A Robust Cooperative Sensing Approach for Incomplete and Contaminated Data

Rui Zhou, Wenqiang Pu, Ming-Yi You, and Qingjiang Shi

***Abstract*—Cooperative sensing utilizes multiple receivers dispersed across different locations, capitalizing on the advantages of multiple antennas and spatial diversity gain. This mechanism is crucial for monitoring the availability of licensed spectrum for secondary use when free from primary users. However, the efficacy of cooperative sensing relies heavily on the flawless transmission of raw data from cooperating receivers to a fusion center, a condition that may not always be fulfilled in real-world scenarios. This study investigates cooperative sensing in the context where the raw data is compromised by errors introduced during transmission, attributable to a relatively high bit error rate (BER). Consequently, the data received at the fusion center becomes incomplete and contaminated. Conventional multiantenna detectors are not adequately designed to handle such situations. To overcome this, we introduce the missing-data t-distribution generalized likelihood ratio test (mtGLRT) detectors to manage such problematic data at the fusion center. The structured covariance matrices are estimated from this problematic data. Efficient optimization algorithms using the generalized expectation-maximization (GEM) method are developed accordingly. Numerical experiments corroborate the robustness of the proposed cooperative sensing methods.**

***Index Terms*—Cooperative sensing, transmission error, incomplete data, contaminated data, robust detector.**

I. Introduction

COGNITIVE radio (CR) communication is widely regarded as a vital technology for fifth-generation (5G) wireless communication and Internet of Things (IoT) systems, offering increased spectrum efficiency and data transmission speeds by leveraging opportunistic spectrum access to other available networks [1], [2], [3]. This technology enables CR network users to utilize the frequency band allocated to primary users during periods of inactivity [4]. For CR network users, spectrum sensing technology is indispensable in monitoring the occupancy status of the desired frequency band. The cooperative sensing, a spectrum sensing method, leverages the distributed antennas of CR network users to augment spatial diversity gain [5], [6], [7]. For instance, a group of unmanned aerial vehicles (UAVs) equipped with omnidirectional antennas could utilize cooperative sensing to detect primary signals more effectively. Typically, a designated node serves as the fusion center, tasked with analyzing the aggregated data. A significant challenge in cooperative sensing lies in the reliable transmission of data from individual CR nodes to the fusion center. In practical scenarios, achieving near prefect transmission of raw data from multiple receivers to the fusion center is challenging due to the inevitable transmission errors [8], [9]. The imperfections in the data collected at the fusion center are discussed as follows.

Incomplete Data: If transmission errors can be accurately identified, such as being linked to a specific data element or a small block of data elements, these erroneous data elements can be eliminated. This process results in an incomplete dataset, i.e., some data elements are missing at the fusion center, as illustrated in Fig. 1. Numerous studies have explored similar problems arising from incomplete data in fields such as finance [10], [11], radar [12], and wireless sensor networks [13]. In general, there are two primary and accessible techniques for handling missing data: the deletion and the imputation techniques [14]. The deletion method removes any samples containing missing entries from the analysis. In contrast, the imputation technique substitutes missing data with estimated values, after which the completed dataset is utilized for further analysis. For example, the most basic methods involve using adjacent values or the mean to impute missing data [15]. Another option is to employ a Gaussian distribution model for data imputation, as detailed in [16]. More sophisticated methods presume prior data structures to inform the imputation process, such as low-rank data completion [17] and regression imputation [18]. However, both of these methods have their limitations. The deletion method might eliminate informative data. Simple imputation methods can introduce errors or disrupt the statistical properties of the data. Moreover, in typical cooperative sensing scenario, the presence of predominant noise obscures any underlying

Manuscript received 22 April 2024; revised 25 July 2024; accepted 20 August 2024. Date of publication 22 August 2024; date of current version 9 September 2024. This work was supported in part by the National Nature Science Foundation of China (NSFC) under Grant 62201362 and Grant 62101350 and in part by Shenzhen Science and Technology Program under Grant RCBS20221008093126071. The associate editor coordinating the review of this article and approving it for publication was Dr. Shaofeng Zou. *(Corresponding authors: Wenqiang Pu; Ming-Yi You.)*

Rui Zhou and Wenqiang Pu are with Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong-Shenzhen, Shenzhen 518172, China (e-mail: rui.zhou@sribd.cn; wpu@sribd.cn).

Ming-Yi You is with the National Key Laboratory of Electromagnetic Space Security, Jiaxing 314033, China (e-mail: youmingyi@126.com).

Qingjiang Shi is with the School of Software Engineering, Tongji University, Shanghai 200092, China, and also with Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong-Shenzhen, Shenzhen 518172, China (e-mail: shiqj@tongji.edu.cn).

Digital Object Identifier 10.1109/TSP.2024.3448498

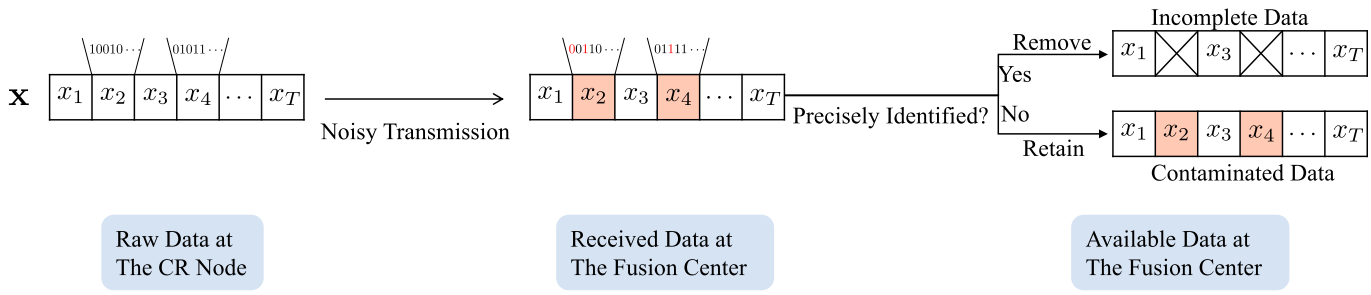


Fig. 1. Illustration of the incomplete and contaminated data at the fusion center.

structural information, thereby hindering the use of imputation techniques that rely on such structures. An alternative approach to handling missing data involves developing tailored methods for specific applications. For example, robust modeling of partially observed data using a vector autoregression model has been examined in [11]. These tailored methods often yield better performance in applications but require additional effort in specific algorithm development.

Contaminated Data: Furthermore, there is also a possibility that certain transmission errors may not be accurately identified, leading to a dataset at the fusion center that is not just incomplete, but contaminated as well. For example, the parity bit, a common error detection mechanism, fails to catch an even number of bit errors, as detailed in [19]. This leads to a contaminated dataset, which can significantly impair the performance of the existing detection algorithms. To address it, typical solutions involve discarding the corrupt data points, employing methods such as the local outlier factor (LOF) and the interquartile range method [20]. These methods assume that contaminated data behave as outliers, which are substantially different from the majority of the data points in a distribution. Nonetheless, these approaches have clear limitations: they are ineffective if the contaminated data do not distinctly differ from other data points, and they often rely on additional hyperparameters to distinguish the contamination.

An alternative strategy to resist contamination is to account for the presence of contaminated data items within the dataset. There has been extensive research in this domain. For example, a robust generalized likelihood ratio test (GLRT) detector has been developed to counter adversarial perturbations in classification tasks [21] and has been evaluated against the minimax approach [22]. Another noteworthy study has focused on the detection of false data injection attacks in electrical grids [23]. In the domain of cooperative sensing, extensive research has been conducted across various levels of data fusion. For instance, the issue of Byzantine attacks is addressed in decision fusion scenarios, where CR users independently determine the presence of primary signals and subsequently transmit their decisions to a fusion center [24]. The challenge of fraudulent data is tackled in the context of power information fusion, with CR users sending their computed energy readings to the fusion center [25], [26]. Likewise, the problem of impulsive noise is examined in [27], which aligns with the focus of this paper on raw data fusion. In this realm, the M-estimator has been introduced to robustly estimate the covariance matrix of the gathered signals, and the detection process utilizes the ratio between the largest eigenvalues of this matrix and the average power of the received signals. It is important to note that these existing robust detection methods cannot be directly applied to scenarios of raw data fusion in the presence of missing data.

The presence of incomplete and contaminated data at the fusion center poses significant challenges for effective spectrum sensing. Conventional non-parametric methods, such as energy detection (ED) [28], the eigenvalue arithmetic-to-geometric mean (AGM) [29], and the eigenvalue-moment-ratio (EMR) detector [30], rely on calculations involving the power or covariance matrix of received signals. Similarly, the standard parametric approach, exemplified by the GLRT detector [31], depends on the estimation of received signals' statistical parameters. When these traditional techniques encounter incomplete and contaminated data, a preprocessing step is required to address missing data points before applying the methods, and the presence of contaminated data inevitably impairs detection performance. The impact of such data anomalies is anticipated to be pronounced, particularly because many extant detection algorithms, including the classical GLRT method, assume that the received data follow a Gaussian distribution—a condition known to be highly susceptible to outliers. To our knowledge, no existing studies have directly addressed the challenges posed by incomplete and contaminated datasets in cooperative sensing, particularly where the fusion center collects raw data from CR nodes.

To this end, the primary goal of this paper is to develop a robust cooperative spectrum sensing approach for incomplete and contaminated data. The main contributions of this paper are as follows:

- A pronounced heavy-tailed distribution is identified within the contaminated dataset. A quantitative analysis is further conducted to demonstrate the proximity of the contaminated data to the t-distribution.
- A customized robust detection methods designated as the missing-data t-distribution GLRT (mtGLRT) is developed for incomplete and contaminated data. Additionally, we propose two specific robust detection methods exclusively for scenarios with either incomplete data or contaminated data.
- The proposed mtGLRT detector involves maximum likelihood estimation for structured covariance matrices within the context of a complex multivariate t-distribution. To ef-

ficiently derive these estimations, we develop optimization algorithm that falls within the category of the generalized expectation maximization (GEM) algorithm.
- Numerical experiments on both synthetic and real recordings are conducted to compare the performance of our proposed cooperative spectrum sensing methods with several benchmark methods. Robustness of the proposed detectors is demonstrated, even in the face of high Bit Error Rate (BER).

This paper is organized as follows. We first provide the signal model and inaccurate transmission issue in Sec. II. In Sec. III, we propose the robust detectors and the corresponding MLE solutions for incomplete and contaminated data. The specific detectors for either incomplete or contaminated data are presented in Sec. IV. Numerical experiments are given in Sec. V. Finally, conclusions are summarized in Sec. VI.

II. Problem Formulation

A. Signal Model

Consider p distributed CR users, each equipped with a single antenna. These users cooperatively sense a frequency band of interest occupied by a primary user. The channels connecting the primary user and each of the CR users exhibit frequency non-selective (flat) fading, and the rank of the corresponding signal subspace is assumed to be r. We define two hypotheses: signal absence and presence events, denoted as $\mathcal{H}_0$ and $\mathcal{H}_1$ respectively. Let $x_j \in \mathbb{C}$ represent the received signal of the j-th $(j = 1, \ldots, p)$ receiver, and $\mathbf{x} = [x_1, \ldots, x_p]^T \in \mathbb{C}^p$. The conventional hypothesis testing problem is formulated as follows:

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{x} = \mathbf{n}, \\ \mathcal{H}_1 &: \mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}, \end{aligned} \tag{1}$$

In (1), $\mathbf{s} \in \mathbb{C}^r$ is the primary signal following an independent and identically distributed (i.i.d.) zero-mean circular complex Gaussian (CCG) distribution[1]. $\mathbf{H} \in \mathbb{C}^{p\times r}$ represents the unknown channel between the primary user and the receivers, and $\mathbf{n} \in \mathbb{C}^p$ is the i.i.d. zero-mean CCG and uncorrelated noise. For the current discussion, we disregard the differences in propagation delays from the primary user to the distributed CR users. The hypothesis testing model remains valid if considering different propagation delays where additional delay compensation procedure is needed.

Given that any spatial correlation and scaling of the primary signal can be incorporated into $\mathbf{H}$, we assume the covariance matrix of $\mathbf{s}$ to be the identity matrix, i.e., $\mathbb{E}\left(\mathbf{s}\mathbf{s}^H\right) = \mathbf{I}_r$ and denote the covariance matrix of the received signal $\mathbf{x}$ as $\mathbf{\Sigma}$. The hypothesis testing problem is thus written as:

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{\Sigma} \in \mathcal{S}_0, \\ \mathcal{H}_1 &: \mathbf{\Sigma} \in \mathcal{S}_1, \end{aligned} \tag{2}$$

[1]The CCG distribution characterizes complex-valued random variables with independently and identically Gaussian-distributed real and imaginary parts, and a uniformly distributed phase from 0 to 2π. This configuration ensures circular symmetry in the complex plane. The CCG distribution is extensively utilized in the fields of signal processing and communications [32].

where $\mathcal{S}_0$ and $\mathcal{S}_1$ are the feasible covariance matrix structure sets, i.e.,

$$\begin{aligned} \mathcal{S}_0 &= \{\mathrm{Diag}\,(\boldsymbol{\psi}) \mid \boldsymbol{\psi} \in [\boldsymbol{\psi}_{\mathrm{L}}, \boldsymbol{\psi}_{\mathrm{U}}]\} \\ \mathcal{S}_1 &= \left\{\mathbf{H}\mathbf{H}^H + \mathrm{Diag}\,(\boldsymbol{\psi}) \mid \mathbf{H} \in \mathbb{C}^{p\times r}, \boldsymbol{\psi} \in [\boldsymbol{\psi}_{\mathrm{L}}, \boldsymbol{\psi}_{\mathrm{U}}]\right\} \end{aligned} \tag{3}$$

with $\boldsymbol{\psi}_{\mathrm{L}}$ and $\boldsymbol{\psi}_{\mathrm{U}}$ being the predetermined lower and upper bounds of the noise power respectively ($\mathbf{0} \le \boldsymbol{\psi}_{\mathrm{L}} \le \boldsymbol{\psi}_{\mathrm{U}}$). Given N observations of the raw data, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]^T$, various methods have been developed for Problem (2) [30], [31], [33] and one of the most popular detector, i.e., GLRT [31], compares the following statistic with a predefined threshold γ to determine the existence of the primary signal, i.e.,

$$\xi_{\mathrm{GLR}} = \frac{f_{\mathcal{G}}\left(\mathbf{X} \mid \hat{\mathbf{\Sigma}}_1\right)}{f_{\mathcal{G}}\left(\mathbf{X} \mid \hat{\mathbf{\Sigma}}_0\right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{4}$$

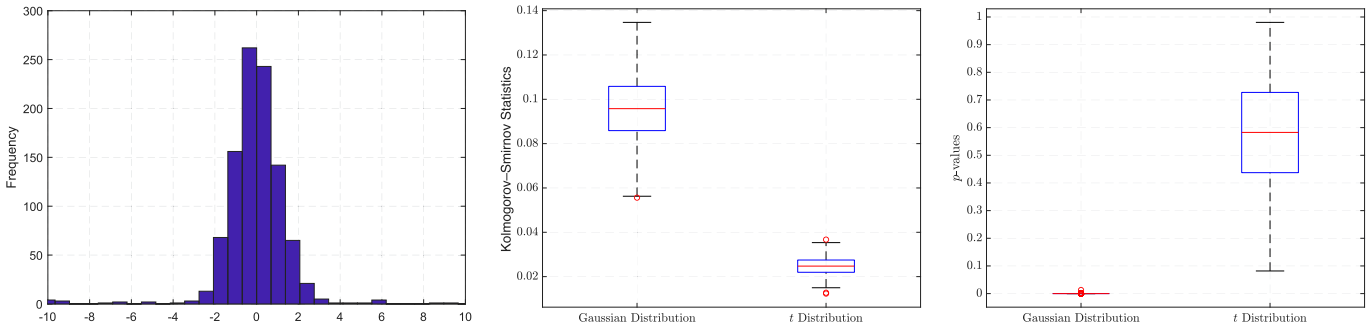
where $\hat{\mathbf{\Sigma}}_0$ and $\hat{\mathbf{\Sigma}}_1$ are the maximum likelihood estimate (MLE) solution from $\mathbf{X}$ under corresponding structures as presented in (2), $f_{\mathcal{G}}\left(\mathbf{X} \mid \mathbf{\Sigma}\right) = \prod_{i=1}^{N} \exp(-\mathbf{x}_i^H \mathbf{\Sigma}^{-1} \mathbf{x}_i) / \left((\pi)^p \det(\mathbf{\Sigma})\right)$ is the probability density function (pdf) of the complex multivariate Gaussian distribution.

B. Challenges of Incomplete and Contaminated Data

Though the effectiveness of the GLRT has been exhibited across various applications, it is applicable only under ideal scenarios where the raw data $\mathbf{X}$ is accurately gathered at the fusion center. However, as previously mentioned in Sec. I, we anticipate that the data at the fusion center will be both **incomplete and contaminated**. A primary concern is that the classical GLRT presumes Gaussian-distributed received data. Contamination can alter the data distribution, thereby impairing detection accuracy.

To illustrate the statistical distribution of the contaminated data, we simulate space electromagnetic signals at the receivers by generating 1000 random numbers following a standard normal distribution $\mathcal{N}(0, 1)$. These signals are sampled using a 24-bit bipolar analog-to-digital converter (ADC) with a ± 10 volts reference voltage. The sampled data are then transmitted to the fusion center through a noisy wireless channel with a BER of 1%. The distribution of the received data, as displayed in Fig. 2(a), exhibits notable heavy tails. The phenomenon can be explained by recognizing that the thin tails of the Gaussian distribution render the inclusion of extreme samples (those with large magnitude values) highly unlikely. However, the magnitude of the samples is dominated by a portion of the bits, such as the top 4 bits. The probability of flipping at least one of these 4 bits is approximately $1 - 0.99^4 \approx 0.04$. Despite the low probability, bit flips can significantly alter $\mathbf{x}$'s values, leading to a distribution with heavier tails. It is acknowledged that Gaussian-based estimates are sensitive to outliers, suggesting that the traditional GLRT's performance may decline with contaminated data.

To identify an appropriate distribution for the contaminated data, we examine the t-distribution, known for effectively capturing heavy-tailed characteristics. The Gaussian distribution is



(a) Data after noisy transmission. (b) K-S statistics. (c) p-values.

Fig. 2. Histogram of the empirical data distribution post noisy transmission (BER = 1%) with K-S test results.

also included as a reference distribution for comparison. We fit the contaminated data to both distributions and then utilized the Kolmogorov–Smirnov (K–S) test to assess the goodness of fit [34]. Using MATLAB's kstest function, we computed the K–S statistics and p-values[2]. After 1000 iterations, the results shown in Fig. 2(b) and 2(c) indicate that the t-distribution more accurately reflects the contaminated data than the Gaussian distribution. The p-values for the t-distribution surpass the conventional standard 5% significance level, endorsing it as a fitting model for the contaminated data. For comparison, the p-values under the Gaussian assumption do not meet this threshold, confirming a poor fit. This can be explained by the fact that t-distributions, which exhibit heavier tails at lower degrees of freedom, are more suitable for modeling heavy-tailed data than the Gaussian distribution.

III. The mtGLRT Detector For Incomplete and Contaminated Data

In this section, we delve into the development of a robust detector specifically designed to handle incomplete and contaminated data. As shown in Sec. II-B, contaminated data often follows a heavy-tailed distribution that aligns closely with the t-distribution profile. Furthermore, the t-distribution has been successfully employed in various applications such as finance [11], radar tracking [35], and multi-sensor fusion [36]. The extensive body of theoretical work on parameter estimation for the t-distribution is expected to simplify the development of associated algorithms [37], while the broader heavy-tailed distributions are likely to entail greater complexity [38], [39]. Therefore, considering these practical considerations, we propose to model the contaminated data with the multivariate t-distribution [37]. Given its base in t-distribution, the ability of directly handing missing data, and its function in the GLRT, we propose to designate it as the missing-data t-distribution (mtGLRT) detector.

[2]The K-S statistic measures the discrepancy between the empirical data distribution and a reference distribution. A small p-value indicates a significant divergence from the reference distribution, whereas a large p-value implies no substantial difference, suggesting equivalence between the two distributions.

A. Preliminaries: The Complex Multivariate t-Distribution

The heavy-tailed distribution known as the multivariate t-distribution is frequently employed in various applications. The p-dimensional complex multivariate t-distribution encompasses a pdf as

$$f_t\left(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu\right) = \frac{\Gamma\left(\nu+p\right)}{(\nu\pi)^p\Gamma\left(\nu\right)\left|\boldsymbol{\Sigma}\right|}\left[1+\frac{1}{\nu}\left(\mathbf{x}-\boldsymbol{\mu}\right)^H\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right)\right]^{-(\nu+p)}, \tag{5}$$

where $\mathbf{x}\in\mathbb{C}^p$ denotes the observation vector, $\nu\geq 1$ is the degrees of freedom, $\boldsymbol{\Sigma}$ is the $p\times p$ positive definite scatter matrix, $\boldsymbol{\mu}$ is the p-dimensional mean vector and $\Gamma\left(a\right)=\int_0^\infty t^{(a-1)}\exp\left(-t\right)dt$ is the gamma function [37]. The smaller ν is, the heavier the tail is. Notably, the complex multivariate Gaussian distribution is a special case of multivariate t-distribution with $\nu\to\infty$. Interestingly, the above multivariate t-distribution can be represented in a hierarchical structure as

$$\begin{aligned}\mathbf{x}|\tau &\overset{i.i.d}{\sim}\mathcal{CN}\left(\boldsymbol{\mu},\frac{1}{\tau}\boldsymbol{\Sigma}\right),\\ \tau &\overset{i.i.d}{\sim}\text{Gamma}\left(\frac{\nu}{2},\frac{\nu}{2}\right),\end{aligned} \tag{6}$$

where $\mathcal{CN}\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right)$ is the complex multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Gamma (a,b) denotes gamma distribution of shape a and rate b with pdf

$$f_{\text{GM}}\left(\tau\right)=b^a\tau^{(a-1)}\frac{\exp\left(-b\tau\right)}{\Gamma\left(a\right)}. \tag{7}$$

The hierarchical structure will play an important role in our analysis and algorithm design, which is discussed in the following.

B. The mtGLRT Detector

We propose to use the complex multivariate t-distribution to model the data. The decision rule therefore can be formulated as follows:

$$\xi_{mt\text{GLR}}=\frac{f_t\left(\mathbf{X}_{\text{o}}\,|\,\hat{\boldsymbol{\Sigma}}_1\right)}{f_t\left(\mathbf{X}_{\text{o}}\,|\,\hat{\boldsymbol{\Sigma}}_0\right)}\underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}}\gamma, \tag{8}$$

where $f_t(\mathbf{X}_\mathsf{o} \mid \mathbf{\Sigma})$ is the pdf of observed part $\mathbf{X}_\mathsf{o}$ given the covariance matrix[3] $\mathbf{\Sigma}$, i.e.,

$$f_t(\mathbf{X}_\mathsf{o} \mid \mathbf{\Sigma}) = \prod_{i=1}^{N} \frac{\Gamma(\nu+p_i)(1+\frac{1}{\nu}\mathbf{x}_{i,\mathsf{o}}^H \mathbf{\Sigma}_{i,\mathsf{oo}}^{-1}\mathbf{x}_{i,\mathsf{o}})^{-(\nu+p_i)}}{(\pi\nu)^{p_i}\Gamma(\nu)\det(\mathbf{\Sigma}_{i,\mathsf{oo}})}, \tag{9}$$

here $\mathbf{x}_{i,\mathsf{o}} \in \mathbb{C}^{p_i}$ is a vector collecting (p_i is the number of observed elements in $\mathbf{x}_i$) observed data elements in $\mathbf{x}_i$, $\mathbf{\Sigma}_{i,\mathsf{oo}} \in \mathbb{S}_+^{p_i}$ ($\mathbb{S}_+^{p_i}$ is the set of all $p_i \times p_i$ symmetric positive semidefinite matrices) is the sub-matrix of $\mathbf{\Sigma}$ representing the covariance matrix of $\mathbf{x}_{i,\mathsf{o}}$, $\hat{\mathbf{\Sigma}}_0$ and $\hat{\mathbf{\Sigma}}_1$ are obtained from observed data $\mathbf{X}_\mathsf{o}$ by solving the following MLE problems:

$$\hat{\mathbf{\Sigma}}_0 = \arg\max_{\mathbf{\Sigma}\in\mathcal{S}_0} \quad f_t(\mathbf{X}_\mathsf{o} \mid \mathbf{\Sigma}), \tag{10}$$

$$\hat{\mathbf{\Sigma}}_1 = \arg\max_{\mathbf{\Sigma}\in\mathcal{S}_1} \quad f_t(\mathbf{X}_\mathsf{o} \mid \mathbf{\Sigma}). \tag{11}$$

The estimation of complex-valued structured covariance matrices confronts significant challenges, particularly when assuming the data to follow a t-distribution amidst instances of missing data, as detailed in Problems (10) and (11). In the work of [40], a comparable constraint to that in Problem (11) was examined, with the solution to the covariance matrix estimation under a real-valued multivariate Gaussian distribution. Subsequent research by [12] delved into the complex-valued covariance matrix estimation problems with different structures under the assumption of multivariate Gaussian distribution. Recently, the authors in [38] explored the estimation of real-valued parameters, including the covariance matrix, from incomplete data within the framework of a mixture of elliptical distributions, a category which encompasses the t-distribution as a special instance. These algorithms are all iterative methods, which address the original problem by solving a sequence of surrogate problems.

However, the methodologies presented in previous studies do not directly translate to solving Problems (10) and (11) that we aim to address. These methods are specifically designed for scenarios that assume either a Gaussian distribution or an unstructured covariance matrix. In the subsequent subsection, we will detail the specialized algorithms developed to tackle these specific Problems.

C. Solving Problems (10) and (11)

In this study, we introduce an optimization algorithm constructed upon the GEM method to solve the aforementioned problems. The introduction of GEM method is also included in Apx. A. More specifically, this method transforms the task of exactly solving the surrogate problem into improving the objective beyond the previous step. We will begin by addressing Problem (11) as it encompasses Problem (10), which is a specific instance of the former. The procedural details are elaborated in the following.

[3]Technically, in the context of the complex t-distribution, the parameter $\mathbf{\Sigma}$ is named as the scatter matrix, while the corresponding covariance matrix is $\frac{\nu}{\nu-2}\mathbf{\Sigma}$. However, for the sake of simplicity in this discussion, we will also refer to it as the covariance matrix in this paper.

Using the hierarchical model (6) of the complex multivariate t-distribution, we have

$$\begin{aligned}(\mathbf{x}_{i,\mathsf{o}}, \mathbf{x}_{i,\mathsf{m}}) \mid \tau_i &\overset{i.i.d}{\sim} \mathcal{CN}\left(\mathbf{0}, \frac{1}{\tau_i}\mathbf{\Sigma}\right),\\ \tau_i &\overset{i.i.d}{\sim} \text{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right),\end{aligned} \tag{12}$$

where $\mathbf{x}_{i,\mathsf{o}}$ and $\mathbf{x}_{i,\mathsf{m}}$ represent the observed and missing components of $\mathbf{x}_i$, $\boldsymbol{\tau} = \{\tau_i\}_{i=1}^N$ collects the introduced latent variables, and ν is the degrees of freedom decided by the user. The complete data log-likelihood of $(\mathbf{X}_\mathsf{o}, \mathbf{X}_\mathsf{m}, \boldsymbol{\tau})$ is simply

$$\ell(\mathbf{X}_\mathsf{o}, \mathbf{X}_\mathsf{m}, \boldsymbol{\tau} \mid \mathbf{\Sigma}) = \sum_{i=1}^{N} \log\left[\frac{\exp(-\mathbf{x}_i^H(\frac{\mathbf{\Sigma}}{\tau_i})^{-1}\mathbf{x}_i)}{(\pi)^p \det(\frac{\mathbf{\Sigma}}{\tau_i})} f_{\text{GM}}(\tau_i)\right], \tag{13}$$

where $\mathbf{X}_\mathsf{m} = \{\mathbf{x}_{i,\mathsf{m}}\}_{i=1}^N$ collects all the missing data, $f_{\text{GM}}(\tau_i)$ is the pdf of τ_i given ν, and is not related to the unknown parameter $\mathbf{\Sigma}$. It should be noted that the equation for the complete data log-likelihood, (13), is expressed differently as both (10) and (11).

a) Expectation (E) step: The task is obtaining the following expression (ignoring the constant items irrelevant to unknown parameter $\mathbf{\Sigma}$):

$$\begin{aligned}&Q\left(\mathbf{\Sigma} \mid \mathbf{\Sigma}^k\right)\\ &= \mathbb{E}_{\mathbf{X}_\mathsf{m},\boldsymbol{\tau}\mid\mathbf{X}_\mathsf{o},\mathbf{\Sigma}^k}\left[\ell(\mathbf{X}_\mathsf{o}, \mathbf{X}_\mathsf{m}, \boldsymbol{\tau} \mid \mathbf{\Sigma})\right]\\ &= T\log\det(\mathbf{\Sigma}^{-1}) - \text{Tr}\Bigg(\mathbf{\Sigma}^{-1}\underbrace{\sum_{i=1}^{N}\mathbb{E}_{\mathbf{X}_\mathsf{m},\boldsymbol{\tau}\mid\mathbf{X}_\mathsf{o},\mathbf{\Sigma}^k}\left[\tau_i\mathbf{x}_i\mathbf{x}_i^H\right]}_{\triangleq\mathbf{A}^{k+1}}\Bigg),\end{aligned} \tag{14}$$

where $\mathbf{A}^{k+1}$ has an analytical form as delineated in Lemma 1. Note that here $\mathbf{A}^{k+1} \succeq \mathbf{0}$ is guaranteed in principle.

Lemma 1: The analytical form of $\mathbf{A}^{k+1}$ can be expressed as

$$\mathbb{E}_{\mathbf{X}_\mathsf{m},\boldsymbol{\tau}\mid\mathbf{X}_\mathsf{o},\mathbf{\Sigma}^k}\left[\tau_i\mathbf{x}_i\mathbf{x}_i^H\right] = \bar{\tau}_i^k\bar{\mathbf{x}}_i^k(\bar{\mathbf{x}}_i^k)^H + \mathbf{\Phi}_i^k, \tag{15}$$

where $\bar{\tau}_i^k = \mathbb{E}_{\mathbf{X}_\mathsf{m},\boldsymbol{\tau}\mid\mathbf{X}_\mathsf{o},\mathbf{\Sigma}^k}[\tau_i] = \frac{\nu+2p_i}{\nu+2\mathbf{x}_{i,\mathsf{o}}^H(\mathbf{\Sigma}_{i,\mathsf{oo}}^k)^{-1}\mathbf{x}_{i,\mathsf{o}}}$, $\bar{\mathbf{x}}_i^k = \mathbb{E}_{\mathbf{X}_\mathsf{m},\boldsymbol{\tau}\mid\mathbf{X}_\mathsf{o},\mathbf{\Sigma}^k}[\mathbf{x}_i]$ whose missing elements will be found using $\mathbb{E}_{\mathbf{X}_\mathsf{m},\boldsymbol{\tau}\mid\mathbf{X}_\mathsf{o},\mathbf{\Sigma}^k}[\mathbf{x}_{i,\mathsf{m}}] = \mathbf{\Sigma}_{i,\mathsf{mo}}^k(\mathbf{\Sigma}_{i,\mathsf{oo}}^k)^{-1}\mathbf{x}_{i,\mathsf{o}}$, and the (l,j)-th element of $\mathbf{\Phi}_i^k \in \mathbb{C}^{p\times p}$ is zero if either $x_{l,i}$ or $x_{j,i}$ is observed, otherwise is the corresponding element of $\mathbf{\Sigma}_{i,\mathsf{mm}}^k - \mathbf{\Sigma}_{i,\mathsf{mo}}^k(\mathbf{\Sigma}_{i,\mathsf{oo}}^k)^{-1}\mathbf{\Sigma}_{i,\mathsf{om}}^k$.

Proof: See Apx. B. ■

b) Maximization (M) step: Given the Q function as (14), the surrogate problem is simply written as

$$\begin{aligned}\max_{\mathbf{H},\boldsymbol{\psi}} \quad & T\log\det(\mathbf{\Sigma}^{-1}) - \text{Tr}\left(\mathbf{\Sigma}^{-1}\mathbf{A}^{k+1}\right)\\ \text{s.t.} \quad & \mathbf{\Sigma} = \mathbf{H}\mathbf{H}^H + \text{Diag}(\boldsymbol{\psi}),\\ & \mathbf{H} \in \mathbb{C}^{p\times r}, \boldsymbol{\psi} \in [\boldsymbol{\psi}_\text{L}, \boldsymbol{\psi}_\text{U}].\end{aligned} \tag{16}$$

This problem allows for a formulation similar to the one considered in [40, Problem (2)] and can be addressed by extending [40, Algorithm 1] to our specific case. The details are documented in Apx. C and the customized algorithm is presented as Alg. 2.

Algorithm 1: A GEM-based algorithm for Problem (10) and Problem (11).

Initialize $\mathbf{H}^0$, $\mathbf{\Psi}^0$ and $\mathbf{\Sigma}^0 = \mathbf{H}^0(\mathbf{H}^0)^H + \mathbf{\Psi}^0$.
for $k = 0, 1, 2, \ldots$ **do**
  E-step: compute $\mathbf{A}^{k+1}$ using (14);
  M-step:
  [Problem (10):] update $\mathbf{\Sigma}_0^{k+1} = \text{Diag}(\mathbf{A}^{k+1}/T)$;
  [Problem (11):] update $\mathbf{\Sigma}_1^{k+1}$ by executing Alg. 2 for a single round ;
  $k \leftarrow k + 1$;
  Terminate when converges;
end for
Return $\mathbf{\Sigma}_0^k$ or $\mathbf{\Sigma}_1^k$.

To tackle Problem (10), we encounter a surrogate problem that shares the same objective function but possesses a different constraint, namely $\mathbf{\Sigma} \in \mathcal{S}_0$. Obtaining the optimal solution to this surrogate problem becomes a straightforward task, as

$$\mathbf{\Sigma}^\star = \text{Diag}(\mathbf{A}^{k+1}/T). \tag{17}$$

c) The Overall Algorithm: In the EM framework, the algorithm iteratively performs the E step and the M step until convergence is reached. During each M step, the iterative Alg. 2, which falls under the majorization-minimization (MM) framework, is traditionally run to convergence, which can be time-intensive. However, as Alg. 2 ensures monotonic improvement of the objective function with each iteration, it is sufficient to execute it for a single iteration within each M step. This modified approach, akin to the GEM method, has the potential to speed up convergence compared to the classical EM algorithm [41], [42]. The comprehensive algorithm for addressing Problem (10) and (11) is detailed in Alg. 1.

d) The Robustness and Convergence: In this subsection, we discuss the robustness of the proposed covariance matrix estimators in (10) and (11) as well as the convergence behavior of Alg. 1.

Since our estimators are the maximum likelihood estimation for data following a t-distribution, their robustness against contamination can be characterized by the so-called influence function [43], which quantifies the estimator's sensitivity to minor disruptions or outliers. A larger value of the influence function indicates higher sensitivity. As delineated in [44, Theorem 8], using a complex multivariate t-distribution to model the data limits the impact of outliers and thus theoretically leads to a bounded influence function. From this perspective, we can assert that our estimators are robust in the face of contaminated data. Regarding the developed Alg. 1, the sequence $\{\mathbf{\Sigma}^k\}$ generated by it converges to a stationary point of Problem (10) or Problem (11). This convergence property is due to the fact that Alg. 1 falls into the generalized expectation-maximization (GEM) framework, whose convergence to stationary points has been well established in [45, Theorem 1].

IV. Specific Detector Design for Incomplete or Contaminated Data

Indeed, the developed mtGLRT detector in Sec. III is designed for scenarios featuring both incomplete and contaminated data. In real-world applications, these problematic data scenarios might not occur conjointly. Considering the real-time requirements in spectrum sensing applications, it becomes compelling to explore specific detectors for either incomplete or contaminated data. In this section, we delve into the development of specialized detectors, tailored for managing either incomplete or contaminated data. They are named as the missing-data GLRT (mGLRT) and t-distribution GLRT (tGLRT), respectively. It is important to note that the mtGLRT detector, simplifies to mGLRT when the degrees of freedom parameter ν tends towards infinity, and to tGLRT in the absence of missing data.

A. The mGLRT Detector for Incomplete Data

Note that if the data are uncontaminated, they can be effectively modeled using the Gaussian distribution, which represents a special case of the t-distribution as $\nu \to \infty$. Building on the proposed mtGLRT method, the missing-data GLRT, now termed mGLRT, can be reformulated as follows:

$$\xi_{\text{mGLR}} = \frac{f_{\mathcal{G}}\left(\mathbf{X}_{\text{o}} \,|\, \hat{\mathbf{\Sigma}}_1\right)}{f_{\mathcal{G}}\left(\mathbf{X}_{\text{o}} \,|\, \hat{\mathbf{\Sigma}}_0\right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{18}$$

where $p\left(\mathbf{X}_{\text{o}} \,|\, \mathbf{\Sigma}\right)$ denotes the Gaussian pdf of the observed data $\mathbf{X}_{\text{o}}$ for a given covariance matrix $\mathbf{\Sigma}$, specifically:

$$f_{\mathcal{G}}\left(\mathbf{X}_{\text{o}} \,|\, \mathbf{\Sigma}\right) = \prod_{i=1}^{N} \frac{\exp(-\mathbf{x}_{i,\text{o}}^H \mathbf{\Sigma}_{i,\text{oo}}^{-1} \mathbf{x}_{i,\text{o}})}{(\pi)^{p_i} \det(\mathbf{\Sigma}_{i,\text{oo}})}, \tag{19}$$

In this context, $\hat{\mathbf{\Sigma}}_0$ and $\hat{\mathbf{\Sigma}}_1$ represent the MLE results derived from $\mathbf{X}_{\text{o}}$ under the assumption of a complex multivariate Gaussian model. By letting $\nu \to \infty$ in Sec. III-C, we can obtain $\hat{\mathbf{\Sigma}}_0$ and $\hat{\mathbf{\Sigma}}_1$ via Alg. 2. This process includes an adjustment where $\bar{\tau}_i^k = 1$ for all i, k during the calculation of $\mathbf{A}^k$.

B. The tGLRT Detector For Contaminated Data

Building on the proposed mtGLRT method, the tGLRT for handling contaminated data can be formally expressed as:

$$\xi_{t\text{GLR}} = \frac{f_t\left(\mathbf{X} \,|\, \hat{\mathbf{\Sigma}}_1\right)}{f_t\left(\mathbf{X} \,|\, \hat{\mathbf{\Sigma}}_0\right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{20}$$

where $\hat{\mathbf{\Sigma}}_0$ and $\hat{\mathbf{\Sigma}}_1$ represent the MLE results derived from the complete data $\mathbf{X}$ under the assumption of a complex multivariate t-distribution. In scenarios without missing data, the estimation of $\hat{\mathbf{\Sigma}}_0$ and $\hat{\mathbf{\Sigma}}_1$ can be efficiently accomplished using Alg. 2. Under such conditions, the computation of $\mathbf{A}^k$ becomes significantly easier.

V. Numerical Experiments

In this section, we will conduct both synthetic data experiments as well as real data experiments to illustrate the performance of our proposed robust cooperative spectrum sensing techniques.

A. Synthetic Data Experiments

In the signal generation phase, an environment encompassing 8 distributed CR users, denoted as $p = 8$, is examined. Each of these users is furnished with an individual omnidirectional antenna. The primary objective here is to identify a sole primary source with a single antenna, specified as $r = 1$. The primary source in question is identified as transmitting Quadrature Phase Shift Keying (QPSK) modulated signals, with the baud rate set precisely to 20 kHz. Each receiver operates at a sampling rate of 100 kHz. The channel between each receiver and the signal source is hypothesized to be an independent Rician fading channel with K-factor being 4. These modulated signals are generated utilizing the MATLAB Communications Toolbox, as referenced in [46]. It is pertinent to note that the noise power of the CR user, denoted as $\hat{\sigma}_j$ $(j = 1 \ldots, p)$, is equal to β, and β expressed in dB, maintains a uniform distribution within an interval of $[-1, 1]$ dB. This distribution is outlined in the study by [29], which examines scenarios characterized by small noise power variations. The noise power is assume to be unknown, i.e., $\psi_{\mathrm{L}} = 0$ and $\psi_{\mathrm{U}} = \infty$. The signal-to-noise ratio (SNR)[4] for the received signal remains constant for each antenna. In the data transmission phase, the process of analog-to-digital conversion (ADC) is simulated to quantify the received data into a format of a 24-bit signed integer. Subsequently, the data is subjected to transmission errors, which are introduced in accordance with a pre-established BER. Each simulation experiment is executed using 500 consecutive samples, equating to a duration of approximately 5 milliseconds. The threshold γ is calculated using the empirical test statistics from 10000 realizations of pure noise data. The probability of detection is determined by repeating the detection procedure 10000 times.

1) Selection of Degrees of Freedom ν: It is essential to specify the degrees of freedom, ν, for the tGLRT and mtGLRT methods. As ν increases towards infinity, the tGLRT method converges to the classical GLRT method theoretically. Conversely, a small ν enables a more robust estimation of the covariance matrix in the presence of contaminated data. However, the impact of ν on detection performance remains uncertain. To investigate this, we performed a numerical experiment evaluating the tGLRT method with varying ν values.

In Fig. 3, we assess the tGLRT method's performance with several ν choices. It is noteworthy that the classical GLRT method is effectively a variant of the tGLRT method in the limit as ν approaches infinity. For error-free transmission, the probability of detection increases slightly with larger ν values.

[4]In this context, the SNR quantifies the quality of the primary signal received by the CR users, whereas the BER evaluates the quality of the wireless data link between the CR users and the fusion center. Consequently, the SNR and BER in the numerical experiments are indirectly related.

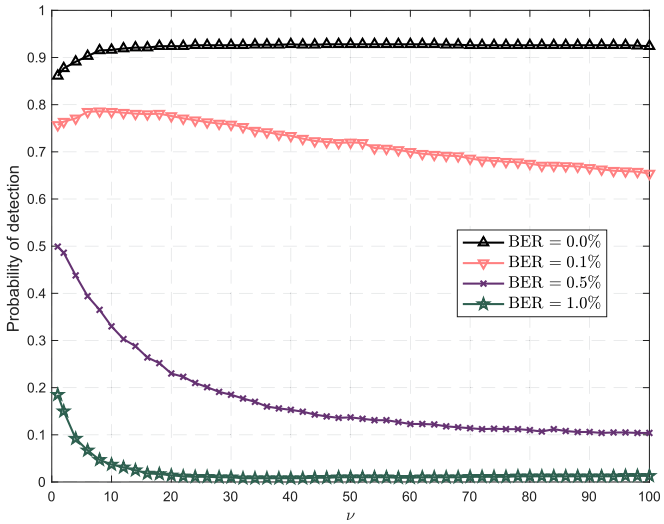


Fig. 3. Performance of tGLRT method under different BER conditions versus ν: $P_{FA} = 1\%$, SNR $= -14$dB.

This slight increase can be ascribed to the fact that data transmitted perfectly approximates a Gaussian distribution, which is a special case of the t-distribution as $\nu \to \infty$. When BER = 0.1%, performance initially improves with higher ν values but then diminishes. Conversely, at higher BER levels, detection performance significantly enhances with smaller ν values. A more detailed empirical explanation of the robustness of the proposed tGLRT method is provided in Apx. D. Therefore, if the BER level is precisely known, the optimal ν can be determined through empirical analysis. In the absence of prior knowledge of the BER level, a conservative choice of ν, such as $\nu = 1$, is recommended for further numerical experiments.

2) Detection Performance: In this section, we illustrate the robust performance of our proposed cooperative detectors by feeding them with imperfectly transmitted data. For comparison, we will report the detecting performance of the following methods:

1) **Classical GLRT with Clean Data:** We utilize the classical GLRT detector as outlined in (4), provided with complete and accurate data. This scenario serves as a benchmark for the optimal performance achievable by GLRT-based methods in our experiments.
2) **Classical GLRT with Imputation/Deletion:** Here, the classical GLRT detector (4) is applied to datasets that have undergone missing data handling through either imputation or deletion.
3) **M-estimator with Data Imputation/Deletion:** This robust detection approach employs an M-estimator, as detailed in [27], coupled with datasets treated for missing values via imputation or deletion.
4) **tGLRT with Data Imputation/Deletion:** Our proposed tGLRT detector, as per (20), is tested on datasets processed to handle missing values by imputation or deletion methods.
5) **mGLRT with LOF:** Our proposed mGLRT detector, i.e., (18), is applied to datasets from which

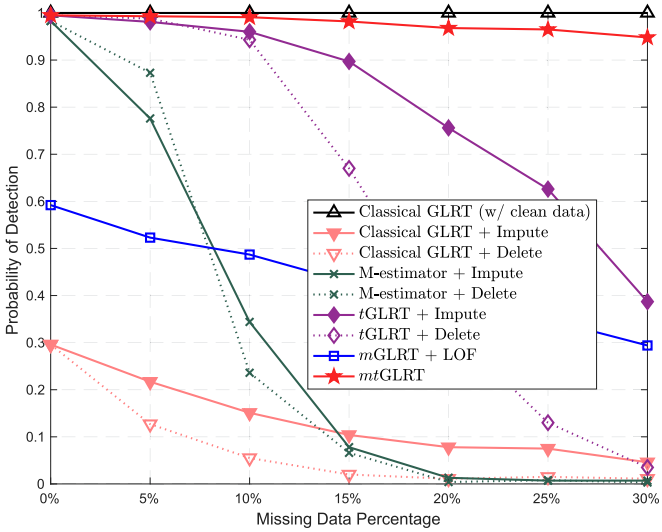


Fig. 4. P_D versus missing data percentage by several detection methods: BER $= 1\%$, $P_{FA} = 1\%$, SNR $= -10$dB.

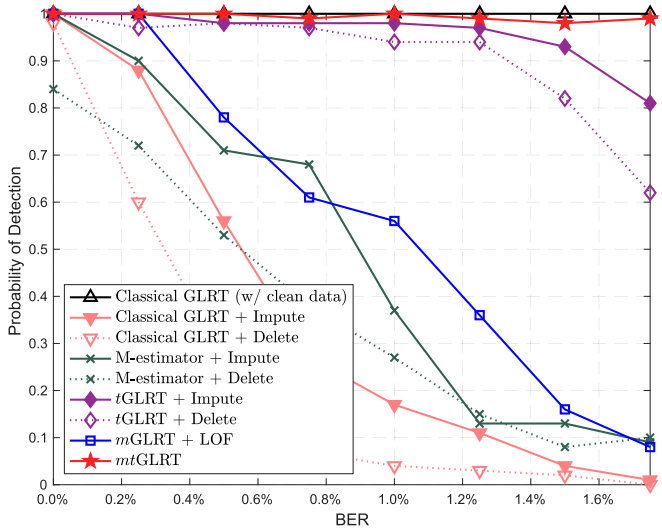


Fig. 5. P_D versus BER by several detection methods: 10% missing data, $P_{FA} = 1\%$, SNR $= -10$dB.

potential outliers have been removed using the LOF method [47].

6) mt**GLRT:** Our proposed mtGLRT detector (8), which is directly applied to the original dataset, inclusive of incomplete and contaminated data.

It is important to note that data imputation was performed using stochastic imputation based on a Gaussian model, as elaborated in [16]. Specifically, the variance for each random variable was first estimated as the sample covariance derived from its observed components. Subsequently, random samples were drawn from the Gaussian distribution in accordance with the estimated covariance. The LOF method was applied using MATLAB's lof function, and, based on empirical observation, the top 1% of data points with the highest anomaly scores were removed, following the approach outlined by [47]. In comparison, our proposed mtGLRT method is distinguished by its inherent capability to directly process both incomplete and contaminated data, unlike other methods that rely on basic techniques to manage missing data or address contaminations. The degrees of freedom ν is set to be 1 according to the previous discussion.

Fig. 4 compares detection performance across various methods as a function of missing data percentages, with the initial data generated at a BER of 1%. The analysis yields several interesting insights. With the exception of the ideal classical GLRT with clean data, all detectors experience decreased performance as the missing data percentage increases. Notably, methods designed to directly manage incomplete data, such as mGLRT with LOF and our proposed mtGLRT, exhibit reduced sensitivity to higher rates of missing data. In contrast, the classical GLRT, which relies on imputation or deletion to handle missing data, demonstrates significantly weaker performance. The Gaussian-based mGLRT method, after outlier removal via the LOF approach, shows marked improvement. The M-estimator based detector maintains commendable performance at lower levels of missing data due to its innate robustness against data contamination. However, its performance deteriorates rapidly as the missing data percentage grows, reflecting a limited capacity to handle incomplete datasets. Our proposed mtGLRT detector consistently outperforms the competition, maintaining the highest level of detection effectiveness. This is attributed to the detector's design, which is intrinsically robust to contamination and adept at managing incomplete data. Fig. 5 assesses detection performance versus BER, with datasets featuring 10% missing values. All methods register a drop in performance with increasing BER, yet detectors utilizing the t-distribution exhibit less sensitivity to BER changes. The mtGLRT detector remains superior, achieving the utmost detection accuracy among all methods, with the exception of the ideal GLRT method applied to clean data.

Fig. 6 illustrates the receiver operating characteristic (ROC) curve, which represents the trade-off between the probability of detection (P_D) and the probability of false alarm (P_{FA}), for an experiment with 10% missing data and a 1% BER. Our proposed mtGLRT method stands out, delivering detection performance nearly equivalent to the classical GLRT method with clean data. The tGLRT method, despite using imputation or deletion for missing data, performs marginally worse than the mtGLRT due to its limited capacity to address incomplete data. The M-estimator and mGLRT methods exhibit moderate performance, surpassing the classical GLRT that employs imputation and deletion, thanks to their improved capability in dealing with contaminations and missing data. The synthetic data experiment demonstrates that although basic techniques exist to manage missing data and mitigate contamination, the performance disparity between these techniques and our mtGLRT method is substantial.

B. Real Data Experiments on Software-Defined Radio Testbed

In this section, we validate our proposed robust detector using real data collected from a software-defined radio testbed.

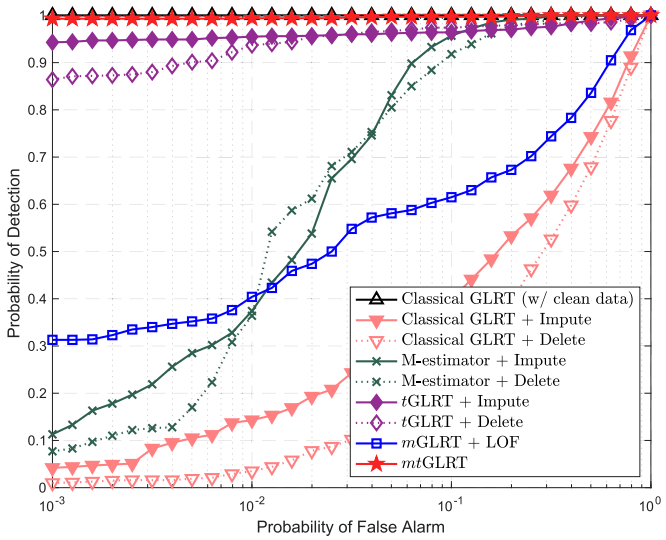


Fig. 6. P_D versus P_{FA} by several detection methods: 20% missing data, BER $= 1\%$, SNR $= -10$dB.

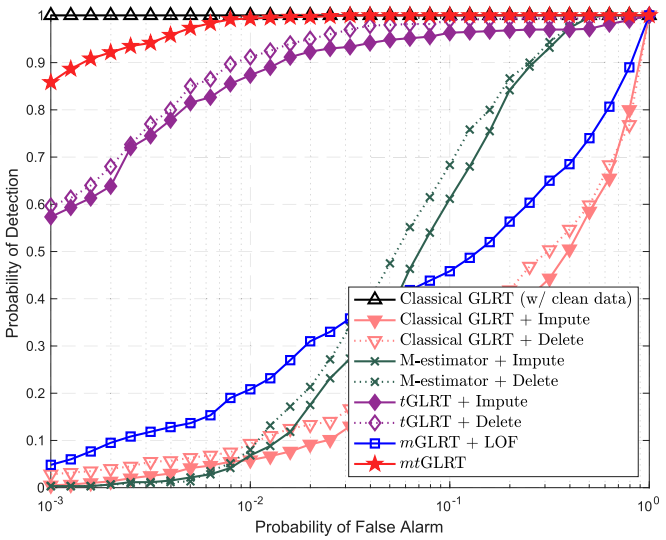


Fig. 8. Probability of detection on software-defined radio testbed (missing percentage: 10%, BER: 1%) by several detection methods.

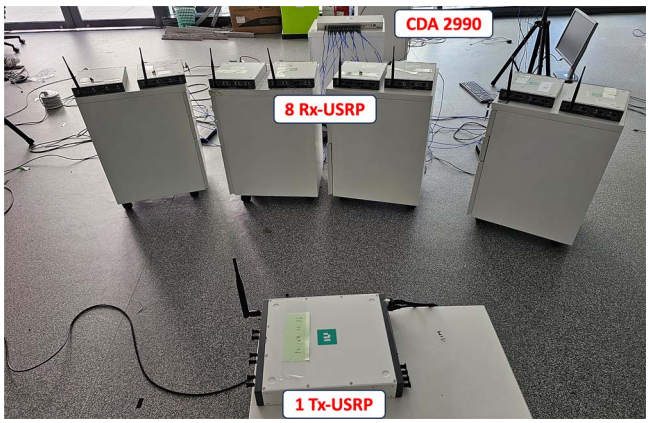


Fig. 7. Experiment platform setup.

Our laboratory experiment employs an NI Universal Software Radio Peripheral (USRP) X410 as the transmitter and eight NI USRP X410s as receivers, each equipped with a single antenna. Conducted in an indoor environment without GPS clock availability, we synchronize the receivers using an NI Clock Distribution Accessory (CDA) 2990 device. Fig. 7 illustrates the experimental setup. The transmitter broadcasts at a carrier frequency of 1.4GHz, using randomly-generated QPSK signals at a precise baud rate of 20 kHz. The transmitter antenna is connected to a 30dB Attenuator to simulate a weak signal scenario. Operating at a sampling rate of 240 kHz and receiving gain of 60 dB, the receivers feature an ADC configuration of 10 bits. The collected data is transmitted to a personal computer via a wired connection. We manually flip the bits with a 1% probability to simulate a 1% BER and randomly remove 10% of the samples. Each detecting experiment is conducted using 1000 consecutive samples. Similar to previous synthetic data experiment, the threshold γ for each method is calculated using 1000 realizations of recorded noise data, with the transmitting gain set to 0 dB.

Fig. 8 depicts the detection probability of various methods against the probability of false alarm. The observed trends and relative performance rankings are similar to those from experiments using synthetic data. The proposed mtGLRT method exhibits superior performance with incomplete and contaminated data. Although the performance of the tGLRT method, when integrated with imputation and deletion techniques for missing data, is marginally lower, it still significantly outperforms the alternative approaches in terms of detection. This suggests that the robustness of our proposed mtGLRT detector in this experiment may primarily stem from the employment of the t-distribution. The real data experiment corroborates the efficacy of the proposed mtGLRT detector in practical scenarios.

VI. Conclusion

In conclusion, this paper has presented robust methodologies for cooperative spectrum sensing designed to handle incomplete and contaminated data. The complex t-distribution is utilized to model data and accommodate missing data elements. Additionally, we have developed the generalized expectation-maximization based algorithms to address the associated maximum likelihood estimation challenges. Our numerical experiments have underscored the robustness of these proposed techniques in the presence of a high bit error rate. This work substantially enhances spectrum sensing techniques and provides more robust solutions capable of overcoming data transmission imperfections.

Appendix

A. The Generalized Expectation-Maximization Algorithm

The generalized expectation-maximization algorithm is a powerful iterative method that addresses challenging MLE

problems involving missing data or latent variables [48]. By incorporating the latent data $\mathbf{Z}$, the EM algorithm enables the conversion of the maximization of the observed data's log-likelihood, $\ell(\mathbf{X}\,|\,\boldsymbol{\theta})$, into the maximization of a sequence of simpler and more tractable problems.

In each iteration, the expectation maximization (EM) algorithm computes $Q(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^k)$, which represents the expected log-likelihood function of the complete data log-likelihood, $\ell(\mathbf{X},\mathbf{Z}\,|\,\boldsymbol{\theta})$, with respect to the current conditional distribution of $\mathbf{Z}$ given the observed data $\mathbf{X}$ and the current estimate of the parameter $\boldsymbol{\theta}^k$. Then, the algorithm finds the updated parameter estimate $\boldsymbol{\theta}^{k+1}$ by maximizing $Q(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^k)$. The EM algorithm can be described as iteratively conducting following steps until convergence:

- **Expectation (E) Step**: calculate the expected value of the complete-data log-likelihood, with respect to the conditional distribution of the latent variable given the observed data under the current parameter estimates. This expectation is denoted as $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^k)$, and can be written as:

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^k) = E_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^k}[\ell(\boldsymbol{\theta};\mathbf{X},\mathbf{Z})]. \tag{21}$$

- **Maximization (M) Step**: find the parameter that maximizes this quantity:

$$\boldsymbol{\theta}^{k+1} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^k) \tag{22}$$

The convergence is generally determined by the change in either the log-likelihood or the parameter estimates becoming smaller than a preset threshold. By iterating through the expectation (E) and maximization (M) steps, the EM algorithm effectively simplifies the optimization process and provides a means to handle complex MLE problems involving missing data or latent variables.

In lieu of determining a minimizer of the function $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^k)$, it is proposed that we ascertain a point $\boldsymbol{\theta}^{k+1}$ such that $Q(\boldsymbol{\theta}^{k+1}|\boldsymbol{\theta}^k) \geq Q(\boldsymbol{\theta}^k|\boldsymbol{\theta}^k)$. In other words, we aim for a scenario that merely provides an improvement. This strategy paves the way to the implementation of the generalized expectation-maximization algorithm as outlined in [41].

B. The Analytical Form of $\mathbf{A}^{k+1}$

By applying Lemma 2, we can establish that the conditional distribution of $\mathbf{x}_{i,\mathsf{m}}$ given $(\tau_i,\mathbf{x}_{i,\mathsf{o}},\boldsymbol{\Sigma})$ follows a normal distribution. This distribution has a mean vector defined as $\boldsymbol{\Sigma}^k_{i,\mathsf{mo}}(\boldsymbol{\Sigma}^k_{i,\mathsf{oo}})^{-1}\mathbf{x}_{i,\mathsf{o}}$ and a covariance matrix described by $(\boldsymbol{\Sigma}^k_{i,\mathsf{mm}} - \boldsymbol{\Sigma}^k_{i,\mathsf{mo}}(\boldsymbol{\Sigma}^k_{i,\mathsf{oo}})^{-1}\boldsymbol{\Sigma}^k_{i,\mathsf{om}})/\tau_i$. Then it is straightforward to have $\mathbb{E}_{\mathbf{X}_\mathsf{m},\boldsymbol{\tau}|\mathbf{X}_\mathsf{o},\boldsymbol{\Sigma}^k}\left[\tau_i\mathbf{x}_i\mathbf{x}_i^H\right] = \bar{\tau}_i^k\bar{\mathbf{x}}_i^k(\bar{\mathbf{x}}_i^k)^H + \boldsymbol{\Phi}_i^k$ with $\bar{\tau}_i^k = \mathbb{E}_{\mathbf{X}_\mathsf{m},\boldsymbol{\tau}|\mathbf{X}_\mathsf{o},\boldsymbol{\Sigma}^k}[\tau_i]$, $\bar{\mathbf{x}}_i^k = \mathbb{E}_{\mathbf{X}_\mathsf{m},\boldsymbol{\tau}|\mathbf{X}_\mathsf{o},\boldsymbol{\Sigma}^k}[\mathbf{x}_i]$, and $\boldsymbol{\Phi}_i^k$ is a $p\times p$ matrix whose (l,j)-th element is zero if either $x_{l,i}$ or $x_{j,i}$ is observed, otherwise is the corresponding element of the conditional covariance matrix of $\mathbf{x}_{i,\mathsf{m}}$ but ignoring τ_i.

Lemma 2 (Conditional Gaussian Distribution [49]): Suppose $(\mathbf{x},\mathbf{y}) \sim \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$, then we have

$$\mathbf{y}\,|\,\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{y}|\mathbf{x}},\boldsymbol{\Sigma}_{\mathbf{y}|\mathbf{x}}), \tag{23}$$

where

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{y}|\mathbf{x}} &= \boldsymbol{\mu}_{\mathbf{y}} + \boldsymbol{\Sigma}_{\mathbf{y},\mathbf{x}}\boldsymbol{\Sigma}_{\mathbf{x},\mathbf{x}}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}}),\\ \boldsymbol{\Sigma}_{\mathbf{y}|\mathbf{x}} &= \boldsymbol{\Sigma}_{\mathbf{y}} - \boldsymbol{\Sigma}_{\mathbf{y},\mathbf{x}}\boldsymbol{\Sigma}_{\mathbf{x},\mathbf{x}}^{-1}\boldsymbol{\Sigma}_{\mathbf{x},\mathbf{y}}.\end{aligned} \tag{24}$$

Note that, given $\mathbf{x}_{i,\mathsf{o}}$ and current $\boldsymbol{\Sigma}$, the conditional pdf of τ_i is

$$\begin{aligned}&p(\tau_i\,|\,\mathbf{x}_{i,\mathsf{o}},\boldsymbol{\Sigma})\\ &\propto p(\tau_i,\mathbf{x}_{i,\mathsf{o}}\,|\,\boldsymbol{\Sigma})\\ &= p(\mathbf{x}_{i,\mathsf{o}}\,|\,\tau_i,\boldsymbol{\Sigma})p(\tau_i\,|\,\boldsymbol{\Sigma})\\ &\propto \frac{\exp\left(-\mathbf{x}_{i,\mathsf{o}}^H(\frac{\boldsymbol{\Sigma}_{i,\mathsf{oo}}^{-1}}{\tau_i})^{-1}\mathbf{x}_{i,\mathsf{o}}\right)}{\det(\frac{\boldsymbol{\Sigma}_{i,\mathsf{oo}}}{\tau_i})}(\tau_i)^{\frac{\nu}{2}-1}\exp\left(-\frac{\nu}{2}\tau_i\right)\\ &\propto \tau_i^{\frac{\nu}{2}+p_i-1}\exp\left(-\mathbf{x}_{i,\mathsf{o}}^H\boldsymbol{\Sigma}_{i,\mathsf{oo}}^{-1}\mathbf{x}_{i,\mathsf{o}}\tau_i - \frac{\nu}{2}\tau_i\right).\end{aligned} \tag{25}$$

Comparing the above equation with (7), we can find that $\tau_i\,|\,\mathbf{x}_i,\boldsymbol{\Sigma}$ follows a gamma distribution:

$$\tau_i\,|\,\mathbf{x}_i,\boldsymbol{\Sigma} \sim \text{Gamma}\left(\frac{\nu}{2}+p_i,\frac{\nu}{2}+\mathbf{x}_{i,\mathsf{o}}^H\boldsymbol{\Sigma}_{i,\mathsf{oo}}^{-1}\mathbf{x}_{i,\mathsf{o}}\right). \tag{26}$$

Then the conditional mean of τ_i is simply

$$\mathbb{E}_{\mathbf{X}_\mathsf{m},\boldsymbol{\tau}|\mathbf{X}_\mathsf{o},\boldsymbol{\Sigma}^k}[\tau_i] = \frac{\nu/2+p_i}{\nu/2+\mathbf{x}_{i,\mathsf{o}}^H\boldsymbol{\Sigma}_{i,\mathsf{oo}}^{-1}\mathbf{x}_{i,\mathsf{o}}}. \tag{27}$$

C. Algorithm Development for Problem (16)

Given $\mathbf{S}\in\mathbb{C}^{p\times p}$ a complex-valued positive semidefinite matrix and $r<p$, we consider the following problem:

$$\begin{aligned}\max_{\mathbf{H},\boldsymbol{\psi}}\quad &\log\det(\boldsymbol{\Sigma}^{-1}) - \text{Tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{S}\right)\\ \text{s.t.}\quad &\boldsymbol{\Sigma} = \mathbf{H}\mathbf{H}^H + \text{Diag}\left(\boldsymbol{\psi}\right),\\ &\mathbf{H}\in\mathbb{C}^{p\times r},\boldsymbol{\psi}\in[\boldsymbol{\psi}_\text{L},\boldsymbol{\psi}_\text{U}].\end{aligned} \tag{28}$$

Naturally, this problem presents significant challenges due to its non-convex objectives and constraints. We propose to tackle Problem (28) by extending the algorithm presented in [40] to accommodate complex numbers and taking into account the upper bound constraint of the variable $\boldsymbol{\psi}$.

More specifically, we will initially incorporate the variable $\mathbf{H}$ into the variable $\boldsymbol{\psi}$ by applying Proposition 3.

Proposition 3: For a fixed $\boldsymbol{\psi}$, the optimal $\mathbf{H}$ to Problem (28) is

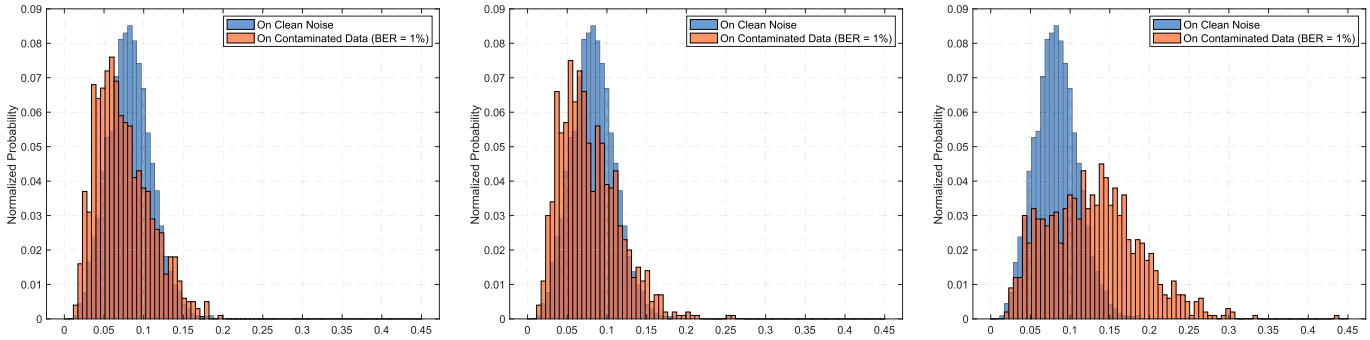
$$\mathbf{H}^\star(\boldsymbol{\psi}) = \text{Diag}(\boldsymbol{\psi}^{\frac{1}{2}})[\mathbf{u}_1,\ldots,\mathbf{u}_r]\text{Diag}(\tilde{\boldsymbol{\lambda}}), \tag{29}$$

where $\{\mathbf{u}_j,\lambda_j\}_{j=1}^p$ are the eigenvectors and eigenvalues of $\text{Diag}(\boldsymbol{\psi}^{-\frac{1}{2}})\mathbf{S}\text{Diag}(\boldsymbol{\psi}^{-\frac{1}{2}})$ in the descending order, and $\tilde{\boldsymbol{\lambda}} = \max\left([\lambda_1,\ldots,\lambda_r]^T - \mathbf{1}_r,\mathbf{0}\right)$.

Proof: See [31, Lemma 4] or [40, Proposition 2]. ■

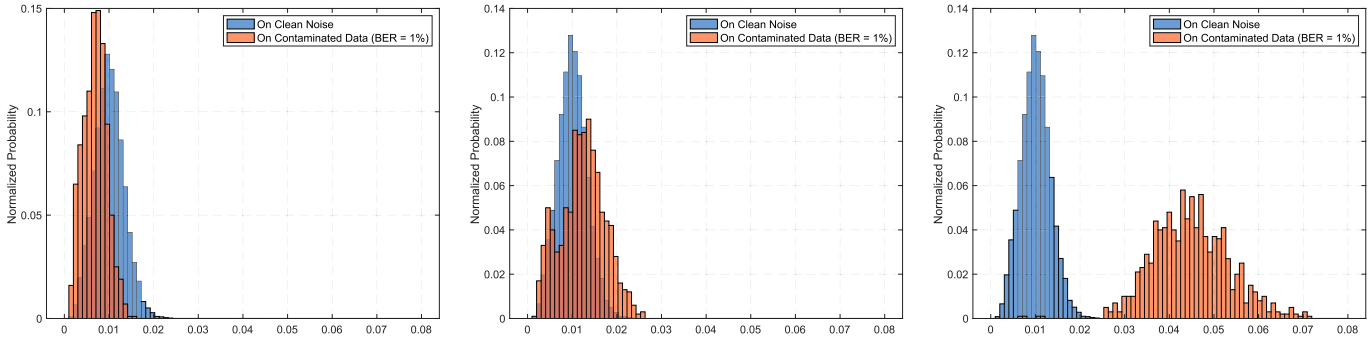
Subsequently, Problem (28) can be reformulated as the following new Problem (30) that depends solely on $\boldsymbol{\psi}$.

$$\begin{aligned}\max_{\boldsymbol{\psi}}\quad &\log\det(\boldsymbol{\Sigma}^{-1}) - \text{Tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{S}\right)\\ \text{s.t.}\quad &\boldsymbol{\Sigma} = \mathbf{H}^\star(\boldsymbol{\psi})(\mathbf{H}^\star(\boldsymbol{\psi}))^H + \text{Diag}\left(\boldsymbol{\psi}\right),\\ &\boldsymbol{\psi}\in[\boldsymbol{\psi}_\text{L},\boldsymbol{\psi}_\text{U}].\end{aligned} \tag{30}$$



(a) SNR = -100 dB. (b) SNR = -14 dB. (c) SNR = -10 dB.

Fig. 9. The empirical distributions of test statistics using classical GLRT under different SNR conditions, where the blue bins are derived from clean noise and are for reference only.



(a) SNR = -100 dB. (b) SNR = -14 dB. (c) SNR = -10 dB.

Fig. 10. The empirical distributions of test statistics using tGLRT under different SNR conditions, where the blue bins are derived from clean noise and are for reference only.

Furthermore, by defining the alternative variable $\boldsymbol{\alpha} = \boldsymbol{\psi}^{-1}$ and utilizing Proposition 4, we can adeptly tackle Problem (30) using the majorization-minimization (MM) technique [50].

Proposition 4: Let $\boldsymbol{\alpha} = \boldsymbol{\psi}^{-1}$, Problem (28) is equivalent to

$$\min_{\boldsymbol{\psi}_{\mathrm{U}}^{-1} \le \boldsymbol{\alpha} \le \boldsymbol{\psi}_{\mathrm{L}}^{-1}} \quad f(\boldsymbol{\alpha}) = f_1(\boldsymbol{\alpha}) - f_2(\boldsymbol{\alpha}), \tag{31}$$

whose majorization problem at $\boldsymbol{\alpha}$ is

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \quad & \check{f}(\boldsymbol{\alpha}) = \sum_{j=1}^{p} \left(S_{jj}\alpha_j - \log \alpha_j - \nabla_j(\boldsymbol{\alpha}^l)\alpha_j \right) \\ \text{s.t.} \quad & \boldsymbol{\psi}_{\mathrm{U}}^{-1} \le \boldsymbol{\alpha} \le \boldsymbol{\psi}_{\mathrm{L}}^{-1}, \end{aligned} \tag{32}$$

where $f_2(\boldsymbol{\alpha}) = \sum_{j=1}^{r} (\max\{1, \lambda_j\} - \log\max\{1, \lambda_j\} - 1)$, $f_1(\boldsymbol{\alpha}) = \sum_{j=1}^{p}(S_{jj}\alpha_j - \log\alpha_j)$, $\mathbf{U}\mathrm{Diag}(\boldsymbol{\lambda})\mathbf{U}^H$ is the eigenvalue decomposition of $\mathbf{S}^* = \mathrm{Diag}(\boldsymbol{\alpha}^{\frac{1}{2}})\mathbf{S}\,\mathrm{Diag}(\boldsymbol{\alpha}^{\frac{1}{2}})$, $\boldsymbol{\nabla}(\boldsymbol{\alpha}) = \mathrm{diag}(\mathrm{Diag}(\boldsymbol{\alpha}^{-1})\mathbf{U}\mathbf{D}\mathbf{U}^H\mathbf{S}^*)$ with

$$D_{jj} = \begin{cases} \max\{0, 1 - \lambda_j^{-1}\} & 1 \le j \le r \\ 0 & \text{otherwise.} \end{cases} \tag{33}$$

Proof: Directly extending [40, Corollary 2] to the complex number case completes the proof. ■

This entails iteratively solving Problem (32) and updating $\boldsymbol{\alpha}$ with its optimal solution. Given that the elements of $\boldsymbol{\alpha}$ are independent and the objective function is convex in the majorization

Algorithm 2: A MM algorithm for Problem (28).

1: Initialize $\mathbf{H}^0$, $\boldsymbol{\psi}^0$ and $\boldsymbol{\alpha}^0 = (\boldsymbol{\psi}^0)^{-1}$ and $\boldsymbol{\Sigma}^0 = \mathbf{H}^0(\mathbf{H}^0)^H + \boldsymbol{\Psi}^0$.
2: **for** $l = 0, 1, 2, \ldots$ **do**
3: $\quad \boldsymbol{\alpha}^{l+1} = \left[\mathrm{diag}(\mathbf{S}) - \boldsymbol{\nabla}(\boldsymbol{\alpha}^l))\right]_{\boldsymbol{\psi}_{\mathrm{U}}^{-1}}^{\boldsymbol{\psi}_{\mathrm{L}}^{-1}}$;
4: $\quad l \leftarrow l + 1$;
5: $\quad$ Terminate when converges;
6: **end for**
7: Recover $\boldsymbol{\psi}^\star = (\boldsymbol{\alpha}^l)^{-1}$ and $\mathbf{H}^\star$ via Proposition 3.
8: Return $\boldsymbol{\Sigma}^\star = \mathbf{H}^\star(\mathbf{H}^\star)^H + \mathrm{Diag}(\boldsymbol{\psi}^\star)$.

Problem (32), we can determine its optimal solution through a closed-form expression. Specifically, the update rule for $\boldsymbol{\alpha}$ is expressed as:

$$\boldsymbol{\alpha}^{l+1} = \left[\mathrm{diag}(\mathbf{S}) - \boldsymbol{\nabla}(\boldsymbol{\alpha}^l))\right]_{\boldsymbol{\psi}_{\mathrm{U}}^{-1}}^{\boldsymbol{\psi}_{\mathrm{L}}^{-1}}. \tag{34}$$

Upon convergence, we can retrieve the optimal $\boldsymbol{\psi}^\star = \boldsymbol{\alpha}^{-1}$ and $\mathbf{H}^\star$ using Proposition 3. The comprehensive algorithm is delineated in Alg. 2, which is guaranteed to converge to a stationary point of Problem (31), i.e., the equivalent formulation of Problem (28), as affirmed in [50].

D. Empirical Distribution Comparison Between ξ_{GLR} and ξ_{tGLR}

We compare the empirical distribution of test statistics from the classical GLRT and proposed tGLRT under different SNR conditions in Figs. 9 and 10. The trend of splitting the groups of test statistics between pure noise and contaminated data validates the superior performance of the tGLRT detector. Interestingly, as shown in Fig. 9(a), when the primary signal is non-existing, the empirical distribution of GLRT test statistics using contaminated data has significantly fatter tails than that using clean data. However, as shown in Fig. 10(a), when the primary signal is non-existing, the empirical distribution of tGLRT ($\nu = 2$) test statistics using contaminated data is significantly left-shifted compared with that using clean data. This can be explained by the fact that the covariance matrix is ill-estimated in the GLRT method with contaminated data, leading to diversely distributed test statistics. The covariance matrix is robustly estimated in the tGLRT method. But the sample covariance matrix tends to be more diagonally dominated, i.e., the diagonal elements become larger with contaminated data. Then, the test statistics tend to be smaller if fed with these contaminated data. This implies that the classical GLRT would require a higher threshold than usual to achieve the desired P_{FA} for contaminated data, which shall further deteriorate its detecting performance in practice.

REFERENCES

[1] X. Hong, J. Wang, C.-X. Wang, and J. Shi, "Cognitive radio in 5G: A perspective on energy-spectral efficiency trade-off," *IEEE Commun. Mag.*, vol. 52, no. 7, pp. 46–53, Jul. 2014.

[2] D. Wang, B. Song, D. Chen, and X. Du, "Intelligent cognitive radio in 5G: AI-based hierarchical cognitive cellular networks," *IEEE Wireless Commun.*, vol. 26, no. 3, pp. 54–61, Jun. 2019.

[3] F. A. Awin, Y. M. Alginahi, E. Abdel-Raheem, and K. Tepe, "Technical issues on cognitive radio-based internet of things systems: A survey," *IEEE Access*, vol. 7, pp. 97887–97908, 2019.

[4] Y. Zeng, Y.-C. Liang, A. T. Hoang, and R. Zhang, "A review on spectrum sensing for cognitive radio: Challenges and solutions," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, pp. 1–15, Jan. 2010.

[5] S. M. Mishra, A. Sahai, and R. W. Brodersen, "Cooperative sensing among cognitive radios," in *Proc. IEEE Int. Conf. Commun.*, vol. 4, Piscataway, NJ, USA: IEEE Press, 2006, pp. 1658–1663.

[6] J. Oksanen, J. Lundén, and V. Koivunen, "Characterization of spatial diversity in cooperative spectrum sensing," in *Proc. Int. Symp. Commun., Control Signal Process. (ISCCSP)*, 2010, pp. 1–5.

[7] J. Lunden, V. Koivunen, and H. V. Poor, "Spectrum exploration and exploitation for cognitive radio: Recent advances," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 123–140, May 2015.

[8] H. Nawaz, H. M. Ali, and A. A. Laghari, "UAV communication networks issues: A review," *Arch. Comput. Methods Eng.*, vol. 28, pp. 1349–1369, Mar. 2021.

[9] L. L. Peterson and B. S. Davie, *Computer Networks: A Systems Approach*. Amsterdam, Netherlands: Elsevier, 2007.

[10] J. Liu, S. Kumar, and D. P. Palomar, "Parameter estimation of heavy-tailed AR model with missing data via stochastic EM," *IEEE Trans. Signal Process.*, vol. 67, no. 8, pp. 2159–2172, Apr. 2019.

[11] R. Zhou, J. Liu, S. Kumar, and D. P. Palomar, "Student's t VAR modeling with missing data via stochastic EM and Gibbs sampling," *IEEE Trans. Signal Process.*, vol. 68, pp. 6198–6211, 2020.

[12] A. Aubry, A. De Maio, S. Marano, and M. Rosamilia, "Structured covariance matrix estimation with missing-(complex) data for radar applications via expectation-maximization," *IEEE Trans. Signal Process.*, vol. 69, pp. 5920–5934, 2021.

[13] S. Rani and A. Solanki, "Data imputation in wireless sensor network using deep learning techniques," in *Proc. Data Anal. Manage.: Proc. ICDAM*. Springer, 2021, pp. 579–594.

[14] G. DiCesare, *Imputation, Estimation and Missing Data in Finance*. ON, Canada: University of Waterloo, 2006.

[15] D. M. P. Murti, U. Pujianto, A. P. Wibawa, and M. I. Akbar, "K-nearest neighbor (k-NN) based missing data imputation," in *Proc. Int. Conf. Sci. Inf. Technol. (ICSITech),* Piscataway, NJ, USA: IEEE Press, 2019, pp. 83–88.

[16] Z. Zhang, "Missing data imputation: Focusing on single imputation," *Ann. Transl. Med.*, vol. 4, no. 1, pp. 1–8, 2016.

[17] P. Jain, P. Netrapalli, and S. Sanghavi, "Low-rank matrix completion using alternating minimization," in *Proc. 45th Annu. ACM Symp. Theory Comput.*, 2013, pp. 665–674.

[18] S. Rana, A. H. John, and H. Midi, "Robust regression imputation for analyzing missing data," in *Proc. Int. Conf. Statist. Sci., Bus. Eng. (ICSSBE)*, 2012, pp. 1–4.

[19] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[20] A. Boukerche, L. Zheng, and O. Alfandi, "Outlier detection: Methods, models, and classification," *ACM Comput. Surveys*, vol. 53, no. 3, pp. 1–37, 2020.

[21] B. Puranik, U. Madhow, and R. Pedarsani, "Generalized likelihood ratio test for adversarially robust hypothesis testing," *IEEE Trans. Signal Process.*, vol. 70, pp. 4124–4139, 2022.

[22] M. Fauß, A. M. Zoubir, and H. V. Poor, "Minimax robust detection: Classic results and recent advances," *IEEE Trans. Signal Process.*, vol. 69, pp. 2252–2283, 2021.

[23] L. Dabush and T. Routtenberg, "Detection of false data injection attacks in unobservable power systems by Laplacian regularization," in *Proc. IEEE 12th Sensor Array Multichannel Signal Process. Workshop (SAM),* Piscataway, NJ, USA: IEEE Press, 2022, pp. 415–419.

[24] A. S. Rawat, P. Anand, H. Chen, and P. K. Varshney, "Collaborative spectrum sensing in the presence of byzantine attacks in cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 774–786, Feb. 2011.

[25] M. Ghaznavi and A. Jamshidi, "A reliable spectrum sensing method in the presence of malicious sensors in distributed cognitive radio network," *IEEE Sensors J.*, vol. 15, no. 3, pp. 1810–1816, Mar. 2015.

[26] S. Zhang, Y. Wang, P. Wan, J. Zhuang, Y. Zhang, and Y. Li, "Clustering algorithm-based data fusion scheme for robust cooperative spectrum sensing," *IEEE Access*, vol. 8, pp. 5777–5786, 2020.

[27] Z. Liu, A. Kammoun, and M.-S. Alouini, "On robust spectrum sensing using m-estimators of covariance matrix," *Sci. China Inf. Sci.*, vol. 63, pp. 1–3, May 2020.

[28] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proc. IEEE*, vol. 55, no. 4, pp. 523–531, 1967.

[29] R. Zhang, T. J. Lim, Y.-C. Liang, and Y. Zeng, "Multi-antenna based spectrum sensing for cognitive radios: A GLRT approach," *IEEE Trans. Commun.*, vol. 58, no. 1, pp. 84–88, Jan. 2010.

[30] L. Huang, J. Fang, K. Liu, H. C. So, and H. Li, "An eigenvalue-moment-ratio approach to blind spectrum sensing for cognitive radio under sample-starving environment," *IEEE Trans. Veh. Technol.*, vol. 64, no. 8, pp. 3465–3480, Aug. 2015.

[31] D. Ramirez, G. Vazquez-Vilar, R. Lopez-Valcarce, J. Via, and I. Santamaria, "Detection of rank-p signals in cognitive radio networks with uncalibrated multiple antennas," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3764–3774, Aug. 2011.

[32] J. G. Proakis and M. Salehi, *Digital Communications*. New York, NY, USA: McGraw-hill, 2008.

[33] Y. Zeng and Y.-C. Liang, "Eigenvalue-based spectrum sensing algorithms for cognitive radio," *IEEE Trans. Commun.*, vol. 57, no. 6, pp. 1784–1793, Jun. 2009.

[34] F. J. Massey Jr, "The Kolmogorov-Smirnov test for goodness of fit," *J. Amer. Statist. Assoc.*, vol. 46, no. 253, pp. 68–78, 1951.

[35] G. Zhang, J. Lan, L. Zhang, F. He, and S. Li, "Filtering in pairwise Markov model with student's t non-stationary noise with application to target tracking," *IEEE Trans. Signal Process.*, vol. 69, pp. 1627–1641, 2021.

[36] T. Li, Z. Hu, Z. Liu, and X. Wang, "Multisensor suboptimal fusion student's t filter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 3, pp. 3378–3387, Jun. 2023.

[37] S. Kotz and S. Nadarajah, *Multivariate t-Distributions and Their Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[38] F. Mouret, A. Hippert-Ferrer, F. Pascal, and J.-Y. Tourneret, "A robust and flexible EM algorithm for mixtures of elliptical distributions with missing data," *IEEE Trans. Signal Process.*, Apr. 20, 2023, doi: 10.1109/TSP.2023.3267994.

[39] W. Breymann and D. Lüthi, "ghyp: A package on generalized hyperbolic distributions," Manual for R Package ghyp, 2013.
[40] K. Khamaru and R. Mazumder, "Computation of the maximum likelihood estimator in low-rank factor analysis," *Math. Program.*, vol. 176, pp. 279–310, Mar. 2019.
[41] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc.: Ser. B (Methodol.)*, vol. 39, no. 1, pp. 1–22, 1977.
[42] R. M. Neal and G. E. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in Graphical Models*, Springer, 1998, pp. 355–368.
[43] J. Law, *Robust Statistics: The Approach Based on Influence Functions*. Wiley Online Library, 1986.
[44] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5597–5625, Nov. 2012.
[45] C. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, Mar. 1983.
[46] The MathWorks, Inc., *Communications Toolbox*, Natick, Massachusetts, United State, 2022. [Online]. Available: https://www.mathworks.com/help/comm/
[47] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: identifying density-based local outliers," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2000, pp. 93–104.
[48] T. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, 1996.
[49] R. J. Harris, *A primer of multivariate statistics*. Psychology Press, 2001.
[50] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, 2016.

Rui Zhou received the B.Eng. degree in information engineering from the Southeast University, Nanjing, China, in 2017, and the Ph.D. degree from Hong Kong University of Science and Technology (HKUST), Hong Kong, in 2021. He is a Research Scientist with Shenzhen Research Institute of Big Data, Shenzhen, China. His research interests include optimization algorithms, statistical signal processing, machine learning, and financial engineering.

Wenqiang Pu received the B.S. and Ph.D. degrees in electrical engineering from Xidian University, Xi'an, China, in 2013 and 2018, respectively. From 2019 to 2020, he was a Postdoctoral Associate with the School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen). Currently, he is a Research Scientist with Shenzhen Research Institute of Big Data. His co-authored paper received Best Student Paper Award from IEEE SAM 2024. He serves as an Associate Editor of IEEE SIGNAL PROCESSING LETTERS. His research interests include signal processing and optimization algorithms.

Ming-Yi You received the B.S. and Ph.D. degrees in mechanical engineering from Shanghai Jiao Tong University, Shanghai, in 2006 and 2012, respectively. From 2007 to 2008, he was invited to visit M. S. Wu Manufacturing Research Center, University of Michigan. Currently, he is a Senior Expert with the No. 36 Research Institute of CETC and leading a Research Group in radio direction finding and localization. He has published over 50 articles and co-authored the book titled *Radio Direction Finding: Theory and Practice*. His research interests include condition-based maintenance, radio direction finding, wireless localization, and multiobject tracking.

Qingjiang Shi received the Ph.D. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2011. From 2009 to 2010, he visited Prof. Z.-Q. (Tom) Luo's Research Group with the University of Minnesota, Twin Cities. In 2011, he worked as a Research Scientist with Bell Labs, China. In 2012, he was with the School of Information and Science Technology, Zhejiang Sci-Tech University. From 2016 to 2017, he worked as a Research Fellow with Iowa State University, USA. Since 2018, he has been with the School of Software Engineering, Tongji University, where he is currently a Full Professor. He is also with Shenzhen Research Institute of Big Data. His research interests include algorithm design and analysis with applications in machine learning, signal processing, and wireless networks. So far, he has published more than 80 IEEE journals and filed about 40 national patents. He was an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was the recipient of the IEEE Signal Processing Society Best Paper Award in 2022, the Huawei Technical Cooperation Achievement Transformation Award (2nd Prize) in 2022, the Huawei Outstanding Technical Achievement Award in 2021, the Golden Medal at the 46th International Exhibition of Inventions of Geneva in 2018, the First Prize of Science and Technology Award from China Institute of Communications in 2017, the National Excellent Doctoral Dissertation Nomination Award in 2013, the Shanghai Excellent Doctoral Dissertation Award in 2012, and the Best Paper Award from the IEEE PIMRC'09 conference.