G_i(t) > G_j(t)$.

and $\lambda_i \geq \lambda_j$ in case 1, we get $\Delta_i > \Delta_j$ in the third column, and the projection is as follows:

$$g(\Delta_i) = \Delta_i G_i(t) - \Delta_j G_j(t) = \Delta_j(\Delta_i)|G_j(t)| - \Delta_i|G_i(t)| < 0.$$

Therefore, the adjustment $(\Delta_i, -\Delta_j)$ in case 1 will never generate an ascent direction. Similarly, case 6 is also infeasible. While in case 2, we have

$$g(\Delta_i) = \Delta_j G_j(t) - \Delta_i G_i(t) = \Delta_i|G_i(t)| - \Delta_j(\Delta_i)|G_j(t)| > 0,$$

which implies that $(-\Delta_i, \Delta_j)$ is feasible in case 2. However, it is not obvious to find the feasibility condition for cases 3, 4, and 5 in Table 4.1. To resolve this issue, we propose the following sufficient conditions for the existence of a feasible direction:

Proposition 4.3 *Let $i^- = \arg \min_i G_i(t)$, and j be any SU other than i^- , then we have the following results:*

- For cases 3 and 4, $(-\Delta_{i^-}, \Delta_j)$ is feasible if $\Delta_{i^-} < \Delta_j < \left| \frac{G_{i^-}(t)}{G_j(t)} \right| \Delta_{i^-}$.
- For case 5, $(\Delta_{i^-}, -\Delta_j)$ is feasible if $\Delta_j > \left| \frac{G_{i^-}(t)}{G_j(t)} \right| \Delta_{i^-} > \Delta_{i^-}$.

Proof: The proof is straightforward by looking at the projection function, i.e., $g(\Delta_{i^-}) = \Delta_{i^-}|G_{i^-}(t)| - \Delta_j(\Delta_{i^-})|G_j(t)|$ in cases 3 and 4, and $g(\Delta_{i^-}) = -\Delta_{i^-}|G_{i^-}(t)| + \Delta_j(\Delta_{i^-})|G_j(t)|$ in case 5. In either case, a feasible direction requires $g(\Delta_{i^-}) > 0$, which directly leads to the results in the proposition. ■

As Δ_j is a function of Δ_{i^-} , Proposition 4.3 implies that SU i^- should choose Δ_{i^-} properly such that the resulting Δ_j lies in the interval $\left[\Delta_{i^-}, \left| \frac{G_{i^-}(t)}{G_j(t)} \right| \Delta_{i^-} \right]$ for cases 3 and 4, and $\left[\left| \frac{G_{i^-}(t)}{G_j(t)} \right| \Delta_{i^-}, \lambda_j - \lambda_{i^-} \right]^3$ for cases 5.

4.3.3 Optimal Step Size

In Section 4.3.2, we have illustrated different cases and the conditions required to ensure a feasible ascent direction. Here we determine the *optimal step size* $\Delta_{i^-}^*$ for these cases, such that the projection $g(\Delta_{i^-})$ on gradient direction is not only positive,

³Here $\Delta_j < \lambda_j - \lambda_{i^-}$ ensures $\lambda_j' > \lambda_{i^-}$ as requested in case 5 of Table 4.1.

but also maximized, i.e., $\Delta_{i-}^* = \arg \max_{\Delta_{i-}} g(\Delta_{i-})$. Take an example in case 2, we have $g(\Delta_{i-}) = \Delta_{i-}|G_{i-}(t)| - \Delta_j(\Delta_{i-})|G_j(t)|$ and the derivative of $g(\Delta_{i-})$ with respect to Δ_{i-} is given by $g'(\Delta_{i-}) = |G_{i-}(t)| - |G_j(t)| \frac{\partial \Delta_j}{\partial \Delta_{i-}}$ where $\frac{\partial \Delta_j}{\partial \Delta_{i-}}$ denotes the rate of change of Δ_j with respect to Δ_{i-} . Therefore we may obtain the optimal step size by the first order optimality condition. Before characterizing the optimal step size Δ_{i-}^* , we have the following property regarding the projection function.

Proposition 4.4 *The projection $g(\Delta_{i-})$ on gradient direction is a concave function for cases 2 – 5.*

The proof for Proposition 4.4 is given in Appendix 4.6. The concavity guarantees the uniqueness of the optimal step size, and we can find its optimal value Δ_{i-}^* based on the first-order optimality condition, i.e., $g'(\Delta_{i-}^*)(\Delta_{i-} - \Delta_{i-}^*) \leq 0$ for any feasible Δ_{i-} satisfying the condition in Proposition 4.3. Specifically, we have the following results for cases 2 – 5.

Theorem 4.3 *If $g'(0) > 0$, there always exists a unique $\Delta_{i-}^u > 0$ such that $g'(\Delta_{i-}^u) = 0^4$, and the optimal step size Δ_{i-}^* is the projection of Δ_{i-}^u on an interval set $\mathbf{Z}$, denoted by $\Delta_{i-}^* = [\Delta_{i-}^u]_{\mathbf{Z}} \triangleq \arg \min_{z \in \mathbf{Z}} \|\Delta_{i-}^u - z\|_2$. Moreover, we have $\mathbf{Z} = [0, |\lambda_{i-} - \lambda_j|]$ for cases 2 and 5 and $\mathbf{Z} = [0, \lambda_{i-}]$ for cases 3 and 4.*

Proof: Note that $g''(\Delta_{i-}) \leq 0$ as proved in Proposition 4.4, there always exists a unique maximizer $\Delta_{i-}^u > 0$ such that $g'(\Delta_{i-}^u) = 0$ when $g'(0) > 0$ and $g(0) = 0$ (we set Δ_{i-}^u as $+\infty$ if the projection function is strictly increasing). Considering different feasible sets of Δ_{i-} in cases 2 – 5, the optimal step size Δ_{i-}^* will be the projection of Δ_{i-}^u on the feasible set $\mathbf{Z}$. Specifically, we have $\Delta_{i-}^* = \min((g')^{-1}(0), \lambda_{i-} - \lambda_j)$ for case 2, $\Delta_{i-}^* = \min((g')^{-1}(0), \lambda_j - \lambda_{i-})$ for case 5, and $\Delta_{i-}^* = \min((g')^{-1}(0), \lambda_{i-})$ for cases 3 and 4. ■

Theorem 4.3 shows a method of computing the optimal step size. However, solving $g'(\Delta_{i-}) = 0$ involves the manipulation of an implicit function. Thus, an explicit closed-form solution may not be available. The good news is that $g'(\Delta_{i-})$ is monotonically

⁴We take $\Delta_{i-}^u = +\infty$ if there is no solution.

decreasing for cases 2 – 5 as shown in Proposition 4.4. Therefore, the rate of change $\frac{\partial \Delta_j}{\partial \Delta_{i-}}$ is also a monotonic function, which motivates the use of a bisection method to search for the optimal step size. In this way, we avoid finding a direct solution to $g'(\Delta_{i-}) = 0$ as required by Theorem 4.3. The detailed procedures are given in Algorithm 3.

The Algorithm starts in Line 3 by checking whether the sufficient condition $g'(0) > 0$ is satisfied as indicated by Theorem 4.3. Line 4 sets the lower and upper boundary points of the feasible set of Δ_{i-} . Note that the feasible set $\mathbf{Z}$ varies according to different cases. Then an iterative process in lines 5 – 13 shrinks the search region until the optimal step size is found. This optimal value Δ_{i-}^* is either a solution to $g'(\Delta_{i-}) = 0$ or one of the two end points of set $\mathbf{Z}$. When updating the decision thresholds in lines 14 and 15, notation $\oplus$ represents either addition (+) or subtraction (−) according to the cases in Table 4.1.

4.4 Numerical Analysis

In the simulation, we consider $N = 3$ SUs, and their received primary signals are denoted by the vector $\boldsymbol{\mu}$. We set $\boldsymbol{\mu} = [3.7, 4.0, 4.3]$ and the maximum false alarm probability $\alpha = 0.1$. The received primary signals are independent at different SUs with an variance matrix $\Sigma = 2\mathbf{I}_N$ where $\mathbf{I}_N$ is an identity matrix with size N . The distribution uncertainty parameters in set (4.2) are $\gamma_\mu = 0.02$ and $\gamma_\sigma = 0.2$. In the following, we first give an example to illustrate the robust performance in our proposed algorithm and then investigate the algorithm's convergence properties with different initial settings.

4.4.1 Comparison between Robust and Nominal Design

In Figure 4.1, we compare the performance of our robust design with that of the nominal design in literature. The nominal design assumes known distribution functions for the received signals, and assigns the same decision thresholds for all SUs [12, 13, 16]. The optimization the decision thresholds in nominal design is based on a multivariate Gaussian distribution $\mathcal{G}_0$ with mean $\boldsymbol{\mu} = [3.7, 4.0, 4.3]$. However, the actual

Algorithm 3 Update Decision Threshold

Input: Current decision threshold $\boldsymbol{\lambda}(t)$ and gradient direction $G(t) = [G_i(t), \dots, G_N(t)]$

Output: New decision threshold vector $\boldsymbol{\lambda}(t+1)$

```

1: initialization:
2:   initiate  $\boldsymbol{\lambda}(t+1)$  by  $\boldsymbol{\lambda}(t)$ , set  $j \in \mathcal{N}$  and  $j \neq i^-$ 
3: if  $g'(0) > 0$  for cases 2 – 5
4:   set the lower and upper end points as  $\Delta_- = \mathbf{Z}_{min}, \Delta^+ = \mathbf{Z}_{max}$ 
5:   while  $|g'((\Delta_- + \Delta^+)/2)| > \epsilon$ , and  $|\Delta_- - \Delta^+| > \epsilon$ 
6:     set  $\bar{\Delta}$  as the middle point between  $\Delta_-$  and  $\Delta^+$ 
7:     if derivative  $g'(\cdot)$  has different sign at  $\bar{\Delta}$  and  $\Delta_-$ , i.e.,  $g'(\bar{\Delta})g'(\Delta_-) < 0$ 
8:       set the new upper point  $\Delta^+ = \bar{\Delta}$ 
9:     else
10:      set the new lower point  $\Delta_- = \bar{\Delta}$ 
11:    end if
12:  end while
13:  set the optimal step size  $\Delta_{i-}^* = (\Delta_- + \Delta^+)/2$ 
14:  update  $\lambda_{i-}(t+1) = \lambda_{i-}(t) \oplus \Delta_{i-}^*$ 
15:  update  $\lambda_j(t+1) = \lambda_j(t) \oplus \Delta_j$  and  $\Delta_j$  is given by (4.7)
16: end if
    
```

distribution may follow either $\mathcal{G}_0$ or $\mathcal{G}_1$, where $\mathcal{G}_1$ is another multivariate Gaussian distribution with the mean $\boldsymbol{\mu} = [3.0, 3.5, 4.0]$. For our proposed robust design, we take such distribution uncertainty into consideration. As shown in Figure 4.1, the nominal design works very well when the actual distribution of primary signals is $\mathcal{G}_0$ (and thus matches with the assumption). However, when the actual distribution is $\mathcal{G}_1$ (and thus there is a mismatch), the nominal design's performance degrades significantly. While for our proposed robust design, though the detection performance is not as good as that of the nominal design in the matched case, it is more robust and has a much better performance when there exists distribution mismatch. This shows that our proposed algorithm achieves a good tradeoff between robustness and performance. In fact, the optimization under the robust design is performed with respect to the uncertainty set in (4.2), instead on a particular distribution.

Figure 4.1: Performance comparison between the nominal and robust designs.

4.4.2 Algorithm Convergence

We have assumed in Section 4.2.1 that the proposed algorithm starts with the equal IDT. Here we show that the algorithm converges and the final performance is not sensitive to this assumption. To illustrate this, we examine three cases with different IDTs: $\lambda_0^1 = [1.82, 1.82, 1.82]$, $\lambda_0^2 = [1.55, 1.82, 2.42]$, and $\lambda_0^3 = [1.52, 1.82, 2.72]$. Here the superscript denotes different initializations, and the subscript indicates that they are initial values (i.e., iteration index = 0). Note that the false alarm probabilities Q_f of all three initial choices are kept at the same level of $\alpha = 0.1$, i.e., $Q_f = \int_{\xi \in S} h(\xi, \lambda) f_0(\xi) d\xi = \alpha$ for any $\lambda \in \{\lambda_0^1, \lambda_0^2, \lambda_0^3\}$.

We plot the dynamics of the projections on gradient, detection performance, and SUs' decision thresholds in Figs. 4.2-4.4, respectively. In each iteration, our algorithm will check all possible strategies listed in Table 4.1, and then determine a search direction and the optimal step size that has the largest projection on its gradient direction as shown in Figure 4.2. A positive projection always improves the detection probability, and the algorithm hits the optimum when the maximum projection is zero. The dynamics of the detection probability is shown in Figure 4.3, where IDTs are set to λ_0^1 , λ_0^2 and λ_0^3 , respectively. Though the initial detection probabilities are not the same, the differences due to different initializations quickly diminish as the iteration

Figure 4.2: The values of projection on gradient direction decrease in each iteration.

increases. With the decrease of gradient projections, the detection probability gradually converges to the maximal value. In Figure 4.4, each SU's decision threshold also converges to a common value under different initialization choices (and different SUs have different values).

4.4.3 Reverse Ordering between λ and μ

Figure 4.4 also shows that our proposed scheme will assign each SU a different decision threshold to improve robustness, compared with the solution of the nominal scheme. Further investigation on Figure 4.4 reveals a counter-intuitive result that an SU receiving a higher signal strength tends to employ a smaller decision threshold. To study the relation between decision thresholds λ and the received signal strengths μ , we consider changing one SU's received signal strength continuously, and check the convergent decision thresholds in our proposed scheme. Specifically, we fix $\mu_1 = 3.7$ and $\mu_2 = 4.0$ as the same in Section 4.4.2, while changing μ_3 from 3.4 to 4.3. The IDTs are set to $\lambda_0^2 = [1.55, 1.82, 2.42]$ during this process. The dynamics of SUs' final decision thresholds are illustrated in Figure 4.5. We can observe that an SU's decision threshold decreases with the increase of the received signal strength. When $\mu_3 = 4.3$, we obtain the same results as in Figure 4.4 with $\lambda_3 < \lambda_2 < \lambda_1$. There are two intersection points among curves in Figure 4.5, which correspond to the signal

Figure 4.3: Detection probabilities converge to the same level irrespective of different IDT settings.

Figure 4.4: Each SU's decision threshold converges to a common value under different IDT settings (and different SUs have different values).

Figure 4.5: Decision thresholds vary with their received signal strength.

mean values of $\hat{\boldsymbol{\mu}} = [3.7, 4.0, 3.7]$ and $\check{\boldsymbol{\mu}} = [3.7, 4.0, 4.0]$. In the first point $\hat{\boldsymbol{\mu}}$, decision thresholds λ_1 and λ_3 converge to the same value, which is higher than λ_2 . We have a similar observation at $\check{\boldsymbol{\mu}}$ where $\lambda_1 > \lambda_2 = \lambda_3$.

4.5 Summary

In this chapter, we consider the problem of robust cooperative sensing in cognitive radio networks. The key contribution is that we incorporate the distribution uncertainty into the sensing performance optimization, so that our solution is more robust and can improve the worst-case detection performance. As it is very difficult to solve the constrained robust optimization problem directly, we propose an approximate algorithm that separates the false alarm probability constraint from the rest of the problem. More specifically, the algorithm starts from an initial choice of decision thresholds that satisfies the false alarm probability constraint, and solves the rest of the problem in a semi-definite program. Then during each iteration, the algorithm searches new decision thresholds in order to improve the system objective.

4.6 Appendix

4.6.1 Proof for Theorem 4.1

To solve the max-min problem, we first consider the inner minimization problem (4.3a)-(4.3b) with fixed decision threshold $\boldsymbol{\lambda}$. By assigning different dual variables to the constraints in uncertainty set (4.3b), we can write the Lagrangian function of the constrained problem (4.3a)-(4.3b) as $\Lambda(f_1(\boldsymbol{\xi}), Z, Q, r)$

$$\begin{aligned}
 &= \mathbb{E}_{f_1(\boldsymbol{\xi})} [h(\boldsymbol{\xi}, \boldsymbol{\lambda})] - \mathbb{E}_{f_1(\boldsymbol{\xi})} \left[\begin{bmatrix} P & p \\ p^T & s \end{bmatrix} \otimes \begin{bmatrix} \Sigma & \boldsymbol{\xi} - \boldsymbol{\mu} \\ (\boldsymbol{\xi} - \boldsymbol{\mu})^T & \gamma_\mu \end{bmatrix} \right] \\
 &\quad - Q \otimes ((1 + \gamma_\sigma)\Sigma - \mathbb{E}_{f_1(\boldsymbol{\xi})} [(\boldsymbol{\xi} - \boldsymbol{\mu})(\boldsymbol{\xi} - \boldsymbol{\mu})^T]) + r \int_{\boldsymbol{\xi} \in S} f_1(\boldsymbol{\xi}) d\boldsymbol{\xi} - r, \\
 &= \mathbb{E}_{f_1(\boldsymbol{\xi})} [h(\boldsymbol{\xi}, \boldsymbol{\lambda}) + r - 2\boldsymbol{\xi}^T(p + Q\boldsymbol{\mu}) + \boldsymbol{\xi}^T Q \boldsymbol{\xi}] \\
 &\quad + (\boldsymbol{\mu}\boldsymbol{\mu}^T - (1 + \gamma_\sigma)\Sigma) \otimes Q - \Sigma \otimes P + 2\boldsymbol{\mu}^T p - \gamma_\mu s - r.
 \end{aligned} \tag{4.8}$$

Here $\otimes$ denotes the Frobenius inner product, i.e., the component-wise inner product of two matrices. Dual variables Z , Q , and r represent the penalties if the corresponding moment constraints are violated. Therefore we can formulate the dual problem $\max_{Z, Q, r} \min_{f_1(\boldsymbol{\xi})} \Lambda(f_1(\boldsymbol{\xi}), Z, Q, r)$ as follows:

$$\max_{Z, Q, r} \quad (\boldsymbol{\mu}\boldsymbol{\mu}^T - (1 + \gamma_\sigma)\Sigma) \otimes Q - \Sigma \otimes P + 2\boldsymbol{\mu}^T p - \gamma_\mu s - r \tag{4.9a}$$

$$\text{s.t.} \quad h(\boldsymbol{\xi}, \boldsymbol{\lambda}) + r - 2\boldsymbol{\xi}^T(p + Q\boldsymbol{\mu}) + \boldsymbol{\xi}^T Q \boldsymbol{\xi} \geq 0 \quad \forall \boldsymbol{\xi} \in S. \tag{4.9b}$$

The constraint (4.9b) guarantees a finite value for the dual objective (4.9a), which is actually the lower bound of the detection probability in (4.3a). Note that the binary decision function $h(\boldsymbol{\xi}, \boldsymbol{\lambda})$ equals 0 only when each term of $\boldsymbol{\xi}$ is no larger than the corresponding term of the decision threshold $\boldsymbol{\lambda}$. Based on this, we can rewrite (4.9b) into two constraints

$$r - 2\boldsymbol{\xi}^T(p + Q\boldsymbol{\mu}) + \boldsymbol{\xi}^T Q \boldsymbol{\xi} \geq 0, \quad \forall \boldsymbol{\xi} \preceq \boldsymbol{\lambda}, \tag{4.10a}$$

$$1 + r - 2\boldsymbol{\xi}^T(p + Q\boldsymbol{\mu}) + \boldsymbol{\xi}^T Q \boldsymbol{\xi} \geq 0, \quad \forall \boldsymbol{\xi} \not\preceq \boldsymbol{\lambda}, \tag{4.10b}$$

where $\boldsymbol{\xi} \not\preceq \boldsymbol{\lambda}$ means that there exist $n \in \mathcal{N}$ such that $\xi_n > \lambda_n$. Constraint (4.10a) also implies that $1 + r - 2\boldsymbol{\xi}^T(p + Q\boldsymbol{\mu}) + \boldsymbol{\xi}^T Q \boldsymbol{\xi} \geq 0$ for $\boldsymbol{\xi} \preceq \boldsymbol{\lambda}$. Together with (4.10b),

we have that $1 + r - 2\boldsymbol{\xi}^T(p + Q\boldsymbol{\mu}) + \boldsymbol{\xi}^T Q \boldsymbol{\xi} \geq 0$ for all $\boldsymbol{\xi} \in S$, which can be written equivalently in the following matrix form

$$\begin{bmatrix} 1 + r & (p + Q\boldsymbol{\mu})^T \\ p + Q\boldsymbol{\mu} & Q \end{bmatrix} \succeq 0 \quad (4.11)$$

by Schur's complement [99]. To summarize, we can replace constraint (4.9b) by (4.10a) and (4.11), and obtain an equivalent form in (4.4a)-(4.4d). Note that the false alarm probability constraint (4.3c) is left intact as (4.4d) in the new problem.

4.6.2 Proof for Proposition 4.1

The key idea is to show that any decision threshold in the interior of $\mathcal{D}$ will be outperformed by some point in the boundary set $\partial\mathcal{D}$. For any $f_1(\boldsymbol{\xi}) \in \mathcal{U}$, the objective function in (4.3a) can be rewritten as $1 - \int_{\boldsymbol{\xi} \preceq \boldsymbol{\lambda}} f_1(\boldsymbol{\xi}) d\boldsymbol{\xi}$, thus (4.3a) intends to find some $\boldsymbol{\lambda}$ that minimizes $\int_{\boldsymbol{\xi} \preceq \boldsymbol{\lambda}} f_1(\boldsymbol{\xi}) d\boldsymbol{\xi}$, which is monotonically increasing in $\boldsymbol{\lambda}$ (as $f_1(\boldsymbol{\xi}) > 0$). Therefore, for any interior point $\boldsymbol{\lambda}$ (which achieves the strict inequality in (4.4d)), we can immediately decrease the objective $\int_{\boldsymbol{\xi} \preceq \boldsymbol{\lambda}} f_1(\boldsymbol{\xi}) d\boldsymbol{\xi}$ by decreasing some SU's decision threshold, i.e., $\lambda'_i = \lambda_i - \Delta_i$ and $\lambda'_j = \lambda_j$ for $j \neq i$. Note that constraint (4.4d) defines the maximum false alarm probability, which actually defines the maximum value for the decrement Δ_i , thus $\boldsymbol{\lambda}'$ is attainable on the boundary set. Also note that (4.5) is a necessary but not sufficient condition for the optimality of Problem (4.3a)-(4.3c). Given $\boldsymbol{\lambda}' \in \partial\mathcal{D}$, we can find infinitely many $\boldsymbol{\lambda}'' \in \partial\mathcal{D}$ by increasing λ'_i and decreasing λ'_j ($j \neq i$) simultaneously.

4.6.3 Proof for Theorem 4.2

In sub-problem (4.4a)-(4.4c), we aim to find an optimal decision threshold $\boldsymbol{\lambda}$ together with the dual variables Z , Q and r , so as to maximize the objective z in (4.4a). Let $z = (\boldsymbol{\mu}\boldsymbol{\mu}^T - (1 + \gamma_\sigma)\Sigma) \otimes Q - \Sigma \otimes P + 2\boldsymbol{\mu}^T p - \gamma_\mu s - r$. As $Z = \begin{bmatrix} P & p \\ p^T & s \end{bmatrix} \succeq 0$, we discuss two cases for the variable s as in [86]: $s > 0$ or $s = 0$. Considering the first case of $s > 0$, we have $P \succeq \frac{1}{s}pp^T$ and

$$\begin{aligned} z &\leq (\boldsymbol{\mu}\boldsymbol{\mu}^T - (1 + \gamma_\sigma)\Sigma) \otimes Q - \frac{1}{s}p^T \Sigma p + 2\boldsymbol{\mu}^T p - \gamma_\mu s - r \\ &\leq (\boldsymbol{\mu}\boldsymbol{\mu}^T - (1 + \gamma_\sigma)\Sigma) \otimes Q + 2\boldsymbol{\mu}^T p - r - 2\sqrt{\gamma_\mu} \|Lp\|_2. \end{aligned} \quad (4.12)$$

Here L is a triangular matrix denoting the Cholesky decomposition of Σ , i.e., $\Sigma = L^T L$. Therefore, the objective in (4.4a) can be represented as the new objective (4.6a) together with a linear constraint (4.6c) and a quadratic constraint (4.6d). Considering the second case of $s = 0$, we must have $p = 0$ by the positive semidefiniteness of $Z = \begin{bmatrix} P & p \\ p^T & 0 \end{bmatrix} \succeq 0$, otherwise ($p \neq 0$) we can construct a vector y by appending any scalar $x > \frac{p^T P p}{2p^T p}$ to the vector p such that

$$y^T Z y = \begin{bmatrix} p \\ x \end{bmatrix}^T \begin{bmatrix} P & p \\ p^T & 0 \end{bmatrix} \begin{bmatrix} p \\ x \end{bmatrix} = p^T P p - 2p^T p x < 0.$$

Therefore, matrix Z with $p \neq 0$ is no longer positive semidefinite, which contradicts with $Z \succeq 0$. When both s and p equal to zero, we further have $P = 0$ in order to maximize the objective in (4.4a). Therefore the new objective degenerates to $z = (\mu\mu^T - (1 + \gamma_\sigma)\Sigma) \otimes Q - r$, which has exactly the same form as (4.12). For simplicity, we combine the resulting formulations from these two cases (i.e., $s > 0$ and $s = 0$) into a general form (4.6a)-(4.6e). In the following, we prove the equivalence between (4.4b) and (4.6b) when the decision threshold λ achieves its optimum.

Note that λ only appears in constraint (4.4b), which is equivalent to see whether $\Phi(\xi^*|Q, p, r) \geq 0$, where $\xi^* = \arg \min_{\xi \preceq \lambda} \Phi(\xi|Q, p, r)$. If ξ^* is optimal to the minimization of $\Phi(\xi|Q, p, r)$ and $\xi_i^* = \lambda_i$ for some $i \in \mathcal{N}$, we say that the i^{th} term of the box constraint (i.e., $\xi \preceq \lambda$)⁵ is active. If $\xi_i^* < \lambda_i$, then this term is inactive. We can consider several different cases:

- Case 1: All terms of the box constraint are inactive. This implies $\xi^* = Q^{-1}p + \mu \prec \lambda$ and $\Phi(\xi^*|Q, p, r) = r - (p + Q\mu)^T Q^{-1}(p + Q\mu) \geq 0$. In this case we have

$$\begin{aligned} z &\leq (\mu\mu^T - (1 + \gamma_\sigma)\Sigma) \otimes Q + 2\mu^T p - r - 2\sqrt{\gamma_\mu} \|Lp\|_2 \\ &\leq -(1 + \gamma_\sigma)\Sigma \otimes Q - 2\sqrt{\gamma_\mu} \|Lp\|_2 - p^T Q^{-1} p \leq 0. \end{aligned}$$

Therefore λ can not be the optimizer since negative detection probability is meaningless in practice.

⁵Box constraint refers to a series of simple inequality constraints on individual variables, e.g., $L_i \leq x_i \leq U_i, \forall i \in \{1, \dots, N\}$.

- Case 2: Some but not all terms of the box constraint are active. Hence there is at least one active term i and one inactive term $j \neq i$. Let $\mathcal{N}_A$ denote the set of active terms, i.e., $\mathcal{N}_A = \{i \in \mathcal{N} | \xi_i^* = \lambda_i\}$, and $\mathcal{N}_{\bar{A}} = \mathcal{N} - \mathcal{N}_A$ be the set of inactive terms. According to Proposition 4.1, we have $\boldsymbol{\lambda} \in \partial\mathcal{D}$ if the current decision threshold $\boldsymbol{\lambda}$ is the optimum. Meanwhile, we can construct an interior point $\boldsymbol{\lambda}' \in \mathcal{D} - \partial\mathcal{D}$ (i.e., strict inequality holds at (4.4d) with $\boldsymbol{\lambda}'$) by properly increasing λ_j for some $j \in \mathcal{N}_{\bar{A}}$. Note that $j \in \mathcal{N}_{\bar{A}}$ is an inactive term, hence this interior point $\boldsymbol{\lambda}'$ will not change the objective. However, given this point $\boldsymbol{\lambda}'$, we can immediately construct the third decision threshold $\boldsymbol{\lambda}'' \in \partial\mathcal{D}$ that improves the detection probability according to the proof of Proposition 4.1. Therefore, the current decision threshold $\boldsymbol{\lambda}$ is not an optimal yet.
- Case 3: We reach the conclusion that all terms of the box constraint must be active when decision threshold $\boldsymbol{\lambda}$ is optimal. In this case we have $\boldsymbol{\xi}^* = \boldsymbol{\lambda} \preceq \boldsymbol{\mu} + Q^{-1}\mathbf{p}$ and $\Phi(\boldsymbol{\xi}^* | Q, \mathbf{p}, r) = \boldsymbol{\lambda}^T Q \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^T (\mathbf{p} + Q\boldsymbol{\mu}) + r$. Therefore, we remove the free variable $\boldsymbol{\xi}$ from sub-problem (4.4a)-(4.4c) and reduce (4.4b) to a linear constraint with respect to $(Q, \mathbf{p}, r)$.

4.6.4 Proof for Proposition 4.2

We first prove the differentiability of the problem (4.6a)-(4.6e) with respect to $\boldsymbol{\lambda}$. Considering problem (4.6a)-(4.6e) as a parametric optimization problem, the decision threshold $\boldsymbol{\lambda}$ serves as the parameter that controls the feasible region $\mathcal{C}(\boldsymbol{\lambda})$ of decision variable $\mathbf{x} = (Q, \mathbf{p}, q, r, z)$, i.e., $\mathbf{x} \in \mathcal{C}(\boldsymbol{\lambda})$. The proposed iterative algorithm decomposes the parametric optimization problem into a sub-problem (4.6a)-(4.6e) with fixed decision threshold $\boldsymbol{\lambda}$, and a master problem that maximizes the optimal value function $z^*(\boldsymbol{\lambda}) = \max_{\mathbf{x} \in \mathcal{C}(\boldsymbol{\lambda})} z$, by updating decision threshold $\boldsymbol{\lambda}$ based on $\mathbf{x}(\boldsymbol{\lambda}) = \arg \max_{\mathbf{x} \in \mathcal{C}(\boldsymbol{\lambda})} z$. For any fixed $\boldsymbol{\lambda}$ satisfying (4.5), it is easy to check that the sub-problem (4.6a)-(4.6e) is a convex problem and can be solved efficiently by interior point methods. It was proved in [100] that the optimal value function $z^*(\boldsymbol{\lambda})$ in such a bi-level decomposition is continuously differentiable with respect to $\boldsymbol{\lambda}$.

Then we show that $G(t) = Q\lambda(t) - Q\mu - p$ is a gradient direction of λ and can improve the objective (4.6a) in iteration $t + 1$. For the ease of presentation, we denote $\mathbf{x}(t) = (Q_t, p_t, q_t, r_t, z_t)$. From (4.6b)-(4.6c), the optimum $z^*(\lambda(t))$ can be set to $\lambda(t)^T Q_t \lambda(t) - 2\lambda(t)^T (p_t + Q_t \mu) + (\mu \mu^T - (1 + \gamma_\sigma) \Sigma) \otimes Q_t + 2\mu^T p_t - q_t$, where the last three terms are independent of $\lambda(t)$ and can be viewed as constants. Now, considering $z^*(\lambda(t))$ as a function of $\lambda(t)$ with gradient $G(t) = Q_t \lambda(t) - Q_t \mu - p_t$, we can increase the objective by updating the decision threshold as $\lambda(t+1) = \lambda(t) + \Delta_t$ with $\Delta_t^T G(t) > 0$. Moreover, we still have $\mathbf{x}(t) \in \mathcal{C}(\lambda(t+1))$, which also implies $z^*(\lambda(t+1)) \geq z^*(\lambda(t))$ in the next iteration.

4.6.5 The Relation Between Δ_i and Δ_j in Table 4.1

Note that Δ_j is a function of λ and Δ_i as in (4.7). To quantify the relation between Δ_j and Δ_i , we have the following proposition that explain the third column in Table 4.1.

Proposition 4.5 *Assuming $\lambda_i > \lambda_j$, and the adjustment $\Delta_i > 0$ and $\Delta_j > 0$, then we have the following statements if the new decision threshold λ' is in the boundary set $\partial\mathcal{D}$:*

- (i) *If $\lambda'_i = \lambda_i + \Delta_i$ and $\lambda'_j = \lambda_j - \Delta_j$, then we have $\Delta_i > \Delta_j$.*
- (ii) *If $\lambda'_i = \lambda_i - \Delta_i$ and $\lambda'_j = \lambda_j + \Delta_j$, then we either have $\Delta_i < \Delta_j$ when $\lambda_j > \lambda'_i$, or have $\Delta_i > \Delta_j$ when $\lambda_j < \lambda'_i$.*

We prove the proposition by constructing a monotonic function, based on which the relation between SUs' decision thresholds and their adjustments are derived. The search in the boundary set implies the α -Preserving property, i.e., $F_0(\lambda_i)F_0(\lambda_j) = F_0(\lambda'_i)F_0(\lambda'_j)$. Given the assumption in (i), we have

$$\frac{F_0(\lambda_i)}{F_0(\lambda_i + \Delta_i)} = \frac{F_0(\lambda_j - \Delta_j)}{F_0(\lambda_j)} = \frac{F_0(\lambda'_j)}{F_0(\lambda'_j + \Delta_j)}.$$

Let $R(x, a) = F_0(x)/F_0(x + a)$ be a function of $x > 0$ with parameter $a > 0$. We can show that $R(x, a)$ is an increasing function with respect to x since

$$\frac{\partial R(x, a)}{\partial x} = \frac{f_0(x)F_0(x + a) - f_0(x + a)F_0(x)}{F_0^2(x + a)} > \frac{f_0(x) - f_0(x + a)}{F_0(x + a)} > 0,$$

where $f_0(x)$ is the standard normal distribution function. When $\lambda_i > \lambda'_j$, we have $R(\lambda_i, \Delta_i) > R(\lambda'_j, \Delta_i)$, that is, $\frac{F_0(\lambda'_j)}{F_0(\lambda'_j + \Delta_j)} = \frac{F_0(\lambda_i)}{F_0(\lambda_i + \Delta_i)} > \frac{F_0(\lambda'_j)}{F_0(\lambda'_j + \Delta_i)}$. Consequently, $F_0(\lambda'_j + \Delta_j) < F_0(\lambda'_j + \Delta_i)$ and $\Delta_j < \Delta_i$, since $F_0(x)$ is also increasing function with respect to x . Similarly, we can obtain the results in 2) when $\lambda'_i = \lambda_i - \Delta_i$ and $\lambda'_j = \lambda_j + \Delta_j$.

4.6.6 Proof for Proposition 4.4

To prove the concavity, we need to show $g''(\Delta_i) < 0$. Since $G_i(t)$ and $G_j(t)$ are known at iteration t , function $g''(\Delta_i)$ only depends on the second-order derivative of Δ_j with respect to Δ_i . In cases 2 – 4, we decrease the decision threshold of SU i and increase that of SU j . Then the projection function has the form of $g(\Delta_i) = \Delta_i |G_i(t)| - \Delta_j |G_j(t)|$, and the rate of change of Δ_j is given by $\frac{\partial \Delta_j}{\partial \Delta_i} = \frac{f_0(\lambda_i - \Delta_i)/F_0(\lambda_i - \Delta_i)}{f_0(\lambda_j + \Delta_j)/F_0(\lambda_j + \Delta_j)}$. After some manipulation, we can show that the second-order derivative of Δ_j with respect to Δ_i satisfying the equation as follows:

$$A \cdot \partial^2 \Delta_j / \partial \Delta_i^2 = 2 - f'_0(\lambda'_j)F_0(\lambda'_j)/f_0^2(\lambda'_j) - f'_0(\lambda'_i)F_0(\lambda'_i)/f_0^2(\lambda_i^2),$$

where $A = \frac{F_0^2(\lambda'_i)f_0(\lambda'_j)}{F_0(\lambda'_j)f_0^2(\lambda'_i)} > 0$, and λ'_i, λ'_j are the updated decision thresholds for SUs i and j , respectively. Since $f_0(\cdot)$ is the density function of Gaussian noise, we have $f'_0(\lambda'_i) < 0$ and $f'_0(\lambda'_j) < 0$. Consequently, we have $\partial^2 \Delta_j / \partial \Delta_i^2 > 0$, which implies the concavity of $g(\Delta_i)$ since $g''(\Delta_i) = -\partial^2 \Delta_j / \partial \Delta_i^2 < 0$. In case 5, we have $g(\Delta_i) = \Delta_j |G_j(t)| - \Delta_i |G_i(t)|$, and we can prove in a similar way that $\partial^2 \Delta_j / \partial \Delta_i^2 < 0$, which also implies the concavity of the projection function $g(\Delta_i)$.

Chapter 5

Robust Power Control with Distribution Uncertainty

In an underlay cognitive radio network, the most important constraint for SUs' power control is to accurately restrict the interference at PU receivers. That means, the SUs have to estimate the worst-case aggregate interference power (AIP) at PU receivers. However, it is unlikely for SUs to have precise channel information without information exchange between PUs and SUs. To describe the uncertainty in AIP estimations, we resort to the reference-based uncertainty model proposed in Chapter 3. That is, we are able to extract a closed-form reference distribution for AIP based on historical channel measurements, and we characterize the difference between the actual AIP distribution and its reference by the KL divergence.

The resulting power control problem is formulated into a chance constrained robust optimization problem that takes distribution function as the uncertain variable, and we develop two iterative algorithms to search for the optimal transmit power in two different cases. In case I, all SUs have the same transmit power. We first find the worst-case AIPs at PU receivers with fixed transmit power by iteratively solving the KKT condition. Then we find that the worst-case AIP is a monotonic function of the transmit power, therefore we can update the transmit power using a bisection method until the interference requirement is met at PU receivers. In case II, each user may choose its own transmit power. We propose an approximate primal-dual method to decompose the original problem into primal and dual sub-problems. When solving the primal problem, it becomes a non-convex problem and we propose the concave-convex

procedure to search for its optimum. Simulation results show that the power control in case II provides better quality of service for SUs than that in case I, however, at the cost of higher computational complexity.

The remainder of this chapter is organized as follows. Section 5.1 describes the network model, time-varying channel and the distribution uncertainty models. Section 5.2 and 5.3 present two iterative algorithms for the power control in two different cases, respectively. In Sections 5.4 and 5.5, we present some numerical results and summarize the whole work.

5.1 System Model

We consider infrastructure-based primary and secondary networks. As shown in Figure 5.1, there is one primary network (e.g., TV service) with M PUs spatially distributed in the coverage of primary base station (PBS), and there are K secondary access points (SAPs) serving N SUs in an underlay manner, sharing the same spectrum band with the primary network simultaneously conditioned on limited interference to PUs. Both primary and secondary networks mainly provide down-link data streaming services, e.g., file downloading and video streaming. Therefore, data transmissions are mainly from the PBS and SAPs to PUs and SUs, respectively. The sets of PUs, SUs and SAPs are denoted by $\mathcal{M}$, $\mathcal{N}$ and $\mathcal{K}$, respectively. Let $\mathcal{N}_k$ be the set of SUs associated with SAP k . Assume $\mathcal{N} = \bigcup_{k \in \mathcal{K}} \mathcal{N}_k$ and $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$ for any $i, j \in \mathcal{K}$ and $i \neq j$.

There is no direct information exchange between SUs and PUs, but a common control channel among SAPs to coordinate with each other and share channel sensing information [101]. After information sharing and processing, the SAPs perform necessary adjustments on transmit power to maximize overall performance of the secondary network, as well as to prevent excessive interference to the primary network. PUs and SUs can be mobile users with relatively low mobility as compared with the convergence speed of a power control algorithm. That is, user mobility does not incur a new round of power adjustments before the convergence of previous adjustments.

Figure 5.1: Primary and secondary networks.

5.1.1 Time-varying Channel Model

In wireless communications, the channel gain is a composite effect of small-scale fading and large-scale path loss, and time-varying as demonstrated in [76]. Over a small time period with fixed physical channel conditions, the channel effects can be captured as a stationary process and we can characterize AIP and SINR by stationary distribution functions, which may change correspondingly if physical channel conditions have changed. Therefore, we model the time-varying channel effects as concatenations of piecewise stationary processes [102]. Each stationary process corresponds to a channel state S_z that is defined by the stationary distributions of AIP ϕ_m for $m \in \mathcal{M}$ and SINR γ_n for $n \in \mathcal{N}$ (Specific forms of AIP ϕ_m and SINR γ_n are to be explained in next section), i.e., $S_z = (f_{\phi_1}^z(x), \dots, f_{\phi_M}^z(x), f_{\gamma_1}^z(x), \dots, f_{\gamma_N}^z(x))$ where $f_{\phi_m}^z(x)$ and $f_{\gamma_n}^z(x)$ denote the pdfs of AIP ϕ_m and SINR γ_n , respectively for state z . Variations of the physical channel conditions result in the transitions between different states. In Figure 5.2, we present a 2-state example. A LOS path between SAP and PU m may exist (e.g., state S_1) for some time and the channel gain follows a Rician distribution, which is different from the Rayleigh distribution when direct transmissions are blocked by surrounding obstructions (e.g., state S_2). Therefore, distributions $f_{\phi_m}^1(x)$ and $f_{\phi_m}^2(x)$ take different forms since AIP ϕ_m is a composite random variable of the channel gain.

5.1.2 Stochastic PU Protection and QoS Provisioning

For each state in Figure 5.2 during a static period, let g_{km} denote the channel gain from SAP k to PU m , and h_{kn} the channel gain from SAP k to SU n . When SAP k

Figure 5.2: Non-stationary channel model with two states.

transmits with power p_k , the AIP at PU $m \in \mathcal{M}$ is given by $\phi_m = \sum_{k=1}^K p_k g_{km}$, and the instantaneous SINR at SU $n \in \mathcal{N}_k$ is $\gamma_n = \frac{p_k h_{kn}}{\sigma_n^2 + \pi_n + \sum_{s=1, s \neq k}^K p_s h_{sn}}$, where σ_n^2 is the noise power received by SU n and π_n denotes the PBS' interference to SU n . Note that the SINR and AIP are composite random variables of the channel gains $\{h_{kn}\}$ and $\{g_{km}\}$, respectively. We characterize PU's interference by an outage probability that AIP is greater than a threshold $\bar{\phi}_m$. Define the stochastic PU protection as the outage probability, for each PU $m \in \mathcal{M}$, to be bounded by a maximum acceptable value η_m , i.e., $\mathbb{P}\{\phi_m \geq \bar{\phi}_m\} \leq \eta_m$, where $\bar{\phi}_m$ is the maximum interference acceptable by a PU receiver (i.e., interference threshold) and $\mathbb{P}\{A\}$ denotes the probability of event A . For an SU n , we consider its transmission failed if the received SINR γ_n is less than a minimum QoS threshold $\bar{\gamma}_n$. Correspondingly, we define the service rate of each SU as the probability of its successful reception, i.e., $\mathbb{P}\{\gamma_n \geq \bar{\gamma}_n\}$. Then, the stochastic QoS provisioning for SUs is to maximize the average service rate of all SUs, subject to a stochastic PU protection level:

$$\max_{\mathbf{p}} \quad \sum_{n \in \mathcal{N}} \omega_n \mathbb{P}\{\gamma_n \geq \bar{\gamma}_n\} \quad (5.1a)$$

$$s.t. \quad \mathbb{P}\{\phi_m \geq \bar{\phi}_m\} \leq \eta_m, \quad m \in \mathcal{M}, \quad (5.1b)$$

$$0 \leq p_k \leq \bar{p}_k, \quad k \in \mathcal{K}, \quad (5.1c)$$

where $\mathbf{p} = [p_1, p_2, \dots, p_K]$ denotes the transmit power of SAPs, and $\bar{p}_k$ is the maximum transmit power of SAP k . Different weights $\{\omega_n\}_{n \in \mathcal{N}}$ represent service priorities associated with different SUs. Most existing works address the problem (5.1a)-(5.1c) with known channel states and look for closed-form approximations for the distributions of AIP and SINR so as to convert problem (5.1a)-(5.1c) into a convex form [79]. However, it is a challenging issue since SAPs are in general unable to detect the state transitions due to their hardware-constrained sensing capabilities. It usually requires

sufficient data samples to have accurate channel estimation, which is often unavailable due to limited sensing time and sample size. Besides, the estimation of AIP requires SAPs to listen to the feedback from PU receivers, which are generally sporadic and cannot provide real-time channel information.

5.1.3 Reference-based Distribution Uncertainty

The SAPs' unawareness to channel state transitions requires SAPs to use a power strategy that is robust against the distribution uncertainty. Robustness implies that once we have a feasible power strategy for state S_1 as in Figure 5.2, it should still be able to provide the required PU protection level when channel state transits to S_2 . To achieve this, we resort to the reference-based uncertainty model. Specifically, through channel measurement, we can obtain closed-form approximations for the AIP and SINR distributions by goodness-of-fit tests. Their actual distributions can differ from the references in the next sensing period. Let $f_{\gamma_n}^0(x)$ and $f_{\phi_m}^0(x)$ be the reference distributions of SINR γ_n and AIP ϕ_m , respectively, and simply denote $f_{\gamma_n}^z(x)$ (or $f_{\phi_m}^z(x)$, respectively) as $f_{\gamma_n}(x)$ (or $f_{\phi_m}(x)$, respectively) since channel state z is uncertain to SAPs. We quantify the difference between the actual distribution $f_{\gamma_n}(x)$ and its reference $f_{\gamma_n}^0(x)$ by $D_{KL}(f_{\gamma_n}(x), f_{\gamma_n}^0(x)) = \mathbb{E}_{f_{\gamma_n}} [\ln f_{\gamma_n}(x) - \ln f_{\gamma_n}^0(x)]$. Here, $f_{\gamma_n}(x)$ represents the real distribution of data through long term observation and precise modeling, $f_{\gamma_n}^0(x)$ is a closed-form approximation based on theoretic assumptions and simplifications. Then, we can define the distribution uncertainty for γ_n as

$$\mathbb{Z}_{\gamma_n}(f_{\gamma_n}^0(x), D_{\gamma_n}) = \{f_{\gamma_n}(x) \mid D_{KL}(f_{\gamma_n}, f_{\gamma_n}^0) \leq D_{\gamma_n}\},$$

where D_{γ_n} represents a distance limit that bounds the probabilistic difference between $f_{\gamma_n}^0(x)$ and $f_{\gamma_n}(x)$. It can be set properly based on historical channel measurements. Similarly, we can define the distribution uncertainty $\mathbb{Z}_{\phi_m}(f_{\phi_m}^0(x), D_{\phi_m})$ for AIP ϕ_m . In practice, we can set both $f_{\gamma_n}^0(x)$ and $f_{\phi_m}^0(x)$ as log-normal distributions [79] and update them online if real-time measurements indicate large deviations, in terms of KL divergence, from their reference distributions.

5.2 Case I: SAPs Transmit with the Same Power

Given the uncertainty sets $\{\mathbb{Z}_{\gamma_n}\}_{n \in \mathcal{N}}$ and $\{\mathbb{Z}_{\phi_m}\}_{m \in \mathcal{M}}$, the robust power control problem in (5.1a)-(5.1c) is rewritten as:

$$\max_{\mathbf{p}} \min_{f_{\gamma} \in \mathbb{Z}_{\gamma}} \sum_{n \in \mathcal{N}} \omega_n \mathbb{E}_{f_{\gamma_n}} [\mathbf{1}(x \geq \bar{\gamma}_n)] \quad (5.2a)$$

$$s.t. \max_{f_{\phi_m} \in \mathbb{Z}_{\phi_m}} \mathbb{E}_{f_{\phi_m}} [\mathbf{1}(x \geq \bar{\phi}_m)] \leq \eta_m, \quad m \in \mathcal{M}, \quad (5.2b)$$

$$\mathbf{0} \preceq \mathbf{p} \preceq \bar{\mathbf{p}}, \quad (5.2c)$$

where $f_{\gamma}(\mathbf{x})$ is a joint distribution function of $\{\gamma_n\}$, $n \in \mathcal{N}$ and $\mathbb{Z}_{\gamma} \triangleq \prod_{n \in \mathcal{N}} \mathbb{Z}_{\gamma_n}$. $\mathbf{1}(A)$ is an indicator function which equals 1 if event A is true and 0 otherwise. The service rate in (5.1a) and outage probability in (5.1b) are represented by expectation functions (denoted by $\mathbb{E}[\cdot]$) averaged over distributions $f_{\gamma_n}(x)$ and $f_{\phi_m}(x)$, respectively. Our objective is to optimize the transmit power at each SAP, so as to maximize the worst-case QoS provisioning for SUs, while maintaining prescribed PU protection level even under worst-case channel conditions.

5.2.1 Protection for Primary Network

Note that constraint (5.2b) requires every PU, $m \in \mathcal{M}$, to be protected by a probability η_m . When all PUs have the same protection levels, i.e., $\eta_1 = \dots = \eta_M = \eta$, we only need to focus on the most vulnerable PU, which experiences a worst-case channel condition [103]. If the most vulnerable PU is protected by η , other PUs can be better protected. Let $\tilde{m}$ denote the most vulnerable PU, constraint (5.2b) is equivalent to

$$\max_{f_{\phi_{\tilde{m}}}} \mathbb{E}_{f_{\phi_{\tilde{m}}}} [\mathbf{1}(x \geq \bar{\phi}_{\tilde{m}})] \leq \eta. \quad (5.3)$$

Constraint (5.2b) or (5.3) is suitable for per-node based power control [104] and data packet is delivered through multiple hops. Within one hop distance, two neighboring SU transceivers can transmit with very low power so as to limit their interference to the most vulnerable PU receiver.

On the other hand, this constraint is rather conservative in our broadcast model where the data streaming is mainly from SAPs to SUs [105]. Each SAP tries to

maximize its coverage area and serve more SUs. But once a PU receiver moves very close to an SAP and experiences high interference, constraint (5.3) will prevent the SAP from serving all associated SUs. For better service in the secondary network, we emphasize on the interference to the whole primary network, rather than to an individual receiver. That is, though PUs may experience different protection levels, the average protection for all PUs is maintained at a constant level η . Therefore, we define a new interference constraint as

$$\max_{f_\phi \in \mathbb{Z}_\phi} \frac{1}{M} \sum_{m \in \mathcal{M}} \mathbb{E}_{f_{\phi_m}} [\mathbf{1}(x \geq \bar{\phi}_m)] \leq \eta, \quad (5.4)$$

where $f_\phi(\mathbf{x})$ is the joint distribution function of $\{\phi_m\}$, $m \in \mathcal{M}$ and $\mathbb{Z}_\phi \triangleq \prod_{m \in \mathcal{M}} \mathbb{Z}_{\phi_m}$. Assuming ϕ_m are independent random variables at different PUs, (5.4) is equivalent to

$$\frac{1}{M} \sum_{m \in \mathcal{M}} \max_{f_{\phi_m}} \mathbb{E}_{f_{\phi_m}} [\mathbf{1}(x \geq \bar{\phi}_m)] \leq \eta.$$

Here, we consider all SAPs transmitting with the same power level, i.e., $p_1 = \dots p_K = p_s$. As a result, the robust power control with constraint (5.4) has only one control variable. It is easy to prove that both objective (5.2a) and the LHS of (5.4) are increasing functions of the transmit power p_s . Therefore, optimal transmit power p_s can be uniquely determined by the equality condition of constraint (5.4). In order to check whether constraint (5.4) is satisfied with a given transmit power p_s , we need to solve, for each $m \in \mathcal{M}$, a robust optimization problem as follows:

$$\max_{f_{\phi_m}} \quad \mathbb{E}_{f_{\phi_m}} [\mathbf{1}(x \geq \bar{\phi}_m)] \quad (5.5a)$$

$$s.t. \quad \mathbb{E}_{f_{\phi_m}} [\ln f_{\phi_m}(x) - \ln f_{\phi_m}^0(x)] \leq D_{\phi_m}, \quad (5.5b)$$

$$\mathbb{E}_{f_{\phi_m}} [1] = 1. \quad (5.5c)$$

Constraint (5.5b) defines the distribution uncertainty of AIP ϕ_m , and (5.5c) restricts $f_{\phi_m}(x)$ to be a valid probability density function. Though we focus on constraint (5.4), the formulation in (5.5a)-(5.5c) can also be used to study (5.3).

5.2.2 Worst-case Interference Power

Let $\eta_m^w(p_s) = \max_{f_{\phi_m}} \mathbb{E}_{f_{\phi_m}}[\mathbf{1}(x \geq \bar{\phi}_m)]$ denote the worst-case interference at PU receiver m when SAPs transmit with power p_s . By the Lagrangian method, we have

$$\eta_m^w(p_s) = \min_{\tau_m, \lambda_m} \max_{f_{\phi_m}} \mathbb{E}_{f_{\phi_m}} [\mathbf{1}(x \geq \bar{\phi}_m) - \lambda_m] + \lambda_m - \tau_m \mathbb{E}_{f_{\phi_m}} [\ln f_{\phi_m}(x) - \ln f_{\phi_m}^0(x)] + \tau_m D_{\phi_m},$$

where non-negative τ_m and λ_m denote the Lagrangian multipliers associated with constraints (5.5b) and (5.5c), respectively. Let

$$\mathcal{P}(p_s, f_{\phi_m}, \tau_m, \lambda_m) = \mathbb{E}_{f_{\phi_m}} \left[\mathbf{1}(x \geq \bar{\phi}_m) - \lambda_m - \tau_m \ln \frac{f_{\phi_m}(x)}{f_{\phi_m}^0(x)} \right],$$

the derivative of $\mathcal{P}(p_s, f_{\phi_m}, \tau_m, \lambda_m)$ with respect to f_{ϕ_m} is given by

$$\frac{\partial \mathcal{P}}{\partial f_{\phi_m}} = \int_{x \in S} \left(\mathbf{1}(x \geq \bar{\phi}_m) - \tau_m \ln \frac{f_{\phi_m}(x)}{f_{\phi_m}^0(x)} - \tau_m - \lambda_m \right) dx,$$

where S denotes the set of all possible observations of AIP ϕ_m . By the Karush-Kuhn-Tucker (KKT) condition, we have

$$\mathbf{1}(x \geq \bar{\phi}_m) - \tau_m \ln \frac{f_{\phi_m}(x)}{f_{\phi_m}^0(x)} - \tau_m - \lambda_m = 0, \quad (5.6a)$$

$$\int_{x \in S} f_{\phi_m}(x) dx = 1, \quad (5.6b)$$

$$D_{\phi_m} - \mathbb{E}_{f_{\phi_m}} \left[\ln \frac{f_{\phi_m}(x)}{f_{\phi_m}^0(x)} \right] \geq 0, \quad (5.6c)$$

$$\tau_m \left(D_{\phi_m} - \mathbb{E}_{f_{\phi_m}} \left[\ln \frac{f_{\phi_m}(x)}{f_{\phi_m}^0(x)} \right] \right) = 0. \quad (5.6d)$$

From (5.6a), the worst-case distribution function is given by

$$f_{\phi_m}^*(x) = f_{\phi_m}^0(x) \exp \left(\frac{\mathbf{1}(x \geq \bar{\phi}_m) - \lambda_m}{\tau_m} - 1 \right), \quad (5.7)$$

where (τ_m, λ_m) satisfies conditions (5.6b)-(5.6d). Specifically, we have the following proposition.

Proposition 5.1 *The choice of (τ_m, λ_m) is a solution to the following equations*

$$H_1(\tau_m, \lambda_m) \triangleq R(\bar{\phi}_m) e^{-\frac{\lambda_m}{\tau_m}} + S(\bar{\phi}_m) e^{\frac{1-\lambda_m}{\tau_m}} - 1 = 0,$$

$$H_2(\tau_m, \lambda_m) \triangleq S(\bar{\phi}_m) e^{\frac{1-\lambda_m}{\tau_m}} - \lambda_m - \tau_m(1 + D_{\phi_m}) = 0,$$

where $S(\bar{\phi}_m) = (1 - F_{\phi_m}^0(\bar{\phi}_m)) e^{-1}$, $R(\bar{\phi}_m) = F_{\phi_m}^0(\bar{\phi}_m) e^{-1}$, and $F_{\phi_m}^0(\bar{\phi}_m) = \int_{x < \bar{\phi}_m} f_{\phi_m}^0(x) dx$.

The proof of Proposition 5.1 is straightforward and omitted here for conciseness. A direct solution to these nonlinear equations is difficult to obtain. Thus, we resort to an iterative method. Firstly, we take linear approximations for $H_1(\tau_m, \lambda_m)$ and $H_2(\tau_m, \lambda_m)$ as follows:

$$\begin{bmatrix} H_1(\tau'_m, \lambda'_m) \\ H_2(\tau'_m, \lambda'_m) \end{bmatrix} \approx \begin{bmatrix} H_1(\tau_m, \lambda_m) \\ H_2(\tau_m, \lambda_m) \end{bmatrix} + \mathbf{J}(\tau_m, \lambda_m) \begin{bmatrix} \Delta\tau_m \\ \Delta\lambda_m \end{bmatrix},$$

where $\tau'_m = \tau_m + \Delta\tau_m$, $\lambda'_m = \lambda_m + \Delta\lambda_m$, and $\mathbf{J}(\tau_m, \lambda_m)$ denotes the Jacobian matrix evaluated at (τ_m, λ_m) . Then we find an update for (τ_m, λ_m) if it is not feasible. The update $[\Delta\tau_m, \Delta\lambda_m]^T$ is a solution to the linear equation $\mathbf{J}(\tau_m, \lambda_m) \begin{bmatrix} \Delta\tau_m \\ \Delta\lambda_m \end{bmatrix} = - \begin{bmatrix} H_1(\tau_m, \lambda_m) \\ H_2(\tau_m, \lambda_m) \end{bmatrix}$. Note that τ_m should be non-negative during iterations.

5.2.3 Controlling Robust Transmit Power

Once we determine (τ_m, λ_m) for each PU receiver m , we can check the feasibility of p_s by comparing $\sum_{m \in \mathcal{M}} \eta_m^w(p_s)$ and the target η , where $\eta_m^w(p_s) = \mathbb{E}_{f_{\phi_m}^*} [\mathbf{1}(x \geq \bar{\phi}_m)] = (1 + D_{\phi_m})\tau_m + \lambda_m$. The second equality is obtained from (5.6a) and (5.6d). Next, we will find the optimal transmit power p_s^* such that $\sum_{m \in \mathcal{M}} \eta_m^w(p_s^*) = \eta$. Note that $\eta_m^w(p_s)$ is an increasing function with respect to p_s since we have

$$\partial \eta_m^w(p_s) / \partial p_s = \frac{\partial}{\partial p_s} \int_{\bar{\phi}_m}^{\infty} f_{\phi_m}^*(x) dx = \frac{\partial}{\partial p_s} \int_{\bar{\phi}_m/p_s}^{\infty} f_{z_m}^*(x) dx = \frac{\bar{\phi}_m}{p_s^2} f_{z_m}^*\left(\frac{\bar{\phi}_m}{p_s}\right) > 0,$$

where $z_m = \sum_{k \in \mathcal{K}} g_{km} = \phi_m/p_s$ is the aggregated channel gain and thus follows a distribution $f_{z_m}^*(x) = f_{\phi_m}^*(xp_s)p_s$ in the worst-case. The monotonicity of $\eta_m^w(p_s)$ allows us to search for p_s^* using a bisection method. With this transmit power p_s^* , objective (5.2a) gives a lower bound of the stochastic QoS provisioning for SUs, and can be processed in the same way as that for problem (5.5a)-(5.5c) when user priorities $\{\omega_n\}_{n \in \mathcal{N}}$ are fixed.

5.3 Case II: SAPs Transmit with Distinct Power

In the second case, each SAP adjusts its own transmit power so that its own utility (to be discussed) can be maximized. Considering SAPs' spatial distribution and different

influences to the PUs, it is likely that they have distinct transmit power levels at optimum. Note that the uncertainty in SUs' SINR γ_n can be processed in the same way as for AIP ϕ_m . Therefore, we only consider the uncertainty set $\mathbb{Z}_{\phi_m}$ in the following. Let $E_{n_k}(\bar{\gamma}_{n_k}, p_k) = \mathbb{E}_{f_{\gamma_{n_k}}}[\mathbf{1}(x \geq \bar{\gamma}_{n_k})]$ denote the service rate of SU n_k , where $f_{\gamma_{n_k}}(\cdot)$ is a known log-normal distribution. Then, each SAP's utility is the weighted sum of service rates of all associated SUs and is given by $u_k(\mathbf{p}_{-k}, p_k) = \sum_{n_k \in \mathcal{N}_k} \omega_{n_k} E_{n_k}(\bar{\gamma}_{n_k}, p_k)$, where $\mathbf{p}_{-k}$ denotes the power levels for all SAPs other than SAP k . Then our target is to determine the transmit power $\mathbf{p}$ for the following optimization problem:

$$\max_{\mathbf{p}} \quad \sum_{k \in \mathcal{K}} u_k(\mathbf{p}_{-k}, p_k) \quad (5.8a)$$

$$s.t. \quad I(\mathbf{p}_{-k}, p_k) \leq \eta, \quad (5.8b)$$

where $I(\mathbf{p}_{-k}, p_k) \triangleq \frac{1}{M} \sum_{m \in \mathcal{M}} \max_{f_{\phi_m} \in \mathbb{Z}_{\phi_m}} \mathbb{E}_{f_{\phi_m}}[\mathbf{1}(x \geq \bar{\phi}_m)]$ denotes PUs' average outage probability when SAPs transmit with power $\mathbf{p}$.

5.3.1 Primal-Dual Decomposition

Proposition 5.2 *SAP's utility $u_k(\mathbf{p}_{-k}, p_k)$ and PUs' interference $I(\mathbf{p}_{-k}, p_k)$ are both concave functions of p_k .*

The proof of Proposition 5.2 is given in Appendix 5.6. The concavity of $I(\mathbf{p}_{-k}, p_k)$ with respect to p_k makes problem (5.8a)-(5.8b) difficult to solve directly. As an attempt to transform (5.8a)-(5.8b) into an unconstrained optimization problem, we have the Lagrangian function as $\Lambda(\mathbf{p}, \nu) = \sum_{k \in \mathcal{K}} u_k(\mathbf{p}_{-k}, p_k) - \nu I(\mathbf{p}_{-k}, p_k) + \nu \eta$, where ν is the price of interference at PU receivers. Given the price ν , each SAP needs to adjust its transmit power to maximize $\Lambda(\mathbf{p}, \nu)$. However, we still face difficulties to solve the problem. Note that both $u_k(\mathbf{p}_{-k}, p_k)$ and $I(\mathbf{p}_{-k}, p_k)$ are concave functions, and there is no direct way to decouple the interference function $I(\mathbf{p})$. To overcome this difficulty, letting $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_K] \succeq 0$, we attribute the total interference $I(\mathbf{p})$ to each SAP k by a constant share θ_k , and reformulate the Lagrangian function as follows:

$$\Lambda(\mathbf{p}, \nu, \boldsymbol{\theta}) = \sum_{k \in \mathcal{K}} [u_k(\mathbf{p}_{-k}, p_k) - \nu \theta_k I(\mathbf{p}_{-k}, p_k)] + \nu \eta. \quad (5.9)$$

Obviously, we require $\sum_{k \in \mathcal{K}} \theta_k = 1$ and each share θ_k can be viewed as a portion of the total debt to PUs. For different fairness consideration, we can set θ_k accordingly, e.g., we set $\theta_k = 1/K$ to enable equal rights for all SAPs.

To maximize the Lagrangian function (5.9) in a distributed way, each SAP k optimizes its own transmit power p_k^* according to the overall interference $I(\mathbf{p}_{-k}, p_k)$ and its own debt θ_k to PUs, given the knowledge of other SAPs' transmit power $\mathbf{p}_{-k}$, i.e.,

$$p_k^*(\nu, \theta_k) = \arg \max_{p_k} u_k(\mathbf{p}_{-k}, p_k) - \nu \theta_k I(\mathbf{p}_{-k}, p_k). \quad (5.10)$$

The maximization requires calculations of the first derivatives of $u_k(\mathbf{p}_{-k}, p_k)$ and $I(\mathbf{p}_{-k}, p_k)$ with respect to p_k . Given the transmit power $\mathbf{p}$, we can easily obtain these two derivatives from (5.12) and (5.14), respectively. However, we are unable to solve the transmit power p_k^* directly from the first-order optimality condition, i.e., $\partial u_k(\mathbf{p}_{-k}, p_k)/\partial p_k = \nu \theta_k \partial I(\mathbf{p}_{-k}, p_k)/\partial p_k$, since both sides of this equation are implicit functions of p_k . Therefore, it is very difficult to establish a closed-form relation between $\partial I(\mathbf{p}_{-k}, p_k)/\partial p_k$ and p_k . Even though we find such p_k^* from the first-order optimality condition, it is not guaranteed to be the solution as the objective in (5.10) is a summation of a concave function $u_k(\mathbf{p}_{-k}, p_k)$ and a convex function $-\nu \theta_k I(\mathbf{p}_{-k}, p_k)$. The resulting p_k^* can be either a local maximum or a local minimum.

5.3.2 Distributed Power Control Algorithm

To this end, we resort to the concave-convex procedure (CCCP) [106] to construct an iterative algorithm that is guaranteed to maximize the objective function monotonically.

Proposition 5.3 *Maximizing $u_k(\mathbf{p}_{-k}, p_k) - \nu \theta_k I(\mathbf{p}_{-k}, p_k)$ can be performed iteratively by updating power p_k^{t+1} , given current power p_k^t , in a way such that*

$$\nabla u_k(\mathbf{p}_{-k}, p_k^{t+1}) = \nu \theta_k \nabla I(\mathbf{p}_{-k}, p_k^t). \quad (5.11)$$

Proof: Given any two feasible p_k^t and p_k^{t+1} , the concavity of $u_k(\mathbf{p}_{-k}, p_k)$ and $I(\mathbf{p}_{-k}, p_k)$ implies the following two inequalities, i.e., $u_k(\mathbf{p}_{-k}, p_k^{t+1}) \geq u_k(\mathbf{p}_{-k}, p_k^t) -$

$\nabla u_k(\mathbf{p}_{-k}, p_k^{t+1})(p_k^t - p_k^{t+1})$ and $-I(\mathbf{p}_{-k}, p_k^{t+1}) \geq -I(\mathbf{p}_{-k}, p_k^t) - \nabla I(\mathbf{p}_{-k}, p_k^t)(p_k^{t+1} - p_k^t)$.

It is straightforward to show that, if we have $\nabla u_k(\mathbf{p}_{-k}, p_k^{t+1}) = \nu \theta_k \nabla I(\mathbf{p}_{-k}, p_k^t)$, the objective function is improved by summing up these two inequalities, i.e., $u_k(\mathbf{p}_{-k}, p_k^{t+1}) - \nu \theta_k I(\mathbf{p}_{-k}, p_k^{t+1}) \geq u_k(\mathbf{p}_{-k}, p_k^t) - \nu \theta_k I(\mathbf{p}_{-k}, p_k^t)$. ■

Note that both $\nabla u_k(\mathbf{p}_{-k}, p_k)$ and $\nabla I(\mathbf{p}_{-k}, p_k)$ are monotonic functions from (5.12) and (5.14), respectively. Given p_k^t and $\nabla I(\mathbf{p}_{-k}, p_k^t)$, we can use a bisection method to search for p_k^{t+1} satisfying (5.11). When each SAP tunes to its preferred transmit power, the interference price ν should be updated accordingly. The detailed power control algorithm is given in Algorithm 4. The distribution uncertainties in line 2 of the algorithm are estimated through channel measurements, and we can update the reference distribution if new measurements reveal that the actual distribution deviates too much from its reference. In line 3 of the algorithm, we can initialize each SAP by its minimum transmit power and choose appropriate interference price ν such that the solution to (5.10) is nontrivial. Actually, any initial transmit power will not affect the algorithms convergence to the same point, however, the choice of minimum transmit power will suppress the interference to PUs even during the algorithm iterations. In lines 6 – 20 of the algorithm, each SAP maximizes its net utility in (5.10) by the concave-convex procedure. In each iteration, we update SAP's transmit power by a bisection method in lines 9 – 18. Then, we increase (or decrease) the interference price ν in line 21 if the resulting power vector $\mathbf{p}$ renders higher (or lower) interference than the prescribed level η . The choice of step size β and convergence property of this iterative process can be found in [92].

One point worth mentioning is that the calculation of $\nabla I(\mathbf{p}_{-k}, p_k^t)$ in line 7 of Algorithm 1 requires the evaluation of $I(\mathbf{p}_{-k}, x)$ at two adjacent points, e.g., $x_1 = p_k^t$ and $x_2 = p_k^t + \Delta$. Then we approximate $\nabla I(\mathbf{p}_{-k}, p_k^t)$ by $(I(\mathbf{p}_{-k}, x_2) - I(\mathbf{p}_{-k}, x_1))/\Delta$ for very small Δ compared with p_k^t . By the definition of $I(\mathbf{p}_{-k}, p_k^t)$ in (5.8b), we need to solve another instance of problem (5.5a)-(5.5c) by the method developed for Case I. Therefore, from this viewpoint, Case I is the core building block for Case II, not just a routine of simplification.

Algorithm 4 Robust Iterative Power Control

Input: Interference and QoS requirements ϕ_m , η and $\bar{\gamma}_n$
Output: Optimal transmit power $\mathbf{p}$ for SAPs

```

1: initialization
2:   estimate distribution uncertainties  $\mathbb{Z}_{\gamma_n}$  and  $\mathbb{Z}_{\phi_m}$ 
3:   set initial transmit power  $\mathbf{p}$  and interference price  $\nu$ 
4: end initialization
5: For each SAP  $k \in \mathcal{K}$ 
6:   while  $|p_k^{t+1} - p_k^t| > \epsilon$ 
7:     set  $p_k^t = p_k^{t+1}$  and calculate  $\nabla I(\mathbf{p}_{-k}, p_k^t)$ 
8:     set  $p_k^U = \bar{p}_k$  and  $p_k^L = 0$ 
9:     while  $|p_k^U - p_k^L| > \epsilon$ 
10:      if  $\nabla u_k(\mathbf{p}_{-k}, \frac{p_k^U + p_k^L}{2}) > \nu \theta_k \nabla I(\mathbf{p}_{-k}, p_k^t)$ 
11:        set  $p_k^L = \frac{p_k^U + p_k^L}{2}$ 
12:      else
13:         $p_k^U = \frac{p_k^U + p_k^L}{2}$ 
14:      end if
15:      if  $|\nabla u_k(\mathbf{p}_{-k}, \frac{p_k^U + p_k^L}{2}) - \nu \theta_k \nabla I(\mathbf{p}_{-k}, p_k^t)| < \epsilon$ ,
16:        break
17:      end if
18:    end while
19:    set  $p_k^{t+1} = \frac{p_k^U + p_k^L}{2}$ 
20:  end while
21: update price  $\nu = [\nu - \beta(\eta - I(\mathbf{p}))]^+$  and go to line 5
    
```

5.3.3 Algorithm Implementation

For implementation of Algorithm 4, we organize it by different functional modules as illustrated in Figure 5.3. The information processing module is in charge of the information gathering and synthesis, providing a knowledge profile for the upper operation module to make informed decisions. This module can be further divided into three units. Each SAP's sensing results and operation parameters (e.g., transmit power) are disseminated to other SAPs by the information exchange unit. The uncertainty extraction unit is specially designed for applications in cognitive radio networks. It helps to characterize the range of channel fluctuations by maintaining a reference dis-

Figure 5.3: System structure for robust power control.

tribution and a distance limit for every channel gain. When new channel information arrives, the unit calculates the KL divergence with respect to the reference distribution and overwrites the stored distance limit by new KL divergence if it becomes larger. Information sharing among SAPs is usually redundant and requires filtering and restructuring for further processing. In the power control problem, an SAP needs to estimate the interference at PU receivers, which requires the knowledge of SAPs' transmit power and the channel gains from SAPs to specific PU receiver. Therefore, the interference estimation unit will put related information together and provide a worst-case interference estimation based on the uncertainty sets extracted in the preceding unit. It also calculates the change of interference with respect to SAP's transmit power as required in line 7 of Algorithm 4. The information processing module interacts with the robust operation module at the interference estimation unit, which triggers the price update and therefore the power control algorithm. For example, PUs' mobility causes changes of the physical channel condition (as illustrated in Figure 5.2) and can be detected by SAPs through overhearing feedback packets from that PU receiver. Then, each SAP updates the uncertainty models and re-evaluates the potential interference to PUs. If current interference estimation is considered low, the robust operation module will suggest to use higher transmit power by tuning down the price of interference.

A distributed implementation of Algorithm 4 would be of paramount significance for practical system. Generally, the distributed design needs to address two problems. The first problem is to decouple secondary network from primary network, limiting

information exchange between SAPs and PBS. While the other one is to design a mechanism that enables the information sharing among SAPs. In our model, both power control methods in Case I and Case II do not require any information exchange with the primary network as each SAP can estimate the channel gains between SAP and PU receivers. However, we need some sort of communications among SAPs to estimate the outage probability at PU receivers as given in (5.4). Assume there is a common control channel among SAPs, then each SAP can collect the channel estimates exchanged from other SAPs and broadcast its own channel estimates as well. After information exchange, each SAP can individually estimate the worst-case outage probability at PU receivers, therefore support a distributed implementation of the power control algorithms. Note that the interference price in Algorithm 4 is a penalty due to interference violation, but it is not necessarily imposed by PBS. Instead, it can be updated and distributed within SAPs through the common control channel.

5.4 Performance Evaluation

To evaluate the proposed robust power control algorithm, a cognitive radio network with 3 SAPs is simulated under the coverage area of a PBS. The PBS transmits with fixed power at 300mW, and each SAP has a maximum transmit power of 200mW. There are 50 PUs uniformly distributed under PBS's coverage area with radius 800m, and each SAP has 10 SUs distributed under its coverage area with radius 200m. The placements of SAPs and SUs are at least 30m away from PBS and PUs. The mean path loss is given as $|h|^2 = \frac{1}{cd^n}$ where d is the distance between transceivers, c and n denote the propagation constant and propagation exponent, respectively. Assuming that the primary and the secondary networks experience a similar channel condition, we set the same $c = 0.5$ and $n = 3.5$ for PBS-PU and SAP-SU channels [107], and set a larger $c = 3.5$ for PBS-SU and SAP-PU interference channels. Due to user mobility, the channel gains are subject to log-normal distributed random variables with zero mean and standard derivation 5dB [108]. Each PU receiver has the same interference power threshold $\bar{\phi} = -65\text{dBw}$, and the outage probability threshold $\eta = 10\%$. All

SUs have the same minimum SINR requirement of 12dB [109], and the noise floor is -130dBw . In the following, we make a comparative study on the performance of our robust methods proposed in Section 5.2 and 5.3, in terms of PU protection levels and QoS provisioning for SUs, respectively. All the simulations are performed in a DELL precision T3500 workstation with Matlab version 7.11.0.584 and Windows 7 Professional SP1.

5.4.1 Channel Uncertainty

Before the comparative study, we embody the existence of distribution uncertainty by demonstrating the changes of AIP's and SINR's distribution functions in different channel measurements. As an illustration, we place the PBS at position $(400m, 400m)$, and three SAPs at $(300m, 300m)$, $(300m, 500m)$ and $(750m, 400m)$, respectively. Note that SAPs 1 and 2 are close to each other and to the PBS, while SAP 3 is far away from the PBS. Initially, all SAPs transmit with the same power at 10mW. We simulate the channel gain by log-normal distributed random variable on top of the mean path loss, then measure AIPs ϕ_m at PU receivers and SINRs γ_n at SU receivers, respectively. With a limited sample size (e.g., 5000 in the simulations), channel measurements reveal that data samples in each sensing period may follow a distribution other than a log-normal reference distribution. To quantify their difference, we fit the data samples into a closed-form distribution and calculate its KL divergence with respect to the log-normal reference distribution. The results are presented in Figure 5.4a. We notice that, compared with SINR, AIP has larger KL divergence with respect to the log-normal reference distribution. This can be shown in our experiments that AIP (in dB) distribution is generally *asymmetric* and has larger skewness than that of SINR (in dB), hence deviates more from its log-normal reference distribution $f_{\phi_m}^0(x)$ which is *symmetric* with respect to $10 \log_{10}(x)$.

To overcome the distribution uncertainty, we require a distance limit that bounds the KL divergence in most of the cases. Figure 5.4b plots an empirical relation between distance limit and the sample size based on the simulation settings. The distance limits in Figure 5.4b bound the KL divergence with a probability 90%. Apparently,

(a) Distribution uncertainty

(b) Distance limit vs. sample size

Figure 5.4: Empirical setting for distance limit.

we need to set a larger distance limit for the AIP distribution than that for the SINR distribution, if we use log-normal reference distributions. In the simulation, we set $D_{\phi_m} = 0.02$ and $D_{\gamma_n} = 0.01$, respectively.

To show the concavity of $u_k(\mathbf{p}_{-k}, p_k)$ and $I(\mathbf{p}_{-k}, p_k)$ as proved in Proposition 5.2 with the simulation setting, we plot the changes of $u_k(\mathbf{p}_{-k}, p_k)$ and $I(\mathbf{p}_{-k}, p_k)$ with respect to p_1 in Figure 5.5. When p_1 increases, the average SINR of SUs associated with SAP 1 is correspondingly increased, so does interference at other SU receivers. Therefore, the utility is an increasing function for SAP 1, while a decreasing function for SAPs 2 and 3. Further, we observe that $u_k(\mathbf{p}_{-k}, p_k)$ is concave with respect to p_k and convex with respect to other SAPs' transmit power. The concavity of interference function $I(\mathbf{p}_{-k}, p_k)$ is illustrated in Figure 5.5b where the nominal PU protection is obtained by taking expectation over the reference distribution when calculating the interference $I(\mathbf{p}_k, p_k)$. However, with channel uncertainty, the actual outage probability may be much worse than the nominal PU protection level. The robust PU protection level in Figure 5.5b gives the worst-case outage probability with a distance limit $D_{\phi_m} = 0.02$.

5.4.2 PU Protection

Given the log-normal reference distribution and distance limit, we can obtain the robust transmit power by the proposed algorithms in Sections 5.2 and 5.3. For a clear presentation, we denote these two methods by UTP (universal transmit power) and DTP (distinct transmit power), respectively. To check their protection for PUs, we compare them with a non-robust, i.e., nominal power control method (denoted as NPC) that does not incorporate distribution uncertainty, e.g., [79]. In the NPC method, we simply use the reference distribution as an approximation of the AIP distribution, based on which SAPs estimate the outage probability at PU receivers and adjust their transmit power levels. In our robust methods (UTP or DTP), the AIP may change its distribution in each sensing period and we consider worst-case interference in power control. After we obtain the optimal transmit power from these

 (a) SAP's utility $u_k(\mathbf{p}_{-k}, p_k)$.

 (b) PUs' interference $I(\mathbf{p}_{-k}, p_k)$.

Figure 5.5: Concavity of $u_k(\mathbf{p}_{-k}, p_k)$ and $I(\mathbf{p}_{-k}, p_k)$ w.r.t. p_k .

(a) Outage probability at PU receivers

(b) KL divergence in experiments

Figure 5.6: PU protection with robust and nominal power control.

three methods, we compare the actual interference they introduce to the primary network, respectively.

We have 30 runs in the simulation, and each run contains 500 AIP samples according to the channel fading model, based on which we estimate the outage probabilities at PU receivers. The average outage probability over all PU receivers is given in Figure 5.6a. We observe that the robust methods guarantee the outage probability to be less than 10% in all the runs, while the NPC method only provides a desired PU protection level when AIP samples follow a distribution very close to the reference distribution (i.e., small KL divergence in Figure 5.6b). However, a violation happens when using the NPC method as the channel fluctuates more intensively and the AIP distribution deviates largely from the reference distribution (i.e. large KL divergence in Figure 5.6b). In this case, our robust methods show their significance since the outage probability is still maintained at the prescribed level. Note that the KL divergence in each run is bounded by the empirical distance limit 0.02 as shown in Figure 5.6b.

5.4.3 SU Performance

With the fluctuations of channel gains, SAP's transmit power may lead to transmission failures due to the violations of PUs' interference constraint. It is likely that, in the case of interference violation, SAPs are forced to suspend their transmissions or levied extra payment for their unsupervised spectrum usage. In the simulation, we set SUs' service rates to zero when PUs' interference constraint is violated, and compare the average service rates with the robust and nominal power control methods, respectively. Each time with different outage probability threshold η , we obtain a nominal transmit power p^N and a robust transmit power p^U from the UTP method, respectively. Then, we run the simulation for 30 times with each of the transmit power, and record the average service rate as shown in Figure 5.7a. The results imply that the robust method with UTP may provide more transmit opportunities than that of the NPC method. Even though the NPC method allows SAPs to transmit with higher power and can provide a larger service rate in a single transmission, the

(a) Service rates

(b) Power levels

Figure 5.7: Robust power control vs. nominal power control.

high transmit power is vulnerable to punishment from PUs due to potential violations of PUs' interference constraint, and thus brings down SUs' average service rate. We also note that, at $\bar{\phi}_m = -55\text{dB}$, SUs' average service rate with the NPC method has a steep raise (Figure 5.7a) when the outage probability threshold increases from $\eta = 0.1$ to $\eta = 0.18$, though the transmit power keeps constant at its peak level (Figure 5.7b) during this period. That is because, a higher outage probability threshold reduces the probability of transmission failure or punishment from PUs due to the violations of PUs' interference constraint, allowing a very large space for the nominal method to accumulate SUs' average service rate.

Considering SAPs' spatial distribution and their different contributions to the total interference of primary network, we optimize distinct transmit power $\mathbf{p}^D = [p_1^D, p_2^D, p_3^D]$ for SAPs by the DTP method, and compare its QoS provisioning for SUs with that of the UTP method. With SAP 3 locating far away, i.e., $(750m, 400m)$ from the center of primary network and inducing less interference, the DTP method enables it to transmit with higher power p_3^D than that of the other SAPs as shown in Figure 5.8a, and hence provides a higher service rate than that of UTP method as shown in Figure 5.8b. We also compare these two methods when varying SAPs' locations. For simplicity, let SAPs 1 and 2 be stationary and SAP 3 move horizontally from $(500m, 400m)$ to its current location $(750m, 400m)$. Before reaching the location $(600m, 400m)$, SAP 3 has almost the same distance with that of SAPs 1 and 2. Therefore, they have relatively the same interference to PBS, and the DTP method is no better than the UTP method as shown in Figure 5.9a. As SAP 3 continues to move away from PBS, the DTP method can notably increase the transmit power of SAP 3, resulting in much higher average transmit power than that of the UTP method as shown in Figure 5.9b, and we can observe that the DTP method provides better service rate for SUs.

5.5 Summary

In this chapter, we study the robust power control problem with a new channel uncertainty model. It relies on a reference distribution which is a closed-form approxima-

(a) Power levels

(b) Service rates

Figure 5.8: The comparison between robust UTP and DTP methods.

(a) Service rates

(b) Average power

Figure 5.9: Performance comparison as SAP 3 moves away from PBS.

tion extracted from historical channel measurements, while allowing the actual AIP or SINR to follow a different distribution. By using a probabilistic distance measure, we provide a quantitative way to describe the uncertain distribution. Then, we formulate the power control problem into a chance constrained robust optimization, and propose two iterative algorithms to determine the robust transmit power for two cases, respectively. Simulation results demonstrate that both methods provide better PU protection than the existing work which does not take account of channel uncertainty.

5.6 Appendix: Proof for Proposition 5.2

As utility $u_k(\mathbf{p}_{-k}, p_k)$ is a weighted summation of $F_{n_k}(\bar{\gamma}_{n_k}, p_k)$ for each user $n_k \in \mathcal{N}_k$, our focus shifts to study the concavity of function $F_{n_k}(\bar{\gamma}_{n_k}, p_k)$. Let $g_{kn} = \frac{h_{kn}}{n_0 + \sum_{i \neq k} p_i h_{in}}$ and $f_g(x)$ be the distribution of g_{kn} . With $\gamma_{n_k} = p_k g_{kn}$, given the distribution $f_{\gamma_{n_k}}(x)$ of γ_{n_k} , we have $f_g(x) = p_k f_{\gamma_{n_k}}(p_k x)$ and $F_{n_k}(\bar{\gamma}_{n_k}, p_k) = \int_{\bar{\gamma}_{n_k}}^{\infty} f_{\gamma_{n_k}}(x) dx = \int_{\frac{\bar{\gamma}_{n_k}}{p_k}}^{\infty} f_g(x) dx$. Therefore,

$$\partial F_{n_k}(\bar{\gamma}_{n_k}, p_k) / \partial p_k = f_g\left(\frac{\bar{\gamma}_{n_k}}{p_k}\right) \frac{\bar{\gamma}_{n_k}}{p_k^2} \geq 0, \quad (5.12)$$

and its second-order derivative with respect to p_k is give by

$$\frac{\partial^2 F_{n_k}}{\partial p_k^2} = -\frac{\bar{\gamma}_{n_k}}{p_k^3} \left(2f_g\left(\frac{\bar{\gamma}_{n_k}}{p_k}\right) + \frac{\bar{\gamma}_{n_k}}{p_k} f'_g\left(\frac{\bar{\gamma}_{n_k}}{p_k}\right) \right) = -\frac{\bar{\gamma}_{n_k}}{p_k^2} \left(2f_{\gamma_{n_k}}(\bar{\gamma}_{n_k}) + \bar{\gamma}_{n_k} f'_{\gamma_{n_k}}(\bar{\gamma}_{n_k}) \right). \quad (5.13)$$

Assuming SINR γ_{n_k} follows a log-normal distribution, i.e., $10 \log_{10}(\gamma_{n_k})$ is a normal distribution with mean $\mu_{\gamma_{n_k}, dB}$ and variance $\sigma_{\gamma_{n_k}, dB}^2$, we can find $f_{\gamma_{n_k}}(x)$ by applying the change-of-variables rule on the density function of a normal distribution, i.e.,

$$f_{\gamma_{n_k}}(x) = \frac{\kappa}{x \sqrt{2\pi \sigma_{\gamma_{n_k}, dB}^2}} \exp \left(-\frac{(10 \log_{10}(x) - \mu_{\gamma_{n_k}, dB})^2}{2\sigma_{\gamma_{n_k}, dB}^2} \right),$$

where $\kappa = 10 \log_{10}(e)$ and its derivative is given by

$$f'_{\gamma_{n_k}}(x) = -\frac{1}{x} f_{\gamma_{n_k}}(x) \left(1 + \frac{10 \log_{10}(x) - \mu_{\gamma_{n_k}, dB}}{\sigma_{\gamma_{n_k}, dB}^2 / \kappa} \right).$$

To prove the concavity of $F_{n_k}(\bar{\gamma}_{n_k}, p_k)$, we require $2f_{\gamma_{n_k}}(\bar{\gamma}_{n_k}) + \bar{\gamma}_{n_k} f'_{\gamma_{n_k}}(\bar{\gamma}_{n_k}) \geq 0$ in (5.13). That is,

$$f_{\gamma_{n_k}}(\bar{\gamma}_{n_k}) \left(1 - \frac{10 \log_{10}(\bar{\gamma}_{n_k}) - \mu_{\gamma_{n_k}, dB}}{\sigma_{\gamma_{n_k}, dB}^2 / \kappa} \right) \geq 0.$$

It implies $10 \log_{10}(\bar{\gamma}_{n_k}) \leq \mu_{\gamma_{n_k}, dB} + \sigma_{\gamma_{n_k}, dB}^2 / \kappa$. This is generally true in a practical system since we usually require the average received SINR $\mu_{\gamma_{n_k}, dB}$ to be greater than a QoS threshold $\bar{\gamma}_{n_k}$ (in dB). Moreover, if the average received SINR is far below the QoS threshold, we can have admission control such that the above condition is always satisfied. In Section 5.4 via simulation, we find that $\mu_{\gamma_{n_k}, dB}$ is in the range of $[-30, 10]$ dB depending on SUs' spatial locations while $\sigma_{\gamma_{n_k}, dB}$ falls around 25dB. A typical SINR requirement $\bar{\gamma}_{n_k}$ (in dB) for WLAN can be set to 12dB, which is far less than $\mu_{\gamma_{n_k}, dB} + \sigma_{\gamma_{n_k}, dB}^2 / \kappa$.

To study the second-order derivative of PUs' interference with respect to p_k , we define $I_m(\mathbf{p}) = \mathbb{E}_{f_{\phi_m}^w}[\mathbf{1}(x \geq \bar{\phi}_m)]$ to be the interference at PU m . If $I_m(\mathbf{p})$ is shown to be concave with respect to p_k , so will be $I(\mathbf{p})$. Note that $\phi_m = z + p_k h_{km}$ where $z = \sum_{i \neq k} p_i h_{im}$. Letting $f_z^w(\cdot)$ and $f_h^w(\cdot)$ denote the worst-case density function of z and h_{km} , respectively, we have $I_m(\mathbf{p}_{-k}, p_k) = 1 - \int_0^{\bar{\phi}_m} f_z^w(y) F_h^w\left(\frac{\bar{\phi}_m - y}{p_k}\right) dy$, where $F_h^w(\cdot)$ denotes the cumulative distribution of $f_h^w(\cdot)$. Consequently,

$$\frac{\partial I_m}{\partial p_k} = \frac{1}{p_k^2} \int_0^{\bar{\phi}_m} f_z^w(y) f_h^w\left(\frac{\bar{\phi}_m - y}{p_k}\right) dy \quad (5.14)$$

and its second-order derivative is given by

$$\frac{\partial^2 I_m}{\partial p_k^2} = -\frac{1}{p_k^2} \int_0^{\bar{\phi}_m} f_z^w(\bar{\phi}_m - t) \left(2f_y^w(t) + t(f_y^w)'(t)\right) dt, \quad (5.15)$$

where $f_y^w(\cdot)$ denotes the density distribution of $y = p_k h_{km}$. Through a similar reasoning for (5.13), we have $\frac{\partial^2 I_m}{\partial p_k^2} \leq 0$ if $\bar{\phi}_m(dB) \leq \mu_{y, dB} + \sigma_{y, dB}^2 / \kappa$. In the same simulation setting, the critical point $\mu_{y, dB} + \sigma_{y, dB}^2 / \kappa$ generally falls in the range $[-30, 10]$ dB, which is also much larger than $\bar{\phi}_m$ (in dB).

Chapter 6

Conclusions and Future Works

6.1 Conclusions

In this Ph.D. research, we identify a common challenge, i.e., channel uncertainty, in the practical implementation of two important aspects of cognitive radio networks, i.e., spectrum identification and exploitation. We note that, an optimal design parameter (e.g., transmit power level) is always based on some assumption and will achieve maximum performance if the observed system parameter (e.g., estimated interference at PU side) matches with the assumption. When mismatch happens, we normally ignore it and accept a suboptimal performance if the mismatch is still within a tolerable bound.

As our main contribution, we propose a framework to characterize the uncertainty (e.g., parameter mismatch) and take proactive steps in the phase of system design to avoid severe performance degradation in the phase of implementation. Specifically, we apply robust optimization techniques to re-design spectrum sensing and power control algorithms that are robust against the channel uncertainty. Compared with existing stochastic and the worst-case methods, the introduction of distribution uncertainty combines these two methods, providing a trade-off study between conservativeness and practicability.

For ease of analytical study, we characterize the distribution uncertainty by two convex sets, i.e., moment-based uncertainty and reference-based uncertainty. In the moment-based uncertainty, the distribution uncertainty is defined by a series of inequality constraints on the moment statistics. In the reference-based uncertainty, we

extract a reference distribution from the historical measurement data and allow actual signal distribution to shift around from the reference distribution. Their difference is quantified by a probabilistic distance measure. Our experiments show that the reference-based uncertainty model is more flexible in characterizing signal's fluctuations than the moment-based uncertainty model.

As an application in cognitive radio networks, we first study the performance of spectrum sensing under different distribution uncertainty models. Though we have to design solution methods for these models case by case, there exist a common structure that allows us to transform a semi-infinite program into semi-definite program by the Lagrangian method, e.g., the lower detection bound is equivalent to solve a semi-definite program for single-user with moment-based uncertainty. For multi-user case, we approximate the problem into a series of semi-definite programs, which can be solved easily by conic optimization tools. We also apply the reference-based uncertainty model to study the robust power control problem in two difference cases. Our proposed iterative method not only guarantees the protection for PUs under channel fluctuations, but can also improve the throughput performance of SUs as illustrated through numerical evaluations.

6.2 Future Works

In Chapter 2, we present a survey on the optimal design of various sensing parameters. Theoretically, all of these sensing parameters may be subject to some uncertain factors. Therefore, we can extend our work to study any of these parameters if we can identify the sources of uncertainty and characterize them properly. Specifically, we can have several potential extensions in the following directions.

6.2.1 Cooperative Sensing with Reference-based Uncertainty

In Chapter 3, we propose several uncertainty sets to describe the distribution uncertainty. These sets fall into two categories, i.e., moment-based uncertainty and reference-based uncertainty. In chapter 4, we study the robust design of decision thresholds for multi-user cooperative sensing with the moment-based uncertainty

model. In our following study, we may consider to make a better use of the sensing history and extract a closed-form distribution that serves as a reference distribution. This closed-form reference distribution may draw from elaborate channel modeling or online goodness-of-fit test for the historical signal samples. Then, we can re-visit the robust design of decision thresholds with reference-based uncertainty model. We expect it to have a better characterization of the real-world uncertainty and enhance the robust sensing performance.

6.2.2 Joint Design of Sensing Parameters and Power Control

In our study, we notice that channel uncertainty is the common challenge for both spectrum identification and exploitation, and present a separate study in these two aspects. But in fact, these two aspects could be coupled as we explained in Chapter 2. For example, the sensing information can help to determine the optimal power allocation in different spectrum bands or time slots. SUs' power budget also poses constraints on the power allocation in sensing and transmission periods, therefore helps to optimize the sensing parameters. Though lots of works in literature study the joint optimization of sensing parameters and the power control mechanism, very few works take the channel uncertainty into account. As one of the future directions, we may investigate the joint and robust decision on sensing time and power control when channel is subject to uncertainty.

6.2.3 Asymmetry in Distribution Uncertainty

Our research largely relies on the development in robust optimization techniques. Following a typical procedure for robust optimization, our future directions may also include the discovery of new uncertainty models and tractable re-formulations that describe our design problem more practically.

In Chapter 3, we propose two different categories of uncertainty sets, i.e., the moment-based and the reference-based. The moment-based uncertainty regulates the shape of a distribution function by its mean and variance. In practice, we may have some extra knowledge about the signal distribution that relates to a higher moment

value. For example, if the channel gain follows a log-normal distribution (symmetric in log-scale), then the aggregate interference power at PU receiver will follow an asymmetric distribution as it is an additive effect of multiple log-normal variables. To describe such property in our future work, we can resort to an asymmetry measure, e.g., the skewness based on the third-order moment value. Usually, it is easy to add new constraints to the uncertainty sets, but it may significantly complicate the algorithm design or even result in intractable robust formulation. Therefore, we need pay special attention to the choice of asymmetry measure, which not only captures the distribution asymmetry properly, but also enables a closed-form study on the robust performance.

6.2.4 Heterogeneous Uncertainty Set

The current version of robust optimization problem contains at least two variables, i.e., decision parameter (e.g., power control level) and uncertain parameter (e.g., AIP distribution). For any decision parameter in its feasible set, we can anticipate a worst-case performance when the uncertain parameter takes any instance in the uncertainty set. A solution to the robust optimization problem is to find such a decision parameter that gives the optimal worst-case performance. In this framework, the uncertainty set does not change with different choices of decision parameter, which we call homogeneous uncertainty. It implies that the uncertain parameter is irrelevant with the decision parameter.

Another variant of this framework is to define heterogeneous uncertainty for decision variables, i.e., specify different uncertainty sets when decision parameter takes different values. The physical meaning of this new variant can be well explained by the fact that we usually employ an iterative procedure to find the robust solution. Before we start an iterative algorithm, we have an initial uncertainty set that bears the most uncertain information. Once we make a tentative choice, we may observe some new information that may help to update the uncertainty set. Most importantly, different choices of decision variable may incur different responses from the system, thus we can customize the uncertainty set for each decision variable. We expect this

new framework to improve the robust performance as it gains new information in the process of decision making.

6.2.5 Distributed Implementation

In our current study, we focus on an analytical study about the robust performance, which guides us to the robust design of the decision variables. As shown in Chapters 3 and 4, we succeed in providing an analytical study. The lower bound of detection probability can be found in a convex semi-definite program, which can be solved by an interior point method. In our future work, we may consider developing distributed algorithms for solving semi-definite programs, especially when the robust design requires information from different SUs. Take cooperative spectrum sensing as an example, current version of robust design requires the fusion center to solve iteratively a series of semi-definite programs. If we can find a distributed implementation, the fusion center can be removed and the network becomes more reliable and scalable.

Another interesting research direction is the combination of robust optimization and game theory, which is inherently strong at analyzing user interaction and has been intensively studied in wireless power control. The resulting combination, namely, robust game model, can be viewed as another solution for the distributed implementation of a robust design. Take power control as an example, SU's best response power update is based on other SUs' transmit power and its channel estimates, which may also be subject to uncertainty. Therefore, each SU's best response becomes a sub-level robust design problem, which is supposed to stabilize at an equilibrium point, i.e., robust Nash equilibrium.

List of Publications

Book Chapter

- [1] **Shimin Gong**, Ping Wang, and Jianwei Huang, “Robust performance of spectrum sensing in cognitive radio networks,” *Cognitive Radio and Interference Management: Technology and Strategy*. IGI Global Publisher, 2012.

Journal Papers

- [1] **Shimin Gong**, Ping Wang, Yongkang Liu, and Weihua Zhuang, “Robust power control in cognitive radio networks with channel uncertainty”, *IEEE J. Sel. Area Commun.*, vol.31, no.11, Nov. 2013.
- [2] **Shimin Gong**, Ping Wang, and Jianwei Huang, “Robust Performance of spectrum sensing in cognitive radio networks,” *IEEE Trans. Wireless Commun.*, vol.12, no.5, pp.2217–2227, May 2013.
- [3] **Shimin Gong**, Ping Wang, “Spectrum Sensing with Uncertainty in Cognitive Radio Networks,” , *IEEE Comsoc MMTC E-Letter*, 2012.

Conference Papers

- [1] **Shimin Gong**, Ping Wang, Lingjie Duan, and Rongxing Lu, “A game theoretic approach for robust power control in cognitive radio networks,” submitted to *IEEE INFOCOM*, Toronto Canada, Apr. 2014.
- [2] **Shimin Gong**, Ping Wang, Wei Liu, and Weihua Zhuang, “Performance bounds of energy detection with signal uncertainty in cognitive radio networks,” in *Proc. IEEE INFOCOM*, Turin Italy, Apr. 2013.
- [3] **Shimin Gong**, Ping Wang, Yongkang Liu, and Weihua Zhuang, “Robust power control in cognitive radio networks with channel uncertainty,” in *Proc. IEEE WCNC*, Shanghai China, Apr. 2013.

- [4] **Shimin Gong**, Ping Wang, and Jianwei Huang, “Robust threshold design for cooperative sensing in cognitive radio networks,” in Proc. IEEE INFOCOM-mini, Orlando FL USA, Mar. 2012.
- [5] **Shimin Gong**, Ping Wang, and Wei Liu, “Spectrum sensing under distribution uncertainty in cognitive radio networks,” in Proc. IEEE ICC, Ottawa Canada, Jun. 2012.
- [6] **Shimin Gong**, Ping Wang, and Jianwei Huang, “On-demand spectrum sharing by flexible time-slotted cognitive radio networks,” in Proc. IEEE GLOBECOM, California USA, Dec. 2012.
- [7] Qiumin Dong, Dusit Niyato, Ping Wang, and **Shimin Gong**, “Deferrable load scheduling of smart meter with intermittent communication connection,” in Proc. IEEE SmartGridComm, Tainan Taiwan, Nov. 2012.
- [8] **Shimin Gong**, Ping Wang, W. Liu, and W. Yuan, “Maximize secondary user throughput via optimal sensing in multi-channel cognitive radio networks,” in Proc. IEEE GLOBECOM, Dec. 2010.
- [9] **Shimin Gong**, Ping Wang, and Dusit Niyato, “Optimal power control in interference-limited cognitive radio networks,” in Proc. IEEE ICCS invited paper, Nov. 2010.
- [10] W. Liu, **Shimin Gong**, Y. Zhou, and P. Wang, “Two-phase indoor positioning technique in wireless networking environment,” in Proc. IEEE ICC, Cape Town, South Africa, 23-27 May 2010.
- [11] **Shimin Gong**, W. Liu, W. Yuan, W. Cheng, S. Wang, “Threshold-learning in local spectrum sensing of cognitive radio,” in Proc. IEEE VTC Spring, Barcelona Spain, Apr., 2009.
- [12] D. J. Kadhim, **Shimin Gong**, W. Xia, W. Liu, W. Cheng, “Power efficiency maximization in cognitive radio networks,” in Proc. IEEE WCNC, Budapest Hungary, Apr., 2009.

Bibliography

- [1] G. B. O. of Communications, *Spectrum Framework Review: A Consultation on Ofcom's Views as to how Radio Spectrum Should be Managed*. Ofcom, 2004.
- [2] F. Sanders, "Broadband spectrum surveys in Denver, CO, San Diego, CA, and Los Angeles, CA: methodology, analysis, and comparative results," in *Proc. IEEE International Symposium on Electromagnetic Compatibility*, vol. 2, pp. 988–993, Aug. 1998.
- [3] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, "Chicago spectrum occupancy measurements & analysis and a long-term studies proposal," in *Proc. international workshop on Technology and policy for accessing spectrum*, (New York, NY, USA), ACM, 2006.
- [4] D. Chen, S. Yin, Q. Zhang, M. Liu, and S. Li, "Mining spectrum usage data: a large-scale spectrum measurement study," in *Proc. ACM international conference on Mobile computing and networking (MobiCom)*, (New York, NY, USA), pp. 13–24, ACM, 2009.
- [5] M. Islam, C. Koh, S. Oh, X. Qing, Y. Lai, C. Wang, Y.-C. Liang, B. Toh, F. Chin, G. Tan, and W. Toh, "Spectrum survey in singapore: Occupancy measurements and analyses," in *Proc. IEEE International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom)*, pp. 1–7, May. 2008.
- [6] R. Chiang, G. Rowe, and K. Sowerby, "A quantitative analysis of spectral occupancy measurements for cognitive radio," in *Proc. IEEE Vehicular Technology Conference (VTC) Spring*, pp. 3016–3020, Apr. 2007.
- [7] V. N. Q. Bao, L. Q. Cuong, L. Q. Phu, T. D. Thuan, N. T. Quy, and L. M. Trung, "Vietnam spectrum occupancy measurements and analysis for cognitive radio applications," in *Proc. International Conference on Advanced Technologies for Communications (ATC)*, pp. 135–143, Aug. 2011.

- [8] M. Lopez-Benitez, A. Umberto, and F. Casadevall, "Evaluation of spectrum occupancy in Spain for cognitive radio applications," in *Proc. IEEE Vehicular Technology Conference (VTC) Spring*, pp. 1–5, Apr. 2009.
- [9] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, 2006.
- [10] I. Mitola, J., "Cognitive radio for flexible mobile multimedia communications," in *Proc. IEEE International Workshop on Mobile Multimedia Communications (MoMuC)*, pp. 3–10, 1999.
- [11] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Sel. Area. Commun.*, vol. 23, pp. 201–220, Feb. 2005.
- [12] Y.-C. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 1326–1337, Apr. 2008.
- [13] W.-Y. Lee and I. F. Akyildiz, "Optimal spectrum sensing framework for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 3845–3857, October 2008.
- [14] W. Yin, P. Ren, Q. Du, and Y. Wang, "Delay and throughput oriented continuous spectrum sensing schemes in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. PP, no. 99, 2012.
- [15] A. Ghasemi and E. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *Proc. IEEE DySPAN*, pp. 131–136, Nov. 2005.
- [16] F. F. Digham, M.-S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," *IEEE Trans. Commun.*, vol. 55, pp. 21–24, Jan. 2007.
- [17] A. Rao and M. Alouini, "Performance of cooperative spectrum sensing over non-identical fading environments," *IEEE Trans. Commun.*, vol. 59, pp. 3249–3253, Dec. 2011.

- [18] A. Ben-tal and A. Nemirovski, “Robust solutions of linear programming problems contaminated with uncertain data,” *Mathematical Programming*, vol. 88, pp. 411–424, 2000.
- [19] A. Ben-Tal and A. Nemirovski, “Robust convex optimization,” *Mathematics of Operations Research*, vol. 23, pp. 769–805, November 1998.
- [20] A. Ben-tal, L. El, and G. A. Nemirovski, “Robust semidefinite programming,” in *Handbook on Semidefinite Programming*, Kluwer Academic Publishers, pp. 139–162, 1998.
- [21] M. Sim, *Robust Optimization*. Massachusetts Institute of Technology, Sloan School of Management, Operations Research Center, 2004.
- [22] K. M. Teo, *Nonconvex robust optimization*. PhD thesis, 2007.
- [23] G. Zheng, K.-K. Wong, and B. Ottersten, “Robust cognitive beamforming with bounded channel uncertainties,” *IEEE Trans. Signal Process.*, vol. 57, pp. 4871–4881, Dec. 2009.
- [24] Y. Huang, Q. Li, W.-K. Ma, and S. Zhang, “Robust multicast beamforming for spectrum sharing-based cognitive radios,” *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 527–533, 2012.
- [25] A. Sahai, N. Hoven, and R. Tandra, “Some fundamental limits on cognitive radio,” in *Proc. Allerton Conference on Communication, Control, and Computing*, 2004.
- [26] D. Cabric, S. Mishra, and R. Brodersen, “Implementation issues in spectrum sensing for cognitive radios,” in *Proc. Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 772–776, Nov. 2004.
- [27] N. Han, S. Shon, J. H. Chung, and J. M. Kim, “Spectral correlation based signal detection method for spectrum sensing in IEEE 802.22 WRAN systems,” in *Proc. International Conference on Advanced Communication Technology*, vol. 3, Feb. 2006.
- [28] J.-K. Lee, J.-H. Yoon, and J.-U. Kim, “A new spectral correlation approach to spectrum sensing for 802.22 wran system,” in *Proc. IEEE International Conference on Intelligent Pervasive Computing*, pp. 101–104, 2007.

- [29] W. Zhang and Y. Sanada, "Cyclostationarity feature matched detection and application to IFDMA system," in *Proc. IEEE International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWN-COM)*, Jun. 2009.
- [30] J. Li, Z. Tan, S. Xu, and H. Wang, "Cooperative correlation based spectrum sensing for DMB-T systems," in *Proc. IEEE Vehicular Technology Conference (VTC) Spring*, May. 2010.
- [31] Y.-C. Liang, Y. Zeng, E. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 1326–1337, Apr. 2008.
- [32] W. yeol Lee, S. Member, and I. F. Akyildiz, "Akyildiz, optimal spectrum sensing framework for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3845–3857, 2008.
- [33] F. Moghimi, A. Nasri, and R. Schober, " l_p -norm spectrum sensing for cognitive radio networks impaired by non-gaussian noise," in *Proc. IEEE GLOBECOM*, Dec. 2009.
- [34] S. Gong, W. Liu, W. Yuan, W. Cheng, and S. Wang, "Threshold-learning in local spectrum sensing of cognitive radio," in *Proc. IEEE VTC Spring*, Apr. 2009.
- [35] D. Joshi, D. Popescu, and O. Dobre, "Gradient-based threshold adaptation for energy detector in cognitive radio systems," *IEEE Commun. Letters*, vol. 15, pp. 19–21, Jan. 2011.
- [36] X. Ling, B. Wu, H. Wen, P.-H. Ho, Z. Bao, and L. Pan, "Adaptive threshold control for energy detection based spectrum sensing in cognitive radios," *IEEE Wireless Commun. Letters*, vol. 1, pp. 448–451, Oct. 2012.
- [37] Y. Lin, K. Liu, and H. Hsieh, "On using interference-aware spectrum sensing for dynamic spectrum access in cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. PP, no. 99, p. 1, 2012.
- [38] E. Peh, Y.-C. Liang, Y. L. Guan, and Y. Zeng, "Optimization of cooperative sensing in cognitive radio networks: A sensing-throughput tradeoff view," *IEEE Trans. Veh. Technol.*, vol. 58, pp. 5294–5299, Nov. 2009.

- [39] Y. Pei, Y.-C. Liang, K. C. Teh, and K. H. Li, "Sensing-throughput tradeoff for cognitive radio networks: A multiple-channel scenario," in *Proc. IEEE PIMRC*, pp. 1257–1261, Sep. 2009.
- [40] K. Hamdi and K. Ben Letaief, "Power, sensing time, and throughput tradeoffs in cognitive radio systems: A cross-layer approach," in *Proc. IEEE WCNC*, Apr. 2009.
- [41] Y. Zou, Y.-D. Yao, and B. Zheng, "Outage probability analysis of cognitive transmissions: Impact of spectrum sensing overhead," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 2676–2688, Aug. 2010.
- [42] M. Cardenas-Juarez and M. Ghogho, "Spectrum sensing and throughput trade-off in cognitive radio under outage constraints over nakagami fading," *IEEE Commun. Lett.*, vol. 15, pp. 1110–1113, Oct. 2011.
- [43] S. Stotas and A. Nallanathan, "Overcoming the sensing-throughput tradeoff in cognitive radio networks," in *Proc. IEEE ICC*, May 2010.
- [44] S. Stotas and A. Nallanathan, "On the throughput and spectrum sensing enhancement of opportunistic spectrum access cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 97–107, Jan. 2012.
- [45] J. Ma, G. Zhao, and Y. Li, "Soft combination and detection for cooperative spectrum sensing in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 4502–4507, Nov. 2008.
- [46] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Physical Communication*, vol. 4, no. 1, pp. 40–62, 2011.
- [47] W. Zhang, R. Mallik, and K. Letaief, "Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 5761–5766, Dec. 2009.
- [48] K. Umebayashi, J. Lehtomaki, T. Yazawa, and Y. Suzuki, "Efficient decision fusion for cooperative spectrum sensing based on OR-rule," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 2585–2595, Jul. 2012.

- [49] B. Abbasi Bastami and E. Saberinia, “A practical multibit data combining strategy for cooperative spectrum sensing,” *IEEE Trans. Veh. Technol.*, vol. 62, pp. 384–389, Jan. 2013.
- [50] Z. Quan, S. Cui, and A. Sayed, “Optimal linear cooperation for spectrum sensing in cognitive radio networks,” *IEEE J. Sel. Top. Signal Process.*, vol. 2, pp. 28–40, Feb. 2008.
- [51] X. Zhou, J. Ma, G. Li, Y. Kwon, and A. Soong, “Probability-based combination for cooperative spectrum sensing,” *IEEE Trans. Commun.*, vol. 58, pp. 463–466, Feb. 2010.
- [52] G. Taricco, “Optimization of linear cooperative spectrum sensing for cognitive radio networks,” *IEEE J. Sel. Top. Signal Process.*, vol. 5, pp. 77–86, Feb. 2011.
- [53] S. Atapattu, C. Tellambura, and H. Jiang, “Energy detection based cooperative spectrum sensing in cognitive radio networks,” *IEEE Trans. Wireless Commun.*, vol. 10, pp. 1232–1241, Apr. 2011.
- [54] S. Althunibat, R. Palacios, and F. Granelli, “Performance optimisation of soft and hard spectrum sensing schemes in cognitive radio,” *IEEE Commun. Lett.*, vol. 16, pp. 998–1001, Jul. 2012.
- [55] S. Chaudhari, J. Lunden, V. Koivunen, and H. Poor, “Cooperative sensing with imperfect reporting channels: Hard decisions or soft decisions?,” *IEEE Trans. Signal Process.*, vol. 60, pp. 18–28, Jan. 2012.
- [56] B. Kasiri and J. Cai, “Effects of correlated shadowing on soft decision fusion in cooperative spectrum sensing,” Mar. 2010.
- [57] A. Cacciapuoti, I. Akyildiz, and L. Paura, “Correlation-aware user selection for cooperative spectrum sensing in cognitive radio ad hoc networks,” *IEEE J. Sel. Area. Commun.*, vol. 30, pp. 297–306, Feb. 2012.
- [58] L. Khalid and A. Anpalagan, “Cooperative sensing with correlated local decisions in cognitive radio networks,” *IEEE Trans. Veh. Technol.*, vol. 61, pp. 843–849, Feb. 2012.
- [59] W. Ren, Q. Zhao, and A. Swami, “Power control in cognitive radio networks: how to cross a multi-lane highway,” *IEEE J. Sel. Areas Commun.*, vol. 27, pp. 1283–1296, Sep. 2009.

- [60] X. Zhou, J. Ma, G. Li, Y. H. Kwon, and A. Soong, "Probability-based optimization of inter-sensing duration and power control in cognitive radio," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 4922–4927, Oct. 2009.
- [61] S. Almalfouh and G. Stuber, "Joint spectrum-sensing design and power control in cognitive radio networks: A stochastic approach," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 4372–4380, Dec. 2012.
- [62] A. T. Hoang and Y.-C. Liang, "Downlink channel assignment and power control for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 3106–3117, Aug. 2008.
- [63] X. Wang, "Joint sensing-channel selection and power control for cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 958–967, Mar. 2011.
- [64] M. Chiang, P. Hande, T. Lan, and C. W. Tan, "Power control in wireless cellular networks," *Found. Trends Netw.*, vol. 2, pp. 381–533, Apr. 2008.
- [65] L. Gao and S. Cui, "Multi-band power and rate control for cognitive radios with energy constraints: A dynamic programming approach," in *Proc. IEEE ICC*, pp. 3563–3567, 19-23 2008.
- [66] L. Gao and S. Cui, "Power and rate control for delay-constrained cognitive radios via dynamic programming," *IEEE Trans. Veh. Technol.*, vol. 58, pp. 4819–4827, Nov. 2009.
- [67] S. Akin and M. C. Gursoy, "Cognitive radio transmission under interference limitations and QoS constraints," in *Proc. IEEE ICC*, May 2010.
- [68] M.-V. Nguyen and H. S. Lee, "Effective scheduling in infrastructure-based cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 10, pp. 853–867, Jun. 2011.
- [69] S. Huang, X. Liu, and Z. Ding, "Distributed power control for cognitive user access based on primary link control feedback," in *Proc. IEEE INFOCOM*, Mar. 2010.
- [70] S. Huang, X. Liu, and Z. Ding, "Decentralized cognitive radio control based on inference from primary link control information," *IEEE J. Sel. Areas Commun.*, vol. 29, pp. 394–406, Feb. 2011.

- [71] F. E. Lapicciarella, X. Liu, and Z. Ding, "Distributed control of multiple cognitive radio overlay for primary queue stability," *IEEE Trans. Wireless Commun.*, vol. 12, pp. 112–122, Jan. 2013.
- [72] S. Parsaeefard and A. Sharafat, "Robust distributed power control in cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. PP, no. 99, 2012.
- [73] S. Parsaeefard and A. Sharafat, "Robust worst-case interference control in underlay cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 61, pp. 3731–3745, Oct. 2012.
- [74] S. Tang and B. Mark, "Modeling and analysis of opportunistic spectrum sharing with unreliable spectrum sensing," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 1934–1943, Apr. 2009.
- [75] A. W. Min and K. G. Shin, "Impact of mobility on spectrum sensing in cognitive radio networks," in *Proc. ACM workshop on Cognitive radio networks*, pp. 13–18, 2009.
- [76] A. Konrad, B. Y. Zhao, A. D. Joseph, and R. Ludwig, "A Markov-based channel model algorithm for wireless networks," *Wireless Networks*, vol. 9, pp. 189–199, May 2003.
- [77] G. Zheng, S. Ma, K.-K. Wong, and T.-S. Ng, "Robust beamforming in cognitive radio," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 570–576, 2010.
- [78] J. Proakis, *Digital Communications*. McGraw-Hill, Aug. 2000.
- [79] E. Dall'Anese, S.-J. Kim, G. Giannakis, and S. Pupolin, "Power control for cognitive radio networks under channel uncertainty," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3541–3551, 2011.
- [80] S. Sun, W. Ni, and Y. Zhu, "Robust power control in cognitive radio networks: A distributed way," in *Proc. IEEE ICC*, Jun. 2011.
- [81] J. Lunden, S. A. Kassam, and V. Koivunen, "Robust nonparametric cyclic correlation-based spectrum sensing for cognitive radio," *IEEE Trans. Signal Process.*, vol. 58, pp. 38–52, Jan. 2010.
- [82] H. Wang, E.-H. Yang, Z. Zhao, and W. Zhang, "Spectrum sensing in cognitive radio using goodness of fit testing," *IEEE Trans. Commun.*, vol. 8, pp. 5427–5430, Nov. 2009.

- [83] S. Rostami, K. Arshad, and K. Moessner, "Order-statistic based spectrum sensing for cognitive radio," *IEEE Commun. Lett.*, vol. 16, no. 5, pp. 592–595, 2012.
- [84] D. Denkovski, V. Atanasovski, and L. Gavrilovska, "HOS based goodness-of-fit testing signal detection," *IEEE Commun. Lett.*, vol. 16, no. 3, pp. 310–313, 2012.
- [85] L. Lu, H.-C. Wu, and S. S. Iyengar, "A novel robust detection algorithm for spectrum sensing," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 305–315, 2011.
- [86] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Operations Research*, vol. 58, pp. 595–612, May 2010.
- [87] Y. E. Nesterov and A. Nemirovski, *Interior-Point Polynomial Algorithms in Convex Programming (Studies in Applied and Numerical Mathematics)*. Society for Industrial Mathematics, 1994.
- [88] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11–12, pp. 625–653, 1999.
- [89] R. M. Gray, *Entropy and information theory*. New York, NY, USA: Springer-Verlag New York, Inc., 1990.
- [90] S. Herath, N. Rajatheva, and C. Tellambura, "Energy detection of unknown signals in fading and diversity reception," *IEEE Trans. Commun.*, vol. 59, pp. 2443–2453, Sep. 2011.
- [91] E. H. Gismalla and E. Alsusa, "Performance analysis of the periodogram-based energy detector in fading channels," *IEEE Trans. Signal Process.*, vol. 59, pp. 3712–3721, Aug. 2011.
- [92] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.
- [93] Y. Zhang, Q. Zhang, and S. Wu, "Entropy-based robust spectrum sensing in cognitive radio," *IET Communications*, vol. 4, pp. 428–436, 5 2010.

- [94] J. Tsitsiklis and M. Athans, “On the complexity of decentralized decision making and detection problems,” *IEEE Trans. Automatic Control*, vol. 30, no. 5, pp. 440–446, 1985.
- [95] K. Isii, “On sharpness of tchebycheff-type inequalities,” *Annals of the Institute of Statistical Mathematics*, vol. 14, pp. 185–197, 1962.
- [96] I. Popescu, D. Bertsimas, and Insead, *Moment Problems Via Semidefinite Programming: Applications in Probability and Finance*. INSEAD working paper, INSEAD, 2000.
- [97] D. Bertsimas and I. Popescu, “Optimal inequalities in probability theory: A convex optimization approach,” *SIAM Journal on Optimization*, vol. 15, pp. 780–804, March 2005.
- [98] M. Grotschel, L. Lovasz, and A. Schrijver, “The ellipsoid method and its consequences in combinatorial optimization,” *Combinatorica*, vol. 1, pp. 169–197, 1981.
- [99] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, Mar. 2004. Appendix A.5.5.
- [100] V. DeMiguel and F. J. Nogales, “On decomposition methods for a class of partially separable nonlinear programs,” *Mathematics of Operations Research*, vol. 33, no. 1, pp. 119–139, 2008.
- [101] K. Chowdhury and I. Akyldiz, “OFDM-based common control channel design for cognitive radio ad hoc networks,” *IEEE Trans. Mobile Comput.*, vol. 10, no. 2, pp. 228–238, 2011.
- [102] W.-H. Chung and K. Yao, “Modified hidden semi-Markov model for modelling the flat fading channel,” *IEEE Trans. Commun.*, vol. 57, no. 6, pp. 1806–1814, 2009.
- [103] Z. Chen, C.-X. Wang, X. Hong, J. Thompson, S. Vorobyov, X. Ge, H. Xiao, and F. Zhao, “Aggregate interference modeling in cognitive radio networks with power and contention control,” *IEEE Trans. Commun.*, vol. 60, pp. 456–468, Feb. 2012.

- [104] Y. Shi, Y. T. Hou, and H. Zhou, “Per-node based optimal power control for multi-hop cognitive radio networks,” *IEEE Trans. Wireless Commun.*, vol. 8, pp. 5290–5299, Oct. 2009.
- [105] S. Gong, P. Wang, and D. Niyato, “Optimal power control in interference-limited cognitive radio networks,” in *Proc. IEEE Int. Conf. Commun. Syst. (ICCS)*, pp. 82–86, Nov. 2010.
- [106] A. Yuille and A. Rangarajan, “The concave-convex procedure,” *Neural Comput.*, vol. 15, pp. 915–936, Apr. 2003.
- [107] J. Sachs, I. Maric, and A. Goldsmith, “Cognitive cellular systems within the TV spectrum,” in *Proc. IEEE DySPAN*, pp. 1–12, Apr. 2010.
- [108] W. Kumwilaisak and C.-C. J. Kuo, “Adaptive variable length Markov chain for non-stationary fading channel modeling,” in *Proc. IEEE GLOBECOM*, pp. 2046–2050, Nov. 2002.
- [109] H. Islam, Y.-C. Liang, and A. Hoang, “Joint power control and beamforming for cognitive radio networks,” *IEEE Trans. Wireless Commun.*, vol. 7, pp. 2415–2419, Jul. 2008.