Robust Performance of Spectrum Sensing in Cognitive Radio Networks

Shimin Gong, Ping Wang, Jianwei Huang

Abstract

The successful coexistence of secondary users (SUs) and primary users (PUs) in cognitive radio networks requires SUs to be spectrum aware and know which spectrum bands are occupied by PUs. Such awareness can be achieved in several ways, one of which is spectrum sensing. While existing spectrum sensing methods usually assume known distributions of the received primary signals, such an assumption is often too strong and unrealistic, and leads to unreliable detection performance in practical networks. In this paper, we design robust spectrum sensing algorithms under the distribution uncertainty of primary signals. After formulating the optimal sensing design as a robust optimization problem, we decompose it into 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 during the iterations. Numerical results verify that our robust sensing algorithm improves the worst-case detection probability and reduces the system sensitivity on decision variables.

Index Terms

Cognitive radio network, spectrum sensing, robust optimization, distribution uncertainty.

I. Introduction

Traditionally, wireless spectrum is statically allocated to license holders for exclusive use. Thus the spectrum available for new applications has become increasingly scarce and pricey. However, the static allocation mechanism is inefficient and causes great waste of spectrum resources [2]. To this end, cognitive radio (CR) [3] is proposed as a promising technique to improve spectrum utilization, as it enables efficient spectrum sharing between legacy licensed primary users (PUs) and unlicensed secondary users (SUs) as long as the SUs do not generate harmful interferences to the PUs. To realize the vision of cognitive radio, SUs need to have the capability of knowing the real time spectrum

This work is supported by MoE Tier 1 grant (Project number RG8108.M52020079) in Singapore and the General Research Fund (Project Number 412511) established under the University Grant Committee of the Hong Kong Special Administrative Region, China. Part of the results are presented in IEEE INFOCOM mini-conference 2012 [1].

Shimin Gong and Ping Wang are with the Centre for Multimedia and Network Technology (CeMNet), Nanyang Technological University, Email: {gong0012,wangping}@ntu.edu.sg; Jianwei Huang is with the Network Communications and Economics Lab (NCEL), The Chinese University of Hong Kong, Email: jwhuang@ie.cuhk.edu.hk.

usage of the PUs. The knowledge can be acquired either through the information exchange between PUs and SUs (e.g., [4]–[7]), or via SUs' spectrum sensing and prediction (e.g., [8]–[10]). The first method requires the cooperation between PUs and SUs, which introduces extra signaling overhead and sometimes may be infeasible [4], [5]. The second method does not rely on the information provided by PUs, but demands accurate sensing algorithms and collaboration among SUs. In this paper, we will focus on the robust design of spectrum sensing.

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 [11]–[13]. For example, the authors in [8], [14], [15] derived analytical expressions for the detection probability under different channel models. Specifically, in a non-fading environment, the detection statistic follows noncentral chi-square distribution. When the channel follows Rayleigh fading or log-normal shadowing, the detection probability can be further expressed as a finite series [15]. When multiple users sense the channel cooperatively, the authors in [14] summarized the analytical detection probabilities with different data fusion rules and channel conditions.

Unfortunately, knowing precise information regarding the primary signals' probability distribution is a very strong assumption that often does not hold in practice. First, it is very complicated to model the primary signals' distribution in a closed-form without significant simplifications. Second, even if it is possible to do so, the complicated series form such as in [8], [14], [15] usually makes it computational demanding to analyze performance metrics such as system throughput and sensing overhead. Third, a deterministic assumption about the primary signals' distribution may not always match the reality. For example, the received signals exhibit different distributions depending on whether there is a line-of-sight between transmitter and receiver. Besides, the mobility of wireless nodes often significantly change the signal distributions.

Some researchers have noticed this distribution uncertainty, and have tried to design robust sensing strategies that are tolerable to unpredictable channel fluctuations. The authors in [16] considered the noise fluctuation, and proposed a probabilistic method that extracts unchanging characteristic such as information entropy in the noise signal. Such entropy-based method is robust against noise power fluctuation, but it still assumes Gaussian distribution for both noise and primary signal. When the noise statistics are fully unknown, the authors in [17] proposed a nonparametric cyclic correlation detector, which does not require closed-form expressions for the noise and signals' distributions. However, such method still requires extra knowledge about the cyclic frequencies of the primary signals. In our previous work [18], we considered distribution uncertainty of both noise and primary signals in the hypothesis testing problem, and studied the lower and upper performance bounds for single user detection. In [1], we considered robustness in terms of the worst-case performance, and proposed a heuristic algorithm to find the robust decision thresholds for multiple SUs when the primary signals' distribution is unknown.

This paper presents theoretical analysis and solidifies the intuitions behind the heuristic algorithm in [1]. We consider

the parameter uncertainty and potential mismatch of the primary signals' distribution, and study their effects on the detection performance in a robust optimization problem. The main results and contributions of this paper are as follows:

- *Uncertainty Model*: We consider the case when the statistical features (i.e., mean and variance) of the primary signals are contaminated by estimation errors, and the primary signals do not follow any specific distribution function. To this end, we define the uncertainty of the primary signals' distribution based on the sample mean and variance, which are measurable in practice.
- *Max-min Robust Formulation*: We study the robustness in terms of the worst-case performance and formulate the robust design problem as a max-min problem. The minimization is over the uncertainty set for primary signals' distribution, and maximization is over the controllable decision thresholds of all sensing SUs.
- *Iterative Algorithm*: Considering the difficulty in solving the max-min problem, we propose an iterative algorithm that decomposes the problem into a series of semi-definite programs, which can be solved easily by conic optimization tools. Numerical results show that the our algorithm improves the worst-case detection probability and is robust to the changes of primary signals' distribution.

The rest of the paper is organized as follows. Section II describes the basic model and our robust design problem. Then we propose the iterative algorithm in III, and detail two main steps of the Algorithm in Section IV. Section V shows some numerical results and Section VI concludes this paper.

II. System Model

We consider a cognitive radio network with a set $\mathcal{N} = \{1, ..., N\}$ of SUs and multiple PUs operating under the same spectrum channel. The SUs have no pre-knowledge about the PUs' transmission characteristics and application types. Due to the change of active PUs on that channel, the received signals at SUs' receivers would be highly dynamic in terms of the statistic information. We assume that there is 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 [14], [19], [20]. Without loss of generality, we assume that all N SUs join the cooperation and each SU has a local energy detector.

A. Sensing Structure

The objective of spectrum sensing is to detect whether the primary signal is present on the primary channel. We assume that the local detector deploys a threshold-based decision scheme, which is commonly employed in hypothesis testing problem [15], [20]. 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 and PUs' regulations. 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, 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 (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 reinitialization 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 express the decision function as follows:

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

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 $I(A)$ 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 (which is denoted by the hypothesis H_1), otherwise it reports an idle channel (which is denoted by the hypothesis H_0).

The detection performance is measured by the receiver operating characteristic (ROC), which denotes the tradeoff between detection probability Q_d (i.e., probability of sensing a busy channel as busy) and false alarm probability Q_f (i.e., probability of sensing an idle channel as busy). When PUs are absent and thus the primary channel is idle, we denote the distribution of the received noise signal as $f_0(\boldsymbol{\xi})$. Thus the false alarm probability is given by

$$Q_f = \int_{\boldsymbol{\xi} \in S} h(\boldsymbol{\xi}, \boldsymbol{\lambda}) f_0(\boldsymbol{\xi})\, d\boldsymbol{\xi},$$

where S denotes the set of all possible sensing results in N-dimensional real space $\mathbb{R}^N$. When the PUs are present, the received signal $\boldsymbol{\xi}$ will exhibit a different distribution function $f_1(\boldsymbol{\xi})$, and we can express the detection probability as follows:

$$Q_d = \int_{\boldsymbol{\xi} \in S} h(\boldsymbol{\xi}, \boldsymbol{\lambda}) f_1(\boldsymbol{\xi})\, d\boldsymbol{\xi}.$$

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.

B. Distribution Uncertainty

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 set, as parameterized by two constants $\gamma_1 \geq 0$ and $\gamma_2 \geq 1$ [21]:

$$\mathcal{U} = \left\{ f_1(\boldsymbol{\xi}) \;\middle|\; \begin{array}{l} \mathcal{P}(\boldsymbol{\xi} \in S) = 1 \\ (E[\boldsymbol{\xi}] - \boldsymbol{\mu})^T \Sigma^{-1} (E[\boldsymbol{\xi}] - \boldsymbol{\mu}) \leq \gamma_1 \\ E[(\boldsymbol{\xi} - \boldsymbol{\mu})(\boldsymbol{\xi} - \boldsymbol{\mu})^T] \preceq \gamma_2 \Sigma \end{array} \right\}, \tag{2}$$

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 [21] has demonstrated how to choose proper values for γ_1 and γ_2 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 [16], [17], 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 paper, 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., $f_0(\boldsymbol{\xi}) = (f_0(\xi_1), \cdots, f_0(\xi_N))$. The distribution parameters of $f_0(\boldsymbol{\xi})$ can be obtained through field measurements or empirical data.

C. 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 [22]. 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:

$$\max_{\boldsymbol{\lambda}} \min_{f_1(\boldsymbol{\xi})} \quad \int_{\boldsymbol{\xi}\in S} h(\boldsymbol{\xi},\boldsymbol{\lambda}) f_1(\boldsymbol{\xi})\, d\boldsymbol{\xi} \tag{3a}$$

$$s.t. \quad f_1(\boldsymbol{\xi}) \in \mathcal{U} \tag{3b}$$

$$\int_{\boldsymbol{\xi}\in S} h(\boldsymbol{\xi},\boldsymbol{\lambda}) f_0(\boldsymbol{\xi})\, d\boldsymbol{\xi} \leq \alpha, \tag{3c}$$

where the noise distribution $f_0(\boldsymbol{\xi})$ is assumed to be known in advance. The above problem (3a)-(3c) is actually the robust counter-part of a decentralized detection problem and is therefore NP-hard [23].

However the robust problem involves a solvable sub-structure (3a)-(3b). When we are given a decision threshold $\boldsymbol{\lambda}$ in the non-convex feasible set defined by the false alarm probability constraint (3c), the inner minimization problem (3a)-(3b) is equivalent to finding the lower bound of the detection probability with moment constraints[1]. Some related algorithms were proposed in [24]–[26] to solve such kind of moment constrained problems. The most related work was presented in [21], where the authors proposed a semi-definite transformation to the problem (3a)-(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 (3c).

III. Algorithm for the Robust Design

Nevertheless, the duality transformation to a semi-definite program in [21] 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 (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 1: The max-min problem (3a)-(3c) is equivalent to (4a)-(4d) as follows:

$$\max_{\boldsymbol{\lambda},Z,Q,r} \quad (\boldsymbol{\mu}\boldsymbol{\mu}^T - \gamma_2\Sigma)\otimes Q - \Sigma \otimes P + 2\boldsymbol{\mu}^T p - \gamma_1 s - r \tag{4a}$$

$$s.t. \quad \boldsymbol{\xi}^T Q \boldsymbol{\xi} - 2\boldsymbol{\xi}^T(p + Q\boldsymbol{\mu}) + r \geq 0, \quad \forall \boldsymbol{\xi} \preceq \boldsymbol{\lambda} \tag{4b}$$

$$\begin{bmatrix} 1+r & (p+Q\boldsymbol{\mu})^T \\ p+Q\boldsymbol{\mu} & Q \end{bmatrix} \succeq 0 \tag{4c}$$

$$\int_{\boldsymbol{\xi}\in S} h(\boldsymbol{\xi},\boldsymbol{\lambda}) f_0(\boldsymbol{\xi})\, d\boldsymbol{\xi} \leq \alpha, \tag{4d}$$

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

Proof: To solve the max-min problem, we first consider the inner minimization problem (3a)-(3b) with fixed decision threshold $\boldsymbol{\lambda}$. By assigning different dual variables to the constraints in uncertainty set (3b), we can write

[1]The definition of uncertainty set $\mathcal{U}$ relates to different moments of a distribution function.

$$
\begin{aligned}
\Lambda(f_1(\boldsymbol{\xi}), Z, Q, r) &= \int_{\boldsymbol{\xi}\in S} h(\boldsymbol{\xi}, \boldsymbol{\lambda}) f_1(\boldsymbol{\xi})(\boldsymbol{\xi})\, d\boldsymbol{\xi} - \int_{\boldsymbol{\xi}\in S} Z \otimes \begin{bmatrix} \Sigma & \boldsymbol{\xi}-\boldsymbol{\mu} \\ (\boldsymbol{\xi}-\boldsymbol{\mu})^T & \gamma_1 \end{bmatrix} f_1(\boldsymbol{\xi})\, d\boldsymbol{\xi} \\
&\quad -\gamma_2 Q \otimes \Sigma + Q \otimes \int_{\boldsymbol{\xi}\in S} (\boldsymbol{\xi}-\boldsymbol{\mu})(\boldsymbol{\xi}-\boldsymbol{\mu})^T f_1(\boldsymbol{\xi})\, d\boldsymbol{\xi} + r \int_{\boldsymbol{\xi}\in S} f_1(\boldsymbol{\xi})\, d\boldsymbol{\xi} - r \\
&= \int_{\boldsymbol{\xi}\in S} \left(h(\boldsymbol{\xi}, \boldsymbol{\lambda}) + r - 2\boldsymbol{\xi}^T(p + Q\boldsymbol{\mu}) + \boldsymbol{\xi}^T Q \boldsymbol{\xi}\right) f_1(\boldsymbol{\xi})\, d\boldsymbol{\xi} \\
&\quad +(\boldsymbol{\mu}\boldsymbol{\mu}^T - \gamma_2\Sigma) \otimes Q - \Sigma \otimes Q + 2\boldsymbol{\mu}^T p - \gamma_1 s - r. \qquad (5)
\end{aligned}
$$

---

the Lagrangian function $\Lambda(f_1(\boldsymbol{\xi}), Z, Q, r)$ of the constrained problem (3a)-(3b) as in (5). 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:

$$
\max_{Z,Q,r} \quad (\boldsymbol{\mu}\boldsymbol{\mu}^T - \gamma_2\Sigma) \otimes Q - \Sigma \otimes P + 2\boldsymbol{\mu}^T p - \gamma_1 s - r \qquad (6a)
$$

$$
s.t. \quad h(\boldsymbol{\xi}, \boldsymbol{\lambda}) + r - 2\boldsymbol{\xi}^T(p + Q\boldsymbol{\mu}) + \boldsymbol{\xi}^T Q\boldsymbol{\xi} \geq 0 \quad \forall \boldsymbol{\xi} \in S. \qquad (6b)
$$

The constraint (6b) guarantees a finite value for the dual objective (6a), which is actually the lower bound of the detection probability in (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 (6b) into two constraints

$$
r - 2\boldsymbol{\xi}^T(p + Q\boldsymbol{\mu}) + \boldsymbol{\xi}^T Q\boldsymbol{\xi} \geq 0, \qquad \forall \boldsymbol{\xi} \preceq \boldsymbol{\lambda}, \qquad (7a)
$$

$$
1 + r - 2\boldsymbol{\xi}^T(p + Q\boldsymbol{\mu}) + \boldsymbol{\xi}^T Q\boldsymbol{\xi} \geq 0, \qquad \forall \boldsymbol{\xi} \not\preceq \boldsymbol{\lambda}, \qquad (7b)
$$

where $\boldsymbol{\xi} \not\preceq \boldsymbol{\lambda}$ means that there exist $n \in \mathcal{N}$ such that $\xi_n > \lambda_n$. Constraint (7a) 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 (7b), 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

$$
\begin{bmatrix} 1 + r & (p + Q\boldsymbol{\mu})^T \\ p + Q\boldsymbol{\mu} & Q \end{bmatrix} \succeq 0 \qquad (8)
$$

by Schur's complement [27]. To summarize, we can replace constraint (6b) by (7a) and (8), and obtain an equivalent form in (4a)-(4d). Note that the false alarm probability constraint (3c) is left intact as (4d) in the new problem. ■

The difficulty imposed by the false alarm probability constraint (4d) still exists in the equivalence (4a)-(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 (4d). When we manage to solve the sub-problem (4a)-(4c), we go back to this non-convex constraint (4d), and design specific search methods to improve the objective function. To recap, we solve the sub-problem (4a)-(4c) with a *fixed* decision threshold vector $\boldsymbol{\lambda}$ that satisfies (4d). Then we

will iteratively improve the objective function value in (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.

A. Initial Setting

In order to solve problem (4a)-(4c), we first need to set an initial decision threshold (IDT) in the feasible set $\mathcal{D}$ defined by constraint (4d). We define the boundary of set $\mathcal{D}$ as $\partial\mathcal{D} = \{\boldsymbol{\lambda} \mid \int_{\boldsymbol{\xi} \preceq \boldsymbol{\lambda}} f_0(\boldsymbol{\xi})\, d\boldsymbol{\xi} = 1 - \alpha\}$. 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 1: The optimal decision threshold $\boldsymbol{\lambda}^*$ to the problem (3a)-(3c) is obtained in the boundary set $\partial\mathcal{D}$, i.e.,

$$\int_{\boldsymbol{\xi} \preceq \boldsymbol{\lambda}^*} f_0(\boldsymbol{\xi})\, d\boldsymbol{\xi} = 1 - \alpha. \tag{9}$$

Proof: 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 (3a) can be rewritten as $1 - \int_{\boldsymbol{\xi} \preceq \boldsymbol{\lambda}} f_1(\boldsymbol{\xi})\, d\boldsymbol{\xi}$, thus (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 (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 (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 (9) is a necessary but not sufficient condition for the optimality of Problem (3a)-(3c). Given $\boldsymbol{\lambda}^{'} \in \partial\mathcal{D}$, we can find infinitely many $\boldsymbol{\lambda}^{''} \in \partial\mathcal{D}$ by increasing $\lambda_i^{'}$ and decreasing $\lambda_j^{'}$ $(j \neq i)$ simultaneously. ■

Since there is no need to search the interior of the feasible set $\mathcal{D}$, Proposition 1 implies that we can start the algorithm by an IDT that satisfies (9). 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 (9). 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., $\lambda_1 = \cdots = \lambda_N = \lambda_0$. 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 V 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 (9) 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 $\lambda_0 = F_0^{-1}\left((1-\alpha)^{1/N}\right)$ 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.

B. Iterative Algorithm

Given a fixed decision thresholds in the boundary set $\partial\mathcal{D}$, we now focus on solving the sub-problem (4a)-(4c). According to the equivalence between separation and optimization [25], [28], 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 (4a) is linear and (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 (4a)-(4c) only depends on (4b), and consequently the answers to the following two questions:

1) For any feasible (Q, p, r) to the problem of (4a) subject to (4c), can we check the feasibility of constraint (4b) in polynomial time?
2) If (Q, p, r) is infeasible for (4b), can we find a hyperplane that separates (Q, p, r) from the feasible set of (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 (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 (4a)-(4c) can be solved by the ellipsoid method [28] 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 (4a)-(4c) as follows:

Theorem 2: At the optimum, sub-problem (4a)-(4c) has the following form:

$$\max_{\boldsymbol{\lambda}, Q, p, q, r, z} \quad z \tag{10a}$$

$$s.t. \quad \boldsymbol{\lambda}(\boldsymbol{\lambda} - 2\boldsymbol{\mu})^T \otimes Q - 2\boldsymbol{\lambda}^T p + r \geq 0 \tag{10b}$$

$$(\boldsymbol{\mu}\boldsymbol{\mu}^T - \gamma_2 \Sigma) \otimes Q + 2\mu^T p - z - r - q \geq 0 \tag{10c}$$

$$q - 2\sqrt{\gamma_1}||Lp||_2 \geq 0 \tag{10d}$$

$$\begin{bmatrix} r+1 & (p+Q\mu)^T \\ (p+Q\mu) & Q \end{bmatrix} \succeq 0. \tag{10e}$$

A detailed proof is given in Appendix A. Problem (10a)-(10e) represents a conic optimization problem in its standard dual form, and can be easily solved by some optimization tool package such as SeDuMi [29]. Given the optimal solution Q, p, q, and r in problem (10a)-(10e), the following proposition provides an ascend search direction to update the decision threshold.

Proposition 2: If $\boldsymbol{\lambda}(t+1) = \boldsymbol{\lambda}(t) + \boldsymbol{\Delta}_t$ and $\boldsymbol{\Delta}_t^T G(t) > 0$ where $G(t) = Q \cdot \boldsymbol{\lambda}(t) - Q \cdot \boldsymbol{\mu} - p$, then $\boldsymbol{\lambda}(t+1)$ leads to an improved detection probability in problem (10a)-(10e).

The proof of Proposition 2 is given in Appendix B. Besides the search direction, the magnitude (namely, step size) of $\boldsymbol{\Delta}_t$ is also critical in the update of decision threshold. We will come back to this issue in the next section. Based

on Theorem 2 and Proposition 2, we sketch the process of our robust design in Algorithm 1, which terminates when we can no longer find a better update to improve the detection probability.

Algorithm 1 Gradient Aided Threshold Update Algorithm

Input: Sample estimates $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and parameters γ_1, γ_2 defining the distribution uncertainty

Output: Robust threshold $\boldsymbol{\lambda}$ and detection probability z

1: **initialization:**
2: **set** initial decision threshold $\boldsymbol{\lambda}_0$ according to (9), and step size $\boldsymbol{\Delta}_0$ such that $\boldsymbol{\lambda}_0 + \boldsymbol{\Delta}_0 \in \partial\mathcal{D}$
3: **while** there exists $\boldsymbol{\Delta}_t$ such that $\boldsymbol{\Delta}_t^T G(t) > 0$
4: **solve** problem (10a)-(10e) with the fixed decision threshold $\boldsymbol{\lambda}(t)$
5: **update** detection probability z and gradient $G(t+1)$ according to Proposition 2
6: **find** the optimal ascent search direction $\boldsymbol{\Delta}_{t+1}^* = \arg\max_{\boldsymbol{\Delta}} \boldsymbol{\Delta}^T G(t+1)$
7: **update** decision threshold $\boldsymbol{\lambda}(t+1) = \boldsymbol{\lambda}(t) + \boldsymbol{\Delta}_{t+1}^*$
8: **set** $t = t+1$ for next iteration
9: **end while**

IV. Gradient Aided Decision Threshold Update

To complete the algorithm, we are going to address two major steps unsolved in Algorithm 1. In the Line 3 of Algorithm 1, 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 1, 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 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.

A. Atomic Movement on Boundary

From (9), 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:

$$\Delta_j = \lambda_j - F_0^{-1}\left(F_0(\lambda_j)F_0(\lambda_i)/F_0(\lambda_i + \Delta_i)\right). \tag{11}$$

When applying this adjustment $(\Delta_i, -\Delta_j)$[2] to the current decision threshold $\boldsymbol{\lambda}$, we can ensure that the new decision threshold is still in the boundary set.

B. 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* $\boldsymbol{\Delta}$ to possess positive projection on the gradient, i.e., $\boldsymbol{\Delta}^T G(t) > 0$. 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 I[3]. The third column also shows the relations between Δ_j and Δ_i based on the analysis in Appendix C. Then we can identify those cases that lead to feasible ascent search directions. Note that there are only two non-zero terms in $\boldsymbol{\Delta}$, e.g., adjustments Δ_i and Δ_j coupled in (11), we define the projection on gradient as a function of Δ_i to simplify the analysis. For example, if we have $G_i(t) < G_j(t) < 0$ 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 I. To resolve this issue, we propose the following sufficient conditions for the existence of a feasible direction:

Proposition 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, the adjustment $(-\Delta_{i^-}, \Delta_j)$ is a feasible ascent search direction if $\Delta_{i^-} < \Delta_j < \left|\frac{G_{i^-}(t)}{G_j(t)}\right| \Delta_{i^-}$.
- For case 5, the adjustment $(\Delta_{i^-}, -\Delta_j)$ is a feasible ascent search direction if $\Delta_j > \left|\frac{G_{i^-}(t)}{G_j(t)}\right| \Delta_{i^-} > \Delta_{i^-}$.

[2]The tuple $(\Delta_{i^-}, -\Delta_j)$ represents an N-dimensional zero vector $\boldsymbol{\Delta}$ 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 $\boldsymbol{\lambda} + \boldsymbol{\Delta}$.

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

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 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 $\left[\left|\frac{G_{i^-}(t)}{G_j(t)}\right| \Delta_{i^-}, \lambda_j - \lambda_{i^-}\right]$[4] for cases 5.

C. Optimal Step Size

In Section IV-B, 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, 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 $\Delta^*_{i^-}$, we have the following property regarding the projection function.

Proposition 4: The projection $g(\Delta_{i^-})$ on gradient direction is a concave function for cases $2-5$.

The proof for Proposition 4 is given in Appendix D. The concavity guarantees the uniqueness of the optimal step size, and we can find its optimal value $\Delta^*_{i^-}$ 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 3. Specifically, we have the following results for cases $2-5$.

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

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

Theorem 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 decreasing for cases $2-5$ as shown in Proposition 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 3. The detailed procedures

[4]Here $\Delta_j < \lambda_j - \lambda_{i^-}$ ensures $\lambda^{'}_j > \lambda_{i^-}$ as requested in case 5 of Table I.

[5]We take $\Delta^u_{i^-} = +\infty$ if there is no solution.

are given in Algorithm 2. The Algorithm starts in Line 3 by checking whether the sufficient condition $g^{'}(0) > 0$ is satisfied as indicated by Theorem 3. Line 4 sets the lower and upper boundary points of the feasible set of Δ_{i^-}. Note that the feasible set $\boldsymbol{Z}$ varies according to different cases. Then an iterative process in line $5-10$ shrinks the search region until the optimal step size is found. This optimal value $\Delta_{i^-}^*$ is either a solution to $g^{'}(\Delta_i^-) = 0$ or one of the two end points of set $\boldsymbol{Z}$. When updating the decision thresholds in Line 12 and 13, notation $\oplus$ represents either addition $(+)$ or substraction $(-)$ according to different cases in Table I.

Algorithm 2 Update Decision Threshold

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

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

1: **initialization:**
2: **initiate** $\boldsymbol{\lambda}(t+1)$ by its previous value $\boldsymbol{\lambda}(t)$, and take j as any node in set $\mathcal{N}$ with $j \neq i^-$
3: **if** $g^{'}(0) > 0$ for cases $2-5$
4: **set** the *lower* and *upper* end points of the interval set as $\Delta_- = \boldsymbol{Z}_{min}$, $\Delta^- = \boldsymbol{Z}_{max}$
5: **while** the derivative of the middle point is not close to zero, i.e., $\left|g^{'}\left((\Delta_- + \Delta^-)/2\right)\right| > \epsilon$, and $|\Delta_- - \Delta^-| > \epsilon$
6: **set** $\bar{\Delta}$ as the middle point between Δ_- and Δ^-
7: **if** derivative $g^{'}(\cdot)$ has different sign at $\bar{\Delta}$ and Δ_-, i.e., $g^{'}(\bar{\Delta})g^{'}(\Delta_-) < 0$
8: **then set** the new upper point $\Delta^- = \bar{\Delta}$
9: **otherwise** set the new lower point $\Delta_- = \bar{\Delta}$
10: **end while**
11: **set** the optimal step size $\Delta_{i^-}^* = (\Delta_- + \Delta^-)/2$
12: **update** the decision threshold of user i^-, i.e., $\lambda_{i^-}(t+1) = \lambda_{i^-}(t) \oplus \Delta_{i^-}^*$
13: **update** the decision threshold of user j, i.e., $\lambda_j(t+1) = \lambda_j(t) \oplus \Delta_j$ and Δ_j is given by (11)
14: **end if**

V. 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 $\Sigma = 2I_N$ where $\boldsymbol{I}_N$ is an identity matrix with size N. The distribution uncertainty parameters in set (2) are $\gamma_1 = 0.02$ and $\gamma_2 = 1.2$. 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.

A. Comparison between Robust and Nominal Design

In Fig. 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 [11], [12], [14]. 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 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 Fig. 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 (2), instead on a particular distribution.

We also compare our robust threshold design (denoted as RTD in Fig. 2) with an existing work in [30], which considers a robust fusion weight (denoted as RFW in Fig. 2) approach to combine signal samples from different SUs. The RFW assumes that the signal samples follow the normal distribution with uncertain mean and variance. Let the uncertainty size (parameters γ_1 and γ_2) be the same for RTD and RFW. We test the RTD and RFW with random signal samples generated from different distribution functions (i.e., normal, exponential, and log-normal distributions, denoted by Normal Dist., Exp. Dist., and LogN. Dist. in Fig. 2, respectively), whose mean and variance are constrained in set (2). We note that, the detection performance of the RFW approach can significantly deviate from the normal distribution benchmark when the signal samples are generated from other distributions. This implies that the RFW approach is very sensitive to distribution uncertainty, and the practical detection performance is difficult to predict. However, our RTD approach is robust against the uncertainty of different distribution functions, and its detection performance does not change significantly when signal distribution changes. In the following, we further investigate various properties of our proposed robust threshold design.

B. Algorithm Convergence

We have assumed in Section III-A 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: $\boldsymbol{\lambda}_0^1 = [1.82, 1.82, 1.82]$, $\boldsymbol{\lambda}_0^2 = [1.55, 1.82, 2.42]$, and $\boldsymbol{\lambda}_0^3 = [1.52, 1.82, 2.72]$. 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., $Q_f = \int_{\boldsymbol{\xi} \in S} h(\boldsymbol{\xi}, \boldsymbol{\lambda}) f_0(\boldsymbol{\xi})\, d\boldsymbol{\xi} = \alpha$ for any $\boldsymbol{\lambda} \in \{\boldsymbol{\lambda}_0^1, \boldsymbol{\lambda}_0^2, \boldsymbol{\lambda}_0^3\}$.

We plot the dynamics of the projections on gradient, detection performance, and SUs' decision thresholds in Figs. 3-5, respectively. In each iteration, our algorithm will check all possible strategies listed in Table I, and then determine a search direction and the optimal step size that has the largest projection on its gradient direction as shown in Fig. 3. 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 Fig. 4, where IDTs are set to $\boldsymbol{\lambda}_0^1$, $\boldsymbol{\lambda}_0^2$ and $\boldsymbol{\lambda}_0^3$, respectively. Though the initial detection probabilities are not the same, the differences due to different initializations quickly diminish as the iteration increases. With the decrease of gradient projections, the detection probability gradually converges to the maximal value. In Fig. 5, each SU's decision threshold also converges to a common value under different initialization choices (and different SUs have different values).

C. Reverse Ordering between $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$

Fig. 5 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 Fig. 5 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 $\boldsymbol{\lambda}$ and the received signal strengths $\boldsymbol{\mu}$, 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 V-B, while changing μ_3 from 3.4 to 4.3. The IDTs are set to $\boldsymbol{\lambda}_0^2 = [1.55, 1.82, 2.42]$ during this process. The dynamics of SUs' final decision thresholds are illustrated in Fig. 6. 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 Fig. 5 with $\lambda_3 < \lambda_2 < \lambda_1$. There are two interaction points among curves in Fig. 6, which correspond to the signal 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 $\lambda_1 > \lambda_2 = \lambda_3$.

VI. Conclusion

In this paper, we consider the problem of robust cooperative sensing in cognitive radio networks. The key contribution of this paper 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 solving the constrained robust optimization problem directly is very difficult, 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.

Our future research will consider bringing more historical information into the definition of signal uncertainty. In this paper, the historical information only refers to the signal statistics, and the resulting detection bounds may be too

conservative. To make a better use of the sensing history and empirical data, we can extract a closed-form distribution that serves as a reference. This closed-form reference distribution may draw from elaborate channel modeling as in [14], [15], [31], or online goodness-of-fit test for the historical signal samples. The actual signal distributions are more or less close to this reference, and the uncertainty can be defined in terms of a distance measure in probability. With this uncertainty set, we expect to have a better characterization of the real-world uncertainty and enhance the detection performance.

APPENDIX

A. Proof for Theorem 2

In sub-problem (4a)-(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 (4a). Let $z = (\boldsymbol{\mu}\boldsymbol{\mu}^T - \gamma_2\Sigma) \otimes Q - \Sigma \otimes P + 2\boldsymbol{\mu}^T p - \gamma_1 s - r$. As $Z = \left[\begin{smallmatrix} P & p \\ p^T & s \end{smallmatrix}\right] \succeq 0$, we discuss two cases for the variable s as in [21]: $s > 0$ or $s = 0$. Considering the first case of $s > 0$, we have $P \succeq \frac{1}{s} pp^T$ and

$$\begin{aligned} z &\leq (\boldsymbol{\mu}\boldsymbol{\mu}^T - \gamma_2\Sigma) \otimes Q - \frac{1}{s} p^T \Sigma p + 2\boldsymbol{\mu}^T p - \gamma_1 s - r \\ &\leq (\boldsymbol{\mu}\boldsymbol{\mu}^T - \gamma_2\Sigma) \otimes Q + 2\boldsymbol{\mu}^T p - r - 2\sqrt{\gamma_1}||Lp||_2. \end{aligned} \tag{12}$$

Here L is a triangular matrix denoting the Cholesky decomposition of Σ, i.e., $\Sigma = L^T L$. Therefore, the objective in (4a) can be represented as the new objective (10a) together with a linear constraint (10c) and a quadratic constraint (10d). Considering the second case of $s = 0$, we must have $p = 0$ by the positive semidefiniteness of $Z = \left[\begin{smallmatrix} P & p \\ p^T & 0 \end{smallmatrix}\right] \succeq 0$, otherwise ($p \neq 0$) we can construct a vector y by appending any scalar $x > \frac{p^T P p}{2 p^T p}$ to the vector p such that

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

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 (4a). Therefore the new objective degenerates to $z = (\boldsymbol{\mu}\boldsymbol{\mu}^T - \gamma_2\Sigma) \otimes Q - r$, which has exactly the same form as (12). For simplicity, we combine the resulting formulations from these two cases (i.e., $s > 0$ and $s = 0$) into a general form (10a)-(10e). In the following, we prove the equivalence between (4b) and (10b) when the decision threshold $\boldsymbol{\lambda}$ achieves its optimum.

Note that $\boldsymbol{\lambda}$ only appears in constraint (4b), which is equivalent to see whether $\Phi(\boldsymbol{\xi}^*|Q, p, r) \geq 0$, where $\boldsymbol{\xi}^* = \arg\min_{\boldsymbol{\xi} \preceq \boldsymbol{\lambda}} \Phi(\boldsymbol{\xi}|Q, p, r)$. If $\boldsymbol{\xi}^*$ is optimal to the minimization of $\Phi(\boldsymbol{\xi}|Q, p, r)$ and $\xi_i^* = \lambda_i$ for some $i \in \mathcal{N}$, we say that the i^{th} term of the box constraint (i.e., $\boldsymbol{\xi} \preceq \boldsymbol{\lambda}$)[6] is active. If $\xi_i^* < \lambda_i$, then this term is inactive. We can consider several different cases:

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

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

$$\begin{aligned} z &\leq (\boldsymbol{\mu}\boldsymbol{\mu}^T - \gamma_2\Sigma) \otimes Q + 2\mu^T p - r - 2\sqrt{\gamma_1}||Lp||_2 \\ &\leq -\left(\gamma_2\Sigma \otimes Q + 2\sqrt{\gamma_1}||Lp||_2 + p^T Q^{-1} p\right) \leq 0. \end{aligned}$$

  Therefore $\boldsymbol{\lambda}$ can not be the optimizer since negative detection probability is meaningless in practice.

- 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 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 (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 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 $\boldsymbol{\xi}^* = \boldsymbol{\lambda} \preceq \boldsymbol{\mu} + Q^{-1}p$ and $\Phi(\boldsymbol{\xi}^*|Q,p,r) = \boldsymbol{\lambda}^T Q\boldsymbol{\lambda} - 2\boldsymbol{\lambda}^T(p+Q\boldsymbol{\mu}) + r$. Therefore, we remove the free variable $\boldsymbol{\xi}$ from sub-problem (4a)-(4c) and reduce (4b) to a linear constraint with respect to (Q,p,r).

B. Proof for Proposition 2

We first prove the differentiability of the problem (10a)-(10e) with respect to $\boldsymbol{\lambda}$. Considering problem (10a)-(10e) 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 $\boldsymbol{x} = (Q,p,q,r,z)$, i.e., $\boldsymbol{x} \in \mathcal{C}(\boldsymbol{\lambda})$. The proposed iterative algorithm decomposes the parametric optimization problem into a sub-problem (10a)-(10e) with fixed decision threshold $\boldsymbol{\lambda}$, and a master problem that maximizes the optimal value function $z^*(\boldsymbol{\lambda}) = \max_{\boldsymbol{x}\in\mathcal{C}(\boldsymbol{\lambda})} z$, by updating decision threshold $\boldsymbol{\lambda}$ based on $\boldsymbol{x}(\boldsymbol{\lambda}) = \arg\max_{\boldsymbol{x}\in\mathcal{C}(\boldsymbol{\lambda})} z$. For any fixed $\boldsymbol{\lambda}$ satisfying (9), it is easy to check that the sub-problem (10a)-(10e) is a convex problem and can be solved efficiently by interior point methods. It was proved in [32] 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 $G(t) = Q\boldsymbol{\lambda}(t) - Q\boldsymbol{\mu} - p$ is a gradient direction of $\boldsymbol{\lambda}$ and can improve the objective (10a) in iteration $t+1$. For the ease of presentation, we denote $\boldsymbol{x}(t) = (Q_t, p_t, q_t, r_t, z_t)$. From (10b)-(10c), the optimum $z^*(\boldsymbol{\lambda}(t))$ can be set to $\boldsymbol{\lambda}(t)^T Q_t \boldsymbol{\lambda}(t) - 2\boldsymbol{\lambda}(t)^T(p_t + Q_t\boldsymbol{\mu}) + (\boldsymbol{\mu}\boldsymbol{\mu}^T - \gamma_2\Sigma) \otimes Q_t + 2\mu^T p_t - q_t$, where the last three terms are independent of $\boldsymbol{\lambda}(t)$ and can be viewed as constants. Now, considering $z^*(\boldsymbol{\lambda}(t))$ as a function of $\boldsymbol{\lambda}(t)$ with gradient $G(t) = Q_t\boldsymbol{\lambda}(t) - Q_t\boldsymbol{\mu} - p_t$, we can increase the objective by updating the decision threshold as $\boldsymbol{\lambda}(t+1) = \boldsymbol{\lambda}(t) + \boldsymbol{\Delta}_t$ with $\boldsymbol{\Delta}_t^T G(t) > 0$. Moreover, we still have $\boldsymbol{x}(t) \in \mathcal{C}(\boldsymbol{\lambda}(t+1))$, which also implies $z^*(\boldsymbol{\lambda}(t+1)) \geq z^*(\boldsymbol{\lambda}(t))$ in the next iteration.

C. The Relation Between Δ_i and Δ_j in Table I

Note that Δ_j is a function of $\boldsymbol{\lambda}$ and Δ_i as in (11). To quantify the relation between Δ_j and Δ_i, we have the following proposition that explain the third column in Table I.

Proposition 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 $\boldsymbol{\lambda}^{'}$ is in the boundary set $\partial\mathcal{D}$:

1) If $\lambda_i^{'} = \lambda_i + \Delta_i$ and $\lambda_j^{'} = \lambda_j - \Delta_j$, then $\Delta_i > \Delta_j$.
2) If $\lambda_i^{'} = \lambda_i - \Delta_i$ and $\lambda_j^{'} = \lambda_j + \Delta_j$, then either $\Delta_i < \Delta_j$ if $\lambda_j > \lambda_i^{'}$, or $\Delta_i > \Delta_j$ if $\lambda_j < \lambda_i^{'}$.

We can 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., $F_0(\lambda_i)F_0(\lambda_j) = F_0(\lambda_i^{'})F_0(\lambda_j^{'})$. Therefore in 1) we have

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

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

$$\begin{aligned} R_x^{'}(x, a) &= \left[f_0(x)F_0(x + a) - f_0(x + a)F_0(x)\right] / F_0^2(x + a) \\ &> \left[f_0(x) - f_0(x + a)\right] / F_0(x + a) > 0, \end{aligned}$$

where $f_0(x)$ is the standard normal distribution function. When $\lambda_i > \lambda_j^{'}$, we thus have $R(\lambda_i, \Delta_i) > R(\lambda_j^{'}, \Delta_i)$, i.e.,

$$F_0(\lambda_j^{'})/F_0(\lambda_j^{'} + \Delta_j) = F_0(\lambda_i)/F_0(\lambda_i + \Delta_i) > F_0(\lambda_j^{'})/F_0(\lambda_j^{'} + \Delta_i).$$

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 $\lambda_j^{'} = \lambda_j + \Delta_j$.

D. Proof for Proposition 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 $\lambda_i^{'}$, $\lambda_j^{'}$ 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 $g(\Delta_i)$.

REFERENCES

[1] S. Gong, P. Wang, and J. Huang, "Robust threshold design for cooperative sensing in cognitive radio networks," in *Proceedings of the IEEE International Conference on Computer Communications: Mini-Conference*, Mar. 2012.

[2] M. Mchenry, "Spectrum white space measurements," *New America Foundation Broadband Forum (June 2003)*.

[3] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[4] R. Southwell, J. Huang, and X. Liu, "Spectrum mobility games," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM 2012)*, March 2012.

[5] X. Chen and J. Huang, "Game theoretical approach for database assisted database sharing," *IEEE ICDCS*, June 2012.

[6] L. Duan, L. Gao, and J. Huang, "Contract-based cooperative spectrum sharing," in *Proceeding of the IEEE Dynamic Spectrum Access Networks (DySPAN)*, Aachen, Germany, May 2011.

[7] L. Gao, X. Wang, Y. Xu, and Q. Zhang, "Spectrum trading in cognitive radio networks: A contract-theoretic modeling approach," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, pp. 843–855, April 2011.

[8] A. Ghasemi and E. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *Proceeding of the IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (IEEE DySPAN 2005)*, Nov. 2005, pp. 131–136.

[9] S. Huang, X. Liu, and Z. Ding, "Optimal transmission strategies for dynamic spectrum access in cognitive radio networks," *IEEE Transactions on Mobile Computing*, vol. 8, no. 12, pp. 1636–1648, Dec. 2009.

[10] H. T. Cheng and W. Zhuang, "Simple channel sensing order in cognitive radio networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, pp. 676–688, April 2011.

[11] Y.-C. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 4, pp. 1326–1337, April 2008.

[12] W.-Y. Lee and I. F. Akyildiz, "Optimal spectrum sensing framework for cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 10, pp. 3845–3857, October 2008.

[13] W. Yin, P. Ren, Q. Du, and Y. Wang, "Delay and throughput oriented continuous spectrum sensing schemes in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. PP, no. 99, 2012.

[14] F. F. Digham, M.-S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 21–24, Jan. 2007.

[15] A. Rao and M. Alouini, "Performance of cooperative spectrum sensing over non-identical fading environments," *IEEE Transactions on Communications*, vol. 59, no. 12, pp. 3249–3253, Dec. 2011.

[16] Y. Zhang, Q. Zhang, and S. Wu, "Entropy-based robust spectrum sensing in cognitive radio," *IET Communications*, vol. 4, no. 4, pp. 428–436, 5 2010.

[17] J. Lunden, S. Kassam, and V. Koivunen, "Robust nonparametric cyclic correlation-based spectrum sensing for cognitive radio," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 38–52, Jan. 2010.

[18] S. Gong, P. Wang, and W. Liu, "Spectrum sensing under distribution uncertainty in cognitive radio networks," in *Proceedings of the IEEE International Conference on Communications*, Jun. 2012.

[19] G. Taricco, "Optimization of linear cooperative spectrum sensing for cognitive radio networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 1, pp. 77–86, Feb. 2011.

[20] K. Ben Letaief and W. Zhang, "Cooperative communications for cognitive radio networks," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 878–893, May 2009.

[21] 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.

[22] A. Ben-tal and A. Nemirovski, "Robust solutions of linear programming problems contaminated with uncertain data," *Mathematical Programming*, vol. 88, pp. 411–424, 2000.

[23] J. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Transactions on Automatic Control*, vol. 30, no. 5, pp. 440–446, 1985.

[24] K. Isii, "On sharpness of tchebycheff-type inequalities," *Annals of the Institute of Statistical Mathematics*, vol. 14, pp. 185–197, 1962.

[25] I. Popescu, D. Bertsimas, and Insead, *Moment Problems Via Semidefinite Programming: Applications in Probability and Finance*, ser. INSEAD working paper. INSEAD, 2000.

[26] 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.

[27] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, Mar. 2004, appendix A.5.5.

[28] M. Grotschel, L. Lovasz, and A. Schrijver, "The ellipsoid method and its consequences in combinatorial optimization," *Combinatorica*, vol. 1, pp. 169–197, 1981.

[29] 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.

[30] A. Mutapcic and S.-J. Kim, "Robust signal detection under model uncertainty," *IEEE Signal Processing Letters*, vol. 16, no. 4, pp. 287–290, Apr. 2009.

[31] M. Mohammadkarimi, B. Mahboobi, and M. Ardebilipour, "Optimal spectrum sensing in fast fading rayleigh channel for cognitive radio," *IEEE Communications Letters*, vol. 15, no. 10, pp. 1032–1034, Oct. 2011.

[32] 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.

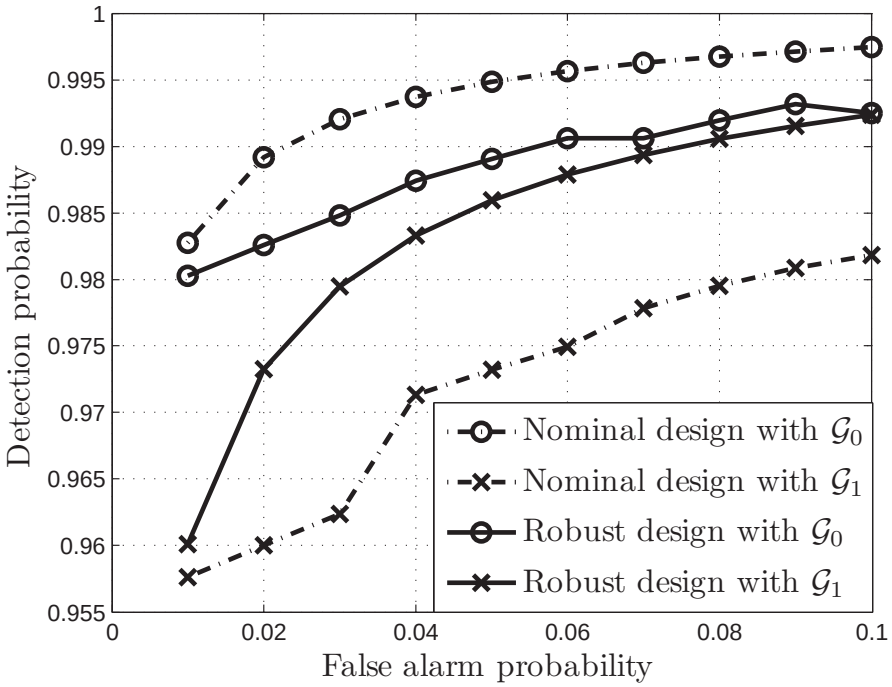

Fig. 1: Performance comparison between the nominal and robust designs.

TABLE I: Strategies given $G_i(t) < G_j(t) < 0$

Thresholds	Adjustment strategies	Outcomes		Cases
$\lambda_i \geq \lambda_j$	$\lambda_i^{'} = \lambda_i + \Delta_i$, $\lambda_j^{'} = \lambda_j - \Delta_j$	$\Delta_i > \Delta_j$		case 1
	$\lambda_i^{'} = \lambda_i - \Delta_i$,	$\lambda_i^{'} > \lambda_j$	$\Delta_i > \Delta_j$	case 2
	$\lambda_j^{'} = \lambda_j + \Delta_j$	$\lambda_i^{'} \leq \lambda_j$	$\Delta_i < \Delta_j$	case 3
$\lambda_i < \lambda_j$	$\lambda_i^{'} = \lambda_i - \Delta_i$, $\lambda_j^{'} = \lambda_j + \Delta_j$	$\Delta_i < \Delta_j$		case 4
	$\lambda_i^{'} = \lambda_i + \Delta_i$,	$\lambda_j^{'} > \lambda_i$	$\Delta_j > \Delta_i$	case 5
	$\lambda_j^{'} = \lambda_j - \Delta_j$	$\lambda_j^{'} < \lambda_i$	$\Delta_j < \Delta_i$	case 6

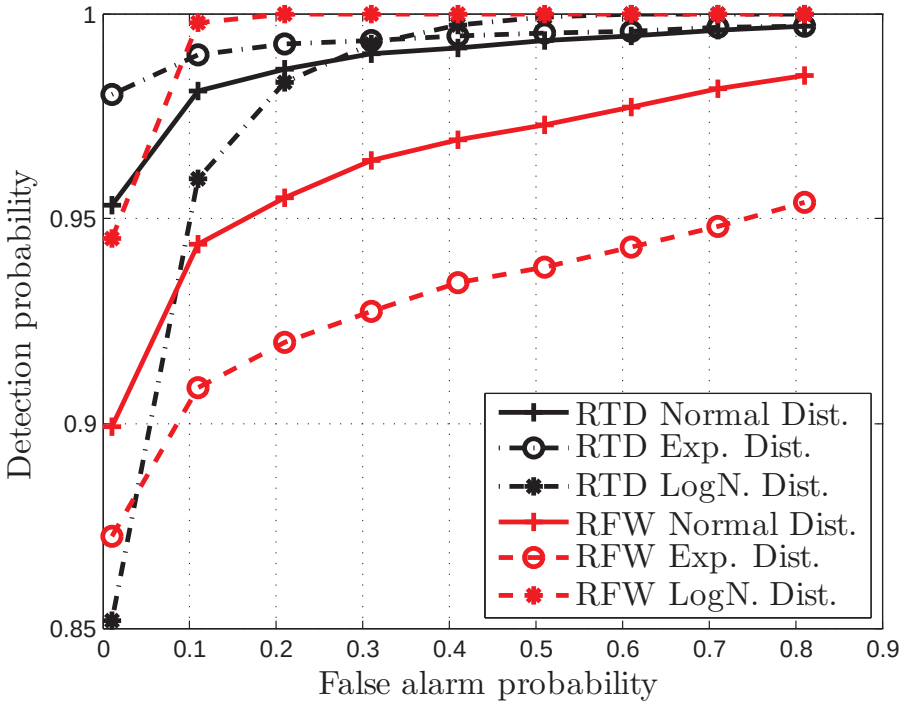

Fig. 2: Performance comparison between RFW and RTD.

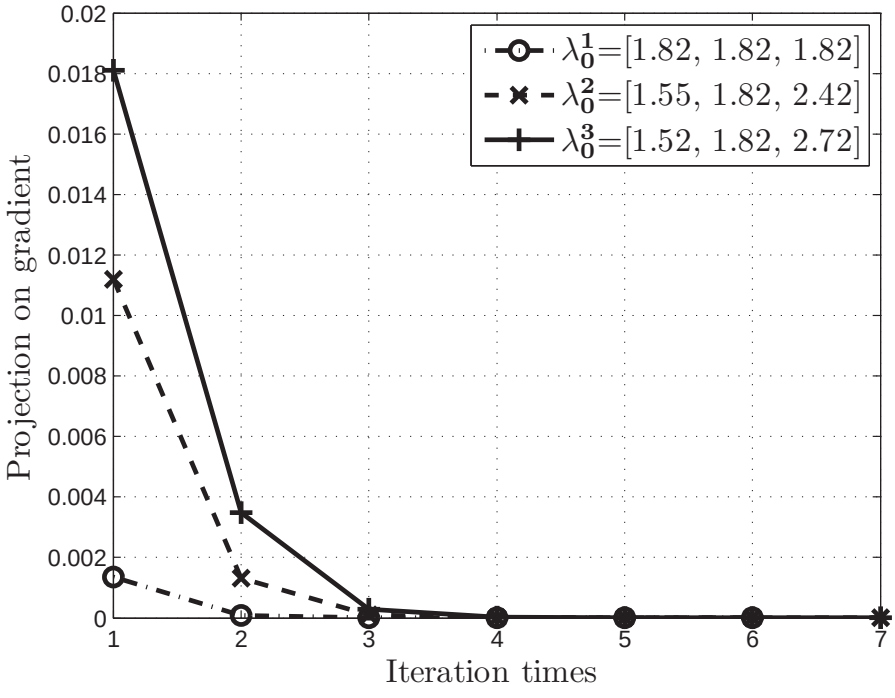

Fig. 3: The values of projection on gradient direction decrease in each iteration.

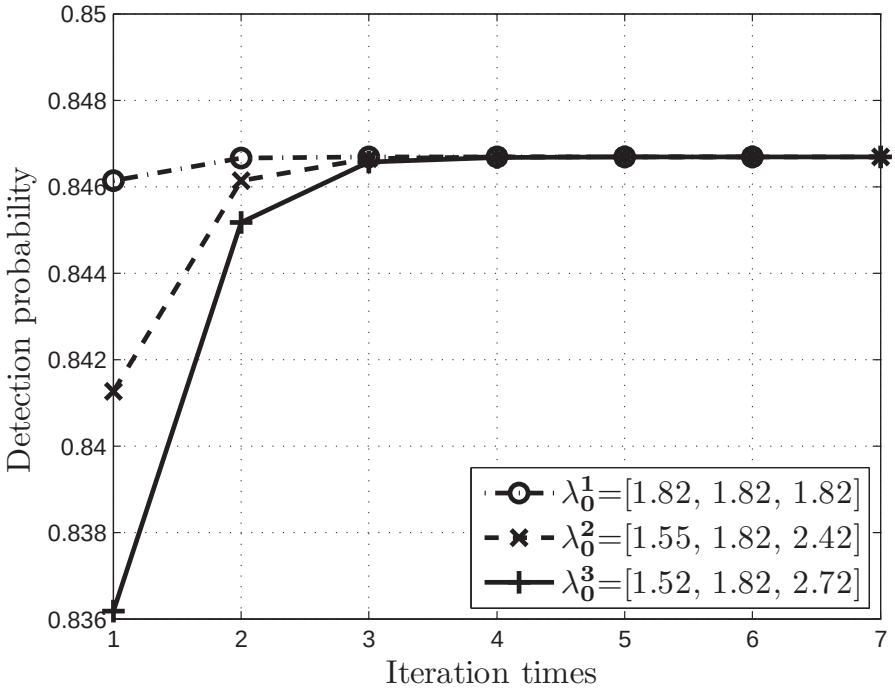

Fig. 4: Detection probabilities converge to the same level irrespective of different IDT settings.

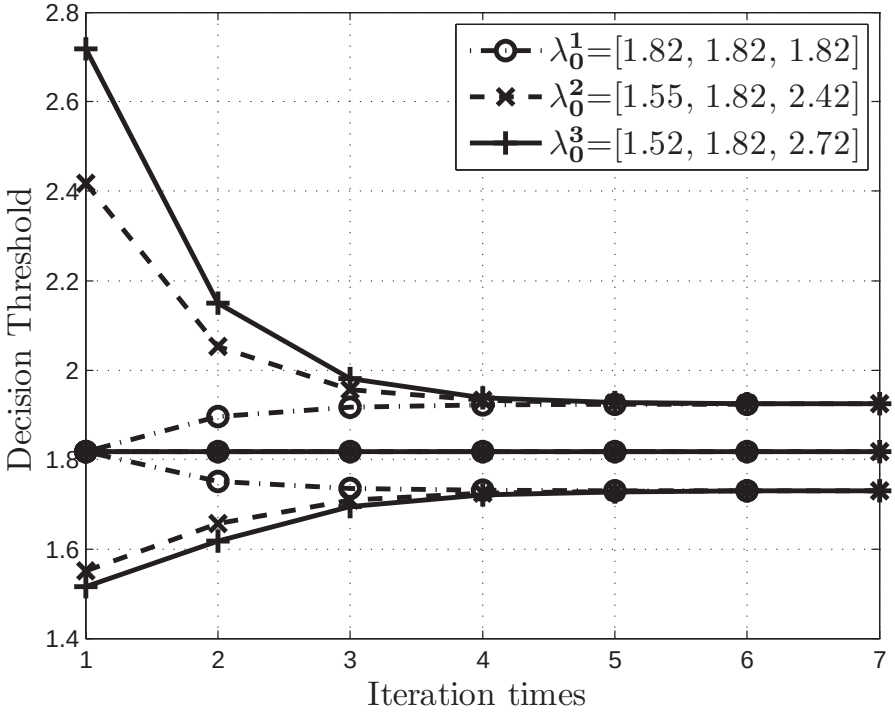

Fig. 5: Each SU's decision threshold converges to a common value under different IDT settings (and different SUs have different values).

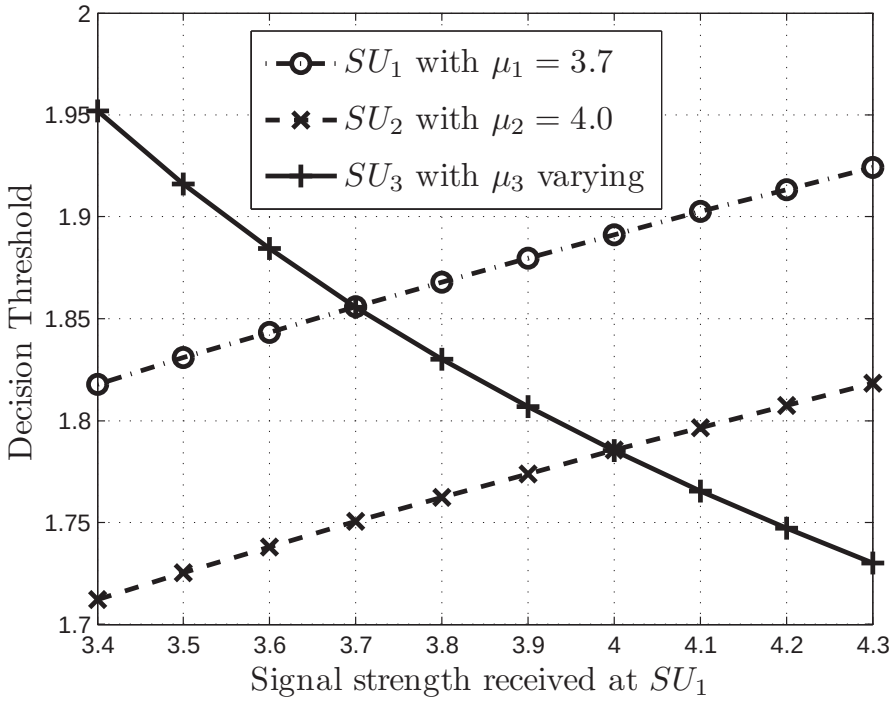

Fig. 6: Decision thresholds vary with their received signal strength.