

A ROBUST GLRT DETECTOR AGAINST MISSING DATA IN COOPERATIVE SENSING

Jinghui Guan[‡], Rui Zhou[†], Wenqiang Pu[†], Qingjiang Shi^{§†}, and Tsung-Hui Chang^{‡†}

[‡]School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, China

[†]Shenzhen Research Institute of Big Data, Shenzhen, China

[§]School of Software Engineering, Tongji University, Shanghai, China

ABSTRACT

Cooperative sensing, a technique employed in cognitive radio (CR) networks for spectrum sensing, exhibits promising potential in bolstering spectrum utilization and enhancing network performance. This approach leverages the information captured by distributed CR users, which is subsequently aggregated at a fusion center. However, the challenges arise when the data are transmitted with low-quality, resulting in the consequential issue of missing data. These factors introduce complexity in detecting primary signals and undermine the reliability of cooperative sensing. In this study, we present a significant advancement in cooperative sensing methodologies by introducing a novel approach: a generalized likelihood ratio test (GLRT) type detector specifically designed to be robust to missing data. More specifically, our proposed robust GLRT detector modifies the computation of the classical GLRT test statistic to accommodate the inherent incompleteness of the data and effectively estimates the desired unknown parameters. Through numerical experiments, we demonstrate the resilience and robustness of our proposed cooperative signal detection method.

Index Terms—Cooperative sensing, generalized likelihood ratio test, missing data, robust detector.

1. INTRODUCTION

The ever-increasing demand for limited spectrum resources, coupled with the simultaneous occurrence of spectrum congestion and underutilization, presents significant managerial challenges in spectrum allocation [1, 2, 3]. To address this issue, cognitive radio (CR) technology has emerged as a promising solution [4]. By leveraging opportunistic spectrum access, CR technology empowers secondary users to detect unoccupied spectrum segments and dynamically reallocate them when primary users are not utilizing them. This approach not only maximizes the utilization of underutilized spectrum but also ensures the primary users' privileges are

safeguarded [5, 6]. The effective implementation of CR technologies relies heavily on the reliable detection of primary signals. Cooperative sensing, a method where devices collaborate to sense the spectrum, plays a crucial role in this regard. In this method, each CR user independently senses its assigned range and shares the acquired data with the fusion center for comprehensive analysis and integration [7, 8]. By pooling together their sensing capabilities, the fusion center can achieve a more accurate and reliable assessment of the spectrum, facilitating optimal utilization and efficient management of the available resources.

The realm of cooperative sensing encompasses a diverse range of approaches, including energy detection (ED) [9], cyclostationary detection (CD) [10], matched filtering detection (MFD) [11], likelihood ratio test (LRT), and generalized likelihood ratio test (GLRT) [12]. Each of these techniques possesses distinct strengths and capabilities, making them suitable for various spectrum sensing tasks. Their widespread application in different settings attests to their versatility. Notwithstanding the advancements achieved in these techniques, certain limitations persist, primarily due to their reliance on accurate raw data. The efficacy of signal detection methods hinges upon the precision and completeness of the underlying data [13]. In the presence of inaccurate or incomplete data, the reliability and accuracy of these methods can be compromised, thus limiting their full potential.

Indeed, in practical applications, the presence of missing data presents a substantial challenge to the accuracy of cooperative sensing [14]. Missing data can stem from various sources, such as the low signal-to-noise ratio (SNR) transmission from CR users to the fusion center or node failures [15]. When such situations arise, they introduce incomplete information into the dataset, significantly compromising the performance of the signal detection methods mentioned earlier. Traditionally, two approaches are commonly employed to address missing data: deletion and imputation [16]. Deletion involves the removal of samples that contain missing values, resulting in a dataset with complete observations. However, when the rate of missing data is relatively high, this method can lead to a substantial reduction in the dataset's size, rendering it potentially insufficient for further analysis or utilization. On the other hand, imputation involves the substitution

This work was supported in part by the National Nature Science Foundation of China (NSFC) under Grant 62201362, Grant 62101350, and in part by the Shenzhen Science and Technology Program (Grant No. RCBS20221008093126071).

of missing entries with derived values that conform to the estimated distribution of the observed data. While this method offers a means to fill in the missing values, it introduces uncertainty and may result in significant deviations from the true underlying data. Hence, it becomes crucial to explore more robust and dependable approaches for handling missing data in the context of cooperative sensing. By doing so, it is expected that we will ensure accurate and reliable signal detection even in the presence of incomplete information.

Therefore, this paper presents a cooperative sensing approach tailored to meet the specific needs of constructing a robust detector capable of tolerating missing data scenario. Our contributions can be delineated as follows:

- We propose a robust GLRT detector which is designed to be fed with data containing direct missing components.
- We employ the EM algorithm to estimate desired covariance matrix from such incomplete dataset.
- The numerical experiments are conducted to demonstrate the effectiveness of our proposed robust detector against missing data.

2. SIGNAL MODEL

Consider a scenario in which a cognitive radio (CR) network system is comprised of P CR nodes and a primary user equipped with r antennas. The signals transmitted by the primary user, denoted as $\{\mathbf{s}_i\}_{i=1}^N$ ($\mathbf{s}_i \in \mathbb{C}^r$), are stacked as a matrix $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_N] \in \mathbb{C}^{r \times N}$. It is commonly assumed that the primary signal follows a circularly symmetric complex Gaussian distribution with a zero mean. In this setup, the P CR nodes, each equipped with an antenna array, independently sense the transmission of the primary signal. They send their sampled data to the fusion center, which makes the decision regarding the presence or absence of the primary signal. The received signal matrix $\mathbf{X} \in \mathbb{C}^{P \times N}$ comprises observations from P CR nodes over N consecutive time slots. The wireless channel $\mathbf{H} \in \mathbb{C}^{P \times r}$, which connects the primary and secondary users, is assumed to follow a wide-sense stationary uncorrelated scattering (WSSUS) model. The system accounts for additive white Gaussian noise (AWGN) as $\mathbf{V} \in \mathbb{C}^{P \times N}$, assuming it to be independent and identically distributed (i.i.d.) across antennas and time. The noise follows a zero-mean Gaussian distribution with a diagonal covariance matrix Σ^2 . In this study, we assume that the parameter r is known, while the noise variances at each cognitive radio (CR) node are considered unknown.

The spectrum sensing issue is framed as a binary hypothesis testing problem [17], aiming to distinguish between,

$$\mathcal{H}_0 : \mathbf{X} = \mathbf{V}, \quad \mathcal{H}_1 : \mathbf{X} = \mathbf{H}\mathbf{S} + \mathbf{V}. \quad (1)$$

Denoting \mathbf{R} as the covariance matrix of the received signal, the hypothesis testing problem can be reformulated as:

$$\mathcal{H}_0 : \mathbf{R} \in \Theta_0, \quad \mathcal{H}_1 : \mathbf{R} \in \Theta_1. \quad (2)$$

Here, Θ_0 and Θ_1 are defined as follows,

$$\begin{aligned} \Theta_0 &:= \{\text{Diag}(\sigma^2) \mid \sigma^2 > 0\} \\ \Theta_1 &:= \{\mathbf{M} + \text{Diag}(\sigma^2) \mid \text{rank}(\mathbf{M}) \leq r, \mathbf{M} \succeq \mathbf{0}, \sigma^2 > 0\} \end{aligned} \quad (3)$$

The classical generalized likelihood ratio test (GLRT) [18] statistic is then calculated as the ratio of the generalized likelihoods, ξ , under the two hypotheses. The signal existence is determined if ξ exceeds a predetermined threshold γ , i.e.,

$$\xi = \frac{f(\mathbf{X}|\hat{\mathbf{R}}_1)}{f(\mathbf{X}|\hat{\mathbf{R}}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \quad (4)$$

where f represents the likelihood of a complex multivariate Gaussian distribution given complete observation \mathbf{X} , $\hat{\mathbf{R}}_0$ and $\hat{\mathbf{R}}_1$ are the estimated covariance matrices from \mathbf{X} under the two hypotheses.

Indeed, when the observed data matrix \mathbf{X} is incomplete, the classical GLRT detector becomes infeasible. In the subsequent section, we will delve into the discussion of a robust detector that can effectively handle missing data in a sequential manner. This approach aims to overcome the limitations posed by incomplete observations, ensuring reliable and accurate signal detection.

3. ROBUST DETECTOR AGAINST MISSING DATA

Denote \mathbf{X}_{obs} as the observed data at the fusion center, we propose to use the following test statistic to decide the existence of the primary signal:

$$\xi = \frac{f(\mathbf{X}_{obs}|\hat{\mathbf{R}}_1)}{f(\mathbf{X}_{obs}|\hat{\mathbf{R}}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \quad (5)$$

where f represents the likelihood of a complex multivariate Gaussian distribution given partial observation \mathbf{X}_{obs} , i.e.,

$$f(\mathbf{X}_{obs} | \mathbf{R}) = \prod_{i=1}^N \frac{\exp(-\mathbf{x}_{i,obs}^H \mathbf{R}_{i,obs}^{-1} \mathbf{x}_{i,obs})}{(\pi)^{P_i} \det(\mathbf{R}_{i,obs})}, \quad (6)$$

$\mathbf{x}_{i,obs} \in \mathbb{C}^{P_i}$ denotes the observed part of \mathbf{x}_i , $\mathbf{R}_{i,obs}$ represents the sub-matrix of \mathbf{R} corresponding to the observed entries of \mathbf{x}_i , $\hat{\mathbf{R}}_0$ and $\hat{\mathbf{R}}_1$ are the estimated covariance matrices from \mathbf{X}_{obs} under the two hypotheses, i.e.,

$$\hat{\mathbf{R}}_0 = \arg \max_{\mathbf{R} \in \Theta_0} f(\mathbf{X}_{obs} | \mathbf{R}), \quad (7)$$

$$\hat{\mathbf{R}}_1 = \arg \max_{\mathbf{R} \in \Theta_1} f(\mathbf{X}_{obs} | \mathbf{R}). \quad (8)$$

Indeed, obtaining accurate estimates for the covariance matrices, namely $\hat{\mathbf{R}}_0$ and $\hat{\mathbf{R}}_1$, from the partially observed matrix \mathbf{X}_{obs} is not a straightforward task due to the complex expression of $f(\mathbf{X}_{obs}|\mathbf{R})$. In the subsequent discussion, we will address the approach to solving problem (7) and problem (8), which pertain to the estimation of these covariance matrices.

3.1. Estimate $\hat{\mathbf{R}}_0$ Under \mathcal{H}_0

Let us consider the optimization problem at hand:

$$\max_{\mathbf{R} \in \Theta_0} f(\mathbf{X}_{obs}|\mathbf{R}). \quad (9)$$

Given that \mathbf{R} is actually a diagonal matrix defined as $\mathbf{R} = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_P^2)$, each element within \mathbf{R} is independent. As a result, it is feasible to estimate each element $\sigma_j^2, j = 1, \dots, P$ independently. To derive the log-likelihood with respect to σ_j^2 and find its maximum, we set the derivative to zero, i.e.,

$$(\sigma_j^2)^* = \frac{1}{N_j} \sum_{i=1}^{N_j} [\mathbf{x}_{i,obs}]_j^H [\mathbf{x}_{i,obs}]_j \quad (10)$$

Here, N_j represents the number of samples for the j -th dimension. This implies that we take the average over all observed values of dimension j to estimate σ_j^2 .

3.2. Estimate $\hat{\mathbf{R}}_1$ Under \mathcal{H}_1

Consider the following optimization problem:

$$\max_{\mathbf{R} \in \Theta_1} f(\mathbf{X}_{obs}|\mathbf{R}). \quad (11)$$

Due to the presence of missing values, directly solving the aforementioned problem becomes challenging. Therefore, we propose employing the expectation-maximization (EM) algorithm [19] to estimate $\hat{\mathbf{R}}_1$ using \mathbf{X}_{obs} . The EM algorithm, which is an iterative process consisting of two steps, is employed to find the maximum likelihood estimates of parameters in probabilistic models that include latent variables or missing data [20]. It is worth noting that the complete data log-likelihood function is expressed as follows:

$$\begin{aligned} \mathcal{L}(\mathbf{R}|\mathbf{X}) &= \log \prod_{i=1}^N f(\mathbf{x}_i|\mathbf{R}) \\ &= -N \log |\mathbf{R}| - \sum_{i=1}^N \text{Tr}(\mathbf{R}^{-1} \mathbf{x}_i \mathbf{x}_i^H). \end{aligned} \quad (12)$$

Expectation (E) Step Let us denote the covariance matrix at the t -th iteration as $\mathbf{R}^{(t)}$. Based on this, we formulate the surrogate function as follows:

$$\begin{aligned} Q(\mathbf{R}|\mathbf{R}^{(t)}) &= \mathbb{E}_{\mathbf{X}_{mis}|\mathbf{X}_{obs}, \mathbf{R}^{(t)}} [\mathcal{L}(\mathbf{R}|\mathbf{X})] \\ &= -N \log |\mathbf{R}| - \text{Tr}(\mathbf{R}^{-1} \mathbf{S}^{(t)}), \end{aligned} \quad (13)$$

where $\mathbf{S}^{(t)} = \sum_{i=1}^n \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^H + \Phi_i$, \mathbf{X}_{mis} represents the missing values in \mathbf{X} , $\hat{\mathbf{x}}_i$ is the conditional mean of \mathbf{x}_i whose missing elements will be found using

$$\begin{aligned} \hat{\mathbf{x}}_{i,mis} &= \mathbb{E}(\mathbf{x}_{i,mis}|\mathbf{x}_{i,obs}, \mathbf{R}^{(t)}) \\ &= \mathbf{R}_{i,mis,obs}^{(t)} (\mathbf{R}_{i,obs,obs}^{(t)})^{-1} \mathbf{x}_{i,obs}, \end{aligned} \quad (14)$$

the (j, k) -th element of $\Phi_i \in \mathbb{C}^{P \times P}$ is zero if either $x_{j,t}$ or $x_{k,t}$ is observed, otherwise is the corresponding element of the conditional covariance matrix of $\mathbf{x}_{t,mis}$, i.e.,

$$\begin{aligned} &\mathbb{E}(\mathbf{x}_{i,mis} \mathbf{x}_{i,mis}^H | \mathbf{x}_{i,obs}, \mathbf{R}^{(t)}) \\ &= \mathbf{R}_{i,mis,mis} - \mathbf{R}_{i,mis,obs} \mathbf{R}_{i,obs,obs}^{-1} \mathbf{R}_{i,obs,mis}. \end{aligned} \quad (15)$$

Maximization (M) Step In the M step, our goal is to solve the following sub-problem:

$$\max_{\mathbf{R} \in \Theta_1} Q(\mathbf{R}|\mathbf{R}^{(t)}). \quad (16)$$

This problem can be effectively resolved by utilizing the algorithm presented in [21, Algorithm 1]. Moreover, we update $\mathbf{R}^{(t+1)}$ using its optimal solution obtained from this algorithm.

The Overall Algorithm To summarize, in the *E step*, we compute the statistics $\mathbf{S}^{(t)}$ based on the current $\mathbf{R}^{(t)}$. In the *M step*, we utilize $\mathbf{S}^{(t)}$ to update $\mathbf{R}^{(t+1)}$. Based on these steps, we can formulate the Algorithm 1 as follows.

Algorithm 1 An EM-based algorithm for Problem (11).

- 1: Initialize \mathbf{R}^0 .
 - 2: **for** $t = 0, 1, 2, \dots$ **do**
 - 3: **E-step:** compute $\mathbf{S}^{(t)}$ using (13);
 - 4: **M-step:** update $\mathbf{R}^{(t+1)}$ by solving problem (16) via [21, Algorithm 1];
 - 5: $t \leftarrow t + 1$;
 - 6: Terminate when converges;
 - 7: **end for**
 - 8: Return \mathbf{R}^t .
-

4. SIMULATION RESULTS

This section presents numerical experiments to demonstrate the efficacy of our proposed robust cooperative spectrum sensing techniques within a defined simulation environment.

We simulate an environment with eight distributed CR users ($P = 8$). The goal is to detect a single primary signal source ($r = 1$), transmitting Quadrature Phase Shift Keying (QPSK) modulated signals at a baud rate of 20 kHz. The receivers, operating at a sampling rate of 100 kHz, connect to the signal source via independent Rician fading channels with a K-factor of 4. The noise power at the CR users, represented

as $\hat{\sigma}_i^2$ ($i = 1, \dots, P$), equals β , uniformly distributed within an interval of $[-1, 1]$ dB. Each simulation uses $N = 500$ consecutive samples, or approximately 5 milliseconds. The threshold γ is derived from the empirical test statistics of 10000 pure noise data realizations, setting the probability of false alarm P_{FA} to 1%.

We initially carry out experiments to evaluate the probability of detection with varying proportions of missing data. Figure 1 depicts the relationship between the detection probability of our proposed robust GLRT method and the SNR. The proportions of missing data are set at 0%, 20%, 40%, and 50%, with the 0% condition serving as a baseline for comparison. The results demonstrate that an increase in SNR consistently enhances detection performance across all conditions, and a lower percentage of missing data correlates with improved performance. Notably, the experimental outcomes reveal that our method experiences only a marginal loss of about 3 to 4 dB in equivalent SNR, even with a data omission rate of 50%.

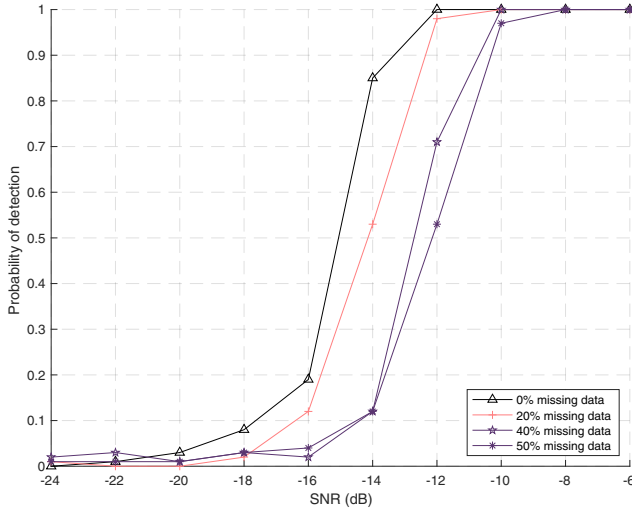


Fig. 1: Probability of detection versus SNR under different percentages of missing data.

In this investigation, we shift our focus to an alternative strategy for handling missing values. We include three other methods as benchmarks to compare their performance with our proposed algorithm, which are:

- Classical GLRT method fed with (accurate) full data;
- Classical GLRT method fed with data after imputation;
- Classical GLRT method fed with data after deletion.

We maintain the settings of the data generation process as before. The appropriate thresholds for detecting the presence of the primary signal in \mathbf{X}_{obs} are pre-calculated using pure noises. We analyze how the probability of detection (P_D) changes with variations in the average SNRs of CR users. To

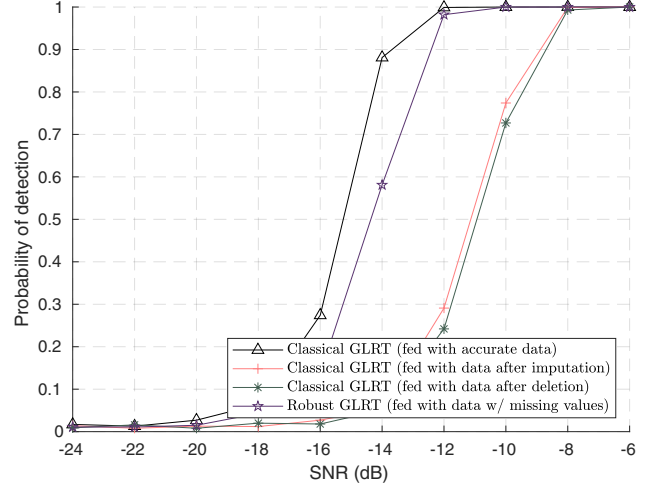


Fig. 2: Probability of detection versus SNR (20% missing data at random).

ensure efficiency and avoid excessive sample elimination, we set the missing rate at a modest value of 20% for the deletion method.

The results obtained from our investigation are presented in Figure 2. It is evident that our robust detector outperforms both the deletion and imputation methods, exhibiting a significantly higher P_D response at the same SNR which is defined in the time domain. Therefore, in practical applications, our algorithm proves to be an effective solution for addressing the missing data problem.

5. CONCLUSION

In this paper, we have presented a novel strategy to tackle the challenge of signal detection in cognitive radio systems, specifically when confronted with missing data. Our approach introduces a robust detecting method that involves estimating the structured covariance matrices from data containing missing values. The performance of this method is commendable, even in the presence of missing data. Through extensive simulations, we have demonstrated the effectiveness and robustness of our proposed detecting method. These results validate the efficacy of our approach and highlight its potential for practical implementation in cognitive radio systems.

6. REFERENCES

- [1] S. M. Mishra, A. Sahai, and R. W. Brodersen, "Cooperative sensing among cognitive radios," in *2006 IEEE International conference on communications*, vol. 4. IEEE, 2006, pp. 1658–1663.
- [2] S. Kim, "Inspection game based cooperative spectrum

- sensing and sharing scheme for cognitive radio iot system,” *Computer Communications*, vol. 105, pp. 116–123, 2017.
- [3] Z. Shi, X. Xie, H. Lu, H. Yang, M. Kadoch, and M. Cheriet, “Deep-reinforcement-learning-based spectrum resource management for industrial internet of things,” *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3476–3489, 2020.
 - [4] W. Lee, M. Kim, and D.-H. Cho, “Deep cooperative sensing: Cooperative spectrum sensing based on convolutional neural networks,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 3005–3009, 2019.
 - [5] D. B. Cabric, “Cognitive radios: System design perspective,” Ph.D. dissertation, University of California, Berkeley, 2007.
 - [6] E. C. Y. Peh, Y.-C. Liang, Y. L. Guan, and Y. Zeng, “Optimization of cooperative sensing in cognitive radio networks: A sensing-throughput tradeoff view,” *IEEE Transactions on Vehicular Technology*, vol. 58, no. 9, pp. 5294–5299, 2009.
 - [7] J. Unnikrishnan and V. V. Veeravalli, “Cooperative sensing for primary detection in cognitive radio,” *IEEE Journal of selected topics in signal processing*, vol. 2, no. 1, pp. 18–27, 2008.
 - [8] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, “Cooperative spectrum sensing in cognitive radio networks: A survey,” *Physical communication*, vol. 4, no. 1, pp. 40–62, 2011.
 - [9] A. Ali and W. Hamouda, “Advances on spectrum sensing for cognitive radio networks: Theory and applications,” *IEEE communications surveys & tutorials*, vol. 19, no. 2, pp. 1277–1304, 2016.
 - [10] K. Kim, I. A. Akbar, K. K. Bae, J.-S. Um, C. M. Spooner, and J. H. Reed, “Cyclostationary approaches to signal detection and classification in cognitive radio,” in *2007 2nd IEEE international symposium on new frontiers in dynamic spectrum access networks*. IEEE, 2007, pp. 212–215.
 - [11] X. Zhang, R. Chai, and F. Gao, “Matched filter based spectrum sensing and power level detection for cognitive radio network,” in *2014 IEEE global conference on signal and information processing (GlobalSIP)*. IEEE, 2014, pp. 1267–1270.
 - [12] Y. Zeng, Y.-C. Liang, A. T. Hoang, and R. Zhang, “A review on spectrum sensing for cognitive radio: challenges and solutions,” *EURASIP journal on advances in signal processing*, vol. 2010, pp. 1–15, 2010.
 - [13] Y. Wang, J. Li, and P. Stoica, *Spectral analysis of signals: The missing data case*. Morgan & Claypool Publishers, 2006.
 - [14] N. Deepan and B. Rebekka, “Performance analysis of wireless powered sensor network with opportunistic scheduling over generalized κ - μ shadowed fading channels,” *Physical Communication*, vol. 53, p. 101727, 2022.
 - [15] B. Sklar, *Digital communications: fundamentals and applications*. Pearson, 2021.
 - [16] F. Cismondi, A. S. Fialho, S. M. Vieira, S. R. Reti, J. M. Sousa, and S. N. Finkelstein, “Missing data in medical databases: Impute, delete or classify?” *Artificial intelligence in medicine*, vol. 58, no. 1, pp. 63–72, 2013.
 - [17] Y. Arjoune and N. Kaabouch, “A comprehensive survey on spectrum sensing in cognitive radio networks: Recent advances, new challenges, and future research directions,” *Sensors*, vol. 19, no. 1, p. 126, 2019.
 - [18] E. Conte, A. De Maio, and G. Ricci, “GLRT-based adaptive detection algorithms for range-spread targets,” *IEEE transactions on signal processing*, vol. 49, no. 7, pp. 1336–1348, 2001.
 - [19] G. J. McLachlan and T. Krishnan, *The EM algorithm and extensions*. John Wiley & Sons, 2007.
 - [20] 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 Transactions on Signal Processing*, vol. 68, pp. 6198–6211, 2020.
 - [21] D. Ramírez, G. Vazquez-Vilar, R. López-Valcarce, J. Vía, and I. Santamaría, “Detection of rank- p signals in cognitive radio networks with uncalibrated multiple antennas,” *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3764–3774, 2011.