2023 IEEE 24th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) | 978-1-6654-9626-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/SPAWC53906.2023.10304532

Weak Signal Detection based on Beta Divergence

Jiahao Liu[‡†], Haoxuan Liu[‡†], Wenqiang Pu[†], Rui Zhou[†], Ming-Yi You[‡], and Qingjiang Shi[§†]

[‡]Science and Technology on Communication Information Security Control Laboratory, Jiaxing, China
[†]Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong, Shenzhen, China
[§]School of Software Engineering, Tongji University, Shanghai, China

***Abstract*—Among different electromagnetic spectrum monitoring tasks, a fundamental one is detecting the presence of weak communication signal under non-cooperative setting. Various detection methods have been studied in the literature. These detection methods usually construct a proper hypothesis testing metric to distinguish the signal-plus-noise and noise-only signals. Information entropy based metric has recently been studied and shown its superiority over traditional energy based metrics. We find that it can be explained as a realization of the general Kullback-Leibler divergence. Inspired by that, in this paper, we propose to detect the signal using a statistic divergence called β-divergence. Different choices of β will lead to several well known divergences, e.g., Itakura-Saito divergence and Euclidean distance. We study its detection performance for non-cooperative signal sources. Both simulation and real experiments demonstrate that β-divergence can improve detection performance for various modulation types of communication signals.**

***Index Terms*—Spectrum monitoring, detection, β-divergence**

I. Introduction

In the past decades, rapidly developed wireless communication technology have produced different types of signals over the space. The limited spectrum space is becoming more and more crowded. Effectively monitoring the spectrum space has become an important issue in cognitive radio. Due to various factors, e.g., channel fading and frequency hopping, the wireless communication signals are weaker even compared with the background noise, i.e., admitting low signal-to-noise ratio (SNR) status. Conducting non-cooperative spectrum monitoring task on these weak signals is difficult, which brings serious challenges to the regulation of electromagnetic spectrum space.

One of the most widely used detection methods is energy detection method [1], which does not need any prior information and can work well in the high SNR case. The energy detection method requires reasonably computational cost and can be applied to a variety of devices. However, when the signal is weak, e.g., SNR is below -10dB, the performance of the energy detection method may become unstable because the noise causes false positives. If statistical parameters of the signal change periodically, detection methods based on the cyclic stationary features [2, 3] can achieve better detection performance than the matched filter detector [4]. However, this method requires prior information about the signal, leading to the limited applications in practice. There are some other methods for blind signal detection, such as the machine learning based technique [5–7] and information entropy based method [8, 9]. Machine learning based methods learn a proper binary classifier from a training set which collects a branch of different types of signals. But their performance on signal types beyond the training set may be unstable. Entropy based methods utilize the entropy of the signal histogram in frequency domain [9] or the normalized power spectrum density [8] as the test statistics, which is shown to work well for different signal types and be robust to the uncertainty of noise power [8, 9]. The entropy based detection method is a promising method for non-cooperative spectrum monitoring.

The three steps of the entropy based detection method are: i) reconstructing a proper discrete probability distribution, e.g., normalized signal histogram in frequency domain [9] or the normalized power spectrum density [8], from the received signal; ii) calculating the entropy of the obtained probability distribution; iii) comparing the achieved entropy with the predefined threshold to determine the existence of the signal. From a mathematical point of view, entropy of a certain probability can be regarded as the Kullback-Leibler (KL) divergence between this probability and a reference distribution. It motivates us to ask whether there is another divergence measurement rather than entropy method that may have better detection performance. Toward this end, this paper will study a class of statistical divergence called β-divergence for detection task. Both simulation and real experiments demonstrate that β-divergence improves detection performance over existing methods.

This paper is organized as follows. We first provide the signal model and our considered problem scenario in Sec. II. In Sec. III, we introduce the proposed detection method. Numerical experiments are given in Sec. IV. Finally, conclusions are summarized in Sec. V.

II. Non-cooperative Spectrum Detection Problem

Considering to detect a given frequency band, the received base-band signal with sampling rate f_s is expressed as:

$$x(n) = s(n) + w(n),\ n = 0, 1, \ldots, N-1, \tag{1}$$

where $x(n)$ is the received signal at the sample index n, $s(n)$ is the primary user signal with band width B ($B < f_s/2$), $w(n)$

Jiahao Liu and Haoxuan Liu are visiting undergraduate students at Shenzhen Research Institute of Big Data. This work is done during their visit and they contribute equally. This work was supported in part by the National Nature Science Foundation of China (NSFC) under Grant 62101350 and 62201362, and Shenzhen Science and Technology Program under Grant No. RCBS20221008093126071. (*Correspondence: Wenqiang Pu and Rui Zhou.*)

represents background noise, which is assumed to be zero-mean White Gaussian Noise (WGN) with a variance of σ_0^2, N is the overall sample size, and $s(n)$ and $w(n)$ are unrelated to each other.

The spectrum detection problem can be formulated as binary hypotheses testing problem, i.e.,

$$\begin{aligned} \mathcal{H}_0 &: x(n) = w(n), & n = 0, 1, \ldots, N-1, \\ \mathcal{H}_1 &: x(n) = s(n) + w(n), & n = 0, 1, \ldots, N-1, \end{aligned} \tag{2}$$

where $\mathcal{H}_0$ and $\mathcal{H}_1$ represent the absence and presence of signal occupation, respectively. For notational simplicity, we denote by $\boldsymbol{x} = [x(0), x(1), \ldots, x(N-1)]^T$ and $\boldsymbol{w} = [w(0), w(1), \ldots, w(N-1)]^T$. Under Neyman-Pearson (NP)-criterion, the optimal hypothesis testing is likelihood ratio test (LRT) [10]

$$T_{LRT}(\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\mathcal{H}_0)}{p(\boldsymbol{x}|\mathcal{H}_1)} \tag{3}$$

But LRT detector may not be applicable since the statistical information of the signal is not accessible, i.e., the exact evaluation of $p(\boldsymbol{x}|\mathcal{H}_0)$ and $p(\boldsymbol{x}|\mathcal{H}_1)$ are impossible, in the non-cooperative spectrum monitoring scenario.

Several existing works, e.g., energy detection [1], entropy based detection [8, 9], attempt to deal this issue by constructing a proper hypothesis testing function $\mathcal{L}(\cdot) : \mathbb{C}^N \mapsto \mathbb{R}$ with $\mathbb{E}[\mathcal{L}(\boldsymbol{w})] = c$, where c is a constant real number. Then the detection decision is made by

$$\mathcal{L}(\boldsymbol{x}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} c + \epsilon, \tag{4}$$

where $\epsilon \geq 0$ is a parameter controlling the false alarm rate P_F, i.e., $\mathbb{P}(\mathcal{L}(\boldsymbol{w}) \geq c + \epsilon) \leq P_F$.

Intuitively, it is expected that the overlap region of $\mathcal{L}(\boldsymbol{x})$ under hypothesises $\mathcal{H}_0$ and $\mathcal{H}_1$ to be as small as possible. It is equivalent to maximizing the probability of detection P_D while satisfying P_F. Thus the NP-criterion for $\mathcal{L}(\cdot)$ is

$$\max_{\mathcal{L}(\cdot), \lambda} \ \mathbb{P}(\mathcal{L}(\boldsymbol{x}) \geq \lambda \mid \mathcal{H}_1) \quad \text{s.t. } \mathbb{P}(\mathcal{L}(\boldsymbol{w}) \geq \lambda) \leq P_F. \tag{5}$$

However, directly solving the above problem faces several challenges, e.g., optimizing high-dimensional function mapping $\mathcal{L}(\cdot)$ is intractable, evaluating $\mathbb{P}(\cdot)$ is also difficult since the statistical information of signal is inaccessible. Machine learning based methods [5–7] parameterize a binary classifier. But they may be unstable if the signal type does not appear in the training data set.

For non-cooperative spectrum monitoring, one representative choice of $\mathcal{L}(\cdot)$ is information entropy [8], whose hypothesis testing metric is

$$\mathcal{L}(\boldsymbol{x}) = -\sum_{\ell=1}^{L} P_\ell(\boldsymbol{x}) \log P_\ell(\boldsymbol{x}) := T_E(\boldsymbol{x}), \tag{6}$$

where $P_\ell(\boldsymbol{x})$ is the normalized power spectrum density (PSD), i.e., $\sum_\ell P_\ell(\boldsymbol{x}) = 1$, represents the value of the ℓ-th frequency bin of signal $\boldsymbol{x}$. It can be estimated by many methods such as the Welch method [8]. The intuition behind this entropy detection method is discussed from information theory view [8]. We will later illustrate that it can also be explained from divergence metric. Note that WGN $\boldsymbol{w}$ should ideally have a uniform constant PSD over different frequency bins. Therefore its estimated normalized PSD $P_\ell(\boldsymbol{w})$ is expected to approximately satisfy $P_\ell(\boldsymbol{w}) \approx 1/L$. The following fact shows that $T_E(\boldsymbol{x})$ is the KL divergence between $P_\ell(\boldsymbol{x})$ and $1/L$ except a constant difference.

Fact 1. *Define hypothesis testing metric $T_{KL}(\boldsymbol{x})$ as*

$$T_{KL}(\boldsymbol{x}) = \sum_{\ell=1}^{L} KL(P_\ell(\boldsymbol{x}), 1/L), \tag{7}$$

It is easy to obtain the following equation:

$$T_{KL}(\boldsymbol{x}) = -T_E(\boldsymbol{x}) + \log(L), \tag{8}$$

where $KL(x, y) : \mathbb{R}_+^2 \mapsto \mathbb{R}_+$ is the generalized KL divergence defined as $KL(x, y) = x \log(x/y) + y - x$.

The statement in Fact 1 implies that $T_{KL}(\boldsymbol{x})$ shall achieve the equivalent detection performance of $T_E(\boldsymbol{x})$ by adding a constant to its detection threshold. Consequentially, the aforementioned equivalence relation poses a question that whether there is other divergence metric that can be used other than the KL divergence. In the next section, we will introduce a statistical divergence called β-divergence[1], which has been shown useful in various fields, e.g., neural signal analysis [11], music signal processing [12], and low rank tensor fitting [13].

III. Proposed Detection Method

The summation form of $T_{KL}(\boldsymbol{x})$, i.e., equation (7), in principle can be extended as the following mathematical form

$$T(\boldsymbol{x}) = \sum_{\ell=1}^{L} d(P_\ell(\boldsymbol{x}), 1/L), \tag{9}$$

where $d(x, y) : \mathbb{R}_+^2 \mapsto \mathbb{R}_+$ is a properly defined function. The β-divergence specifies $d(x, y)$ as:

$$d_\beta(x, y) = \begin{cases} \frac{x}{y} - \log(\frac{x}{y}) - 1, & \beta = 0, \\ x \log(\frac{x}{y}) + y - x, & \beta = 1, \\ \frac{x^\beta + (\beta-1)y^\beta - \beta x y^{\beta-1}}{\beta(\beta-1)}, & \text{o.w.}, \end{cases} \tag{10}$$

where x, y are non-negative numbers. Note that $d_\beta(x, y) = 0$ if and only if $x = y$ and it subsumes the Itakura-Saito divergence ($\beta = 0$), the KL divergence ($\beta = 1$), and the Euclidean distance ($\beta = 2$) as special cases. Fig. 1 shows the function curves corresponding to different β values, where y is fixed at 0.5. It can be seen that when x is equal to y, the function value is 0, and different β values will lead to different slopes on both sides of the curve. Consequentially, β-divergence based hypothesis testing metric $T_\beta(\boldsymbol{x})$ can be constructed as

$$T_\beta(\boldsymbol{x}) = \sum_{\ell=1}^{L} d_\beta(P_\ell(\boldsymbol{x}), 1/L) \tag{11}$$

[1] Our motivation is simply from a metric selection perspective, theoretical justification of the statistical property of using β-divergence would be an interesting future direction.

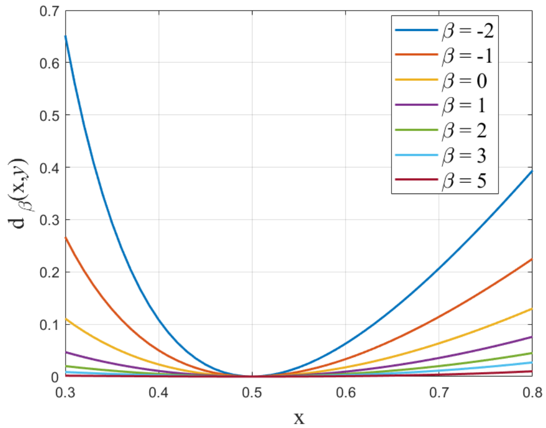


Fig. 1: Metric function curve when $y = 0.5$.

The intuition of our considered β-divergence is: if the received signal $\boldsymbol{x}$ only contains noise, $P_\ell(\boldsymbol{x})$ shall be close to $1/L$, leading to small value of $T_\beta(\boldsymbol{x})$; otherwise $T_\beta(\boldsymbol{x})$ shall be large. Decision is simply made by $T_\beta(\boldsymbol{x}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda$, where $\lambda \geq 0$ is the detection threshold. Note that the choice of λ depends on the desired false alarm rate P_F. For a given β value, if assuming $T_\beta(\boldsymbol{x})$ follows a Gaussian distribution with the mean value μ and the variance σ^2, the threshold λ is determined by

$$\lambda = \sigma Q^{-1}(P_F) + \mu \tag{12}$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} exp(\frac{-t^2}{2})dt$. However, such Gaussian assumption may not hold, i.e., $T_\beta(\boldsymbol{w})$ follows an unmodeled distribution. A practical way to select λ is empirical validating P_F from a branch of noise samples. Finally, we remark that the detection performance of $T_\beta(\boldsymbol{x})$ has an interesting invariant property (see Fact 2), which allows us to arbitrary scale $P_\ell(\boldsymbol{x})$ while the detection performance remains the same.

Fact 2. *Let $\eta > 0$ be a positive constant and define hypothesis testing metric $T_\beta(\boldsymbol{x}, \eta)$ as*

$$T_\beta(\boldsymbol{x}, \eta) = \sum_{\ell=1}^{L} d_\beta(\eta P_\ell(\boldsymbol{x}), \eta/L) \tag{13}$$

Then, we have $T_\beta(\boldsymbol{x}, \eta) = \eta^\beta T_\beta(\boldsymbol{x})$.

The statement in Fact 2 implies that $T_\beta(\boldsymbol{x}, \eta)$ can achieve the same detection performance of $T_\beta(\boldsymbol{x})$ by multiplying a constant η^β to the detection threshold of $T_\beta(\boldsymbol{x})$. The detection performance only depends on the choice of β, which will be studied in Sec. IV-A. Also, simply specifying $\eta = L$ can avoid the calculation of the term $xy^{\beta-1}$, which reduces the computational cost for calculating the hypothesis testing metric especially when L is large.

IV. Simulation and Experiment Results

In this section, we use simulations and real experiments to evaluate the detection performance of the proposed β-divergence detection (β-DD) method. The classical energy detection (ED) method [1] and the power spectrum entropy detection (PSED) method [8] are utilized as baseline methods. To fairly compare the different hypothesis metrics, the two-stage trick in PSED method is not used and the information entropy is utilized as the metric. Different modulation types of communication signals including BPSK, QPSK, 8PSK, PAM4, CPFSK and 16QAM are considered in the evaluation. The false alarm probability P_F is fixed to be 0.01 and probability of detection is compared. The symbol rate is $R_b = 20$ kHz and the sampling rate is $f_s = 120$ kHz. The during time of each signal instance is 0.5s and the Welch method with 1024-FFT points and commonly used 50% overlap is utilized to estimated the PSD, which is further used in β-DD and PSED methods for calculating the hypothesis metrics. The value of η is set to be 0.5 and the total number of independent Monte Carlo experiments is 5000.

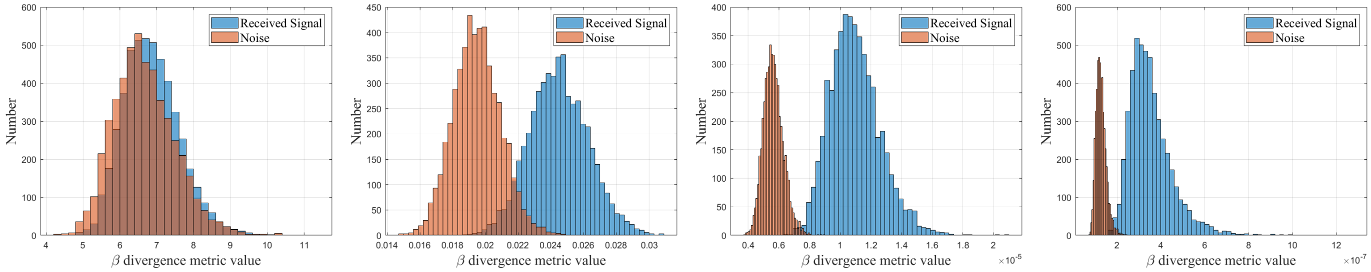
A. Simulation Results

We first evaluate the impact parameter β. The histogram of different hypothesis testing metrics under $\mathcal{H}_0$ and $\mathcal{H}_1$ are compared in Fig. 2, where the SNR is fixed to be -16dB. The modulated signals are generated using the MATLAB Communications Toolbox. The smaller the overlapping area of the received signal and noise histogram, the more beneficial to detection. For different β values, i.e., $\beta = -10, -2, 10, 16$, the size of the overlapping area are different. Particularly, for $P_F = 0.01$, the corresponding P_Ds are 1.10%, 77.70%, 99.60%, and 98.76%, respectively. Therefore, value of β needs to be specified and in the rest part, for a fair comparison, we do not exhaustively tuning β and simply fixed $\beta = 6$ for the simulation and experiment detection.

Next, we change SNR from -20dB to -12dB with 2dB step size. For different modulation signals, the proposed method is compared to ED and PSED and the detection probability curves are shown in the Fig. 3. It can be seen that the ED method is the worst one, the the PSED method has similar performance with β-DD method in high SNR cases, but worse in low SNR cases. As P_F changes, the results show that the P_D of low P_F also decreases, since the threshold is increased. For all considered modulation types, the β-DD method achieves considerable detection gain at low SNR, and its performance is the best.

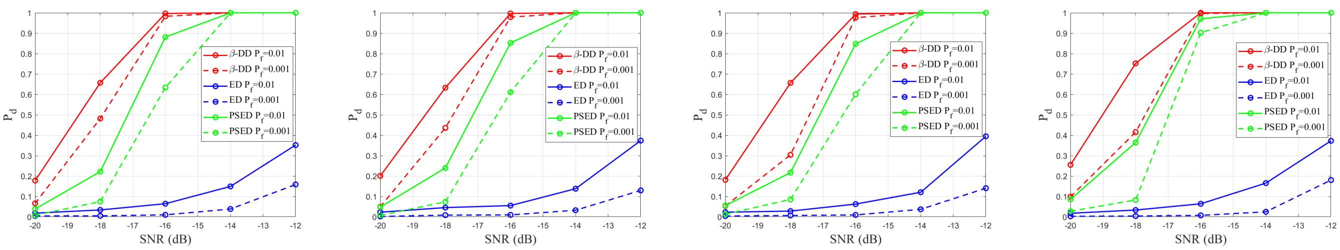
B. Experiment Result

In order to evaluate the performance in practice, we collect the measured noise and signal samples based on the NI USRP X410 software radio platform, see Fig. 4. The transmitter transmits the signal and the receiver receives it. As the received signal is too strong, which is not conducive to performance analysis, we put an attenuator with -30dB at the transmitter. The launch gain is controlled within 6-16dB, the modulation modes include 8PSK, 16QAM, BPSK and QPSK. The false



(a) $\beta = -10$ (b) $\beta = -2$ (c) $\beta = 10$ (d) $\beta = 16$

Fig. 2: Examples of histograms of the received signal and noise under different β. (QPSK signal with -16 dB SNR).



(a) QPSK (b) 16QAM (c) PAM4 (d) CPFSK

Fig. 3: Comparison of probability of detection of different methods.

alarm probability P_F is 0.01, the symbol rate R_b is 20 kHz, the sampling rate f_s is 241kHz, and the number samples of each single instance is 60000. Each experiment is repeated for 600 times. Since the measured data samples contains external

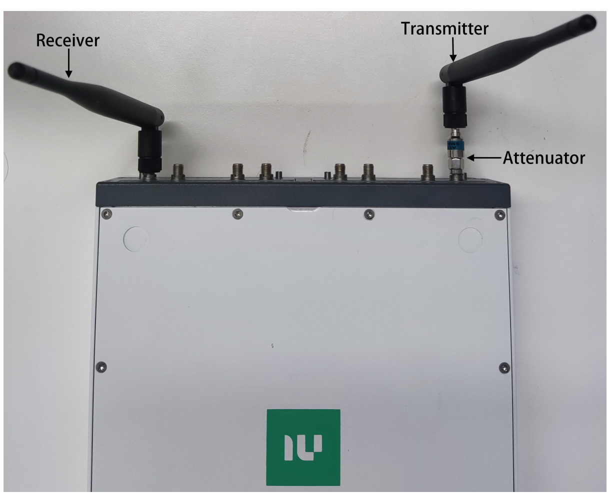
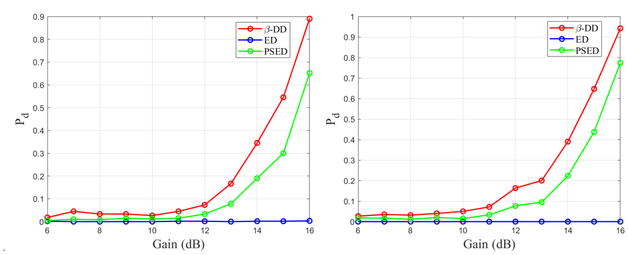


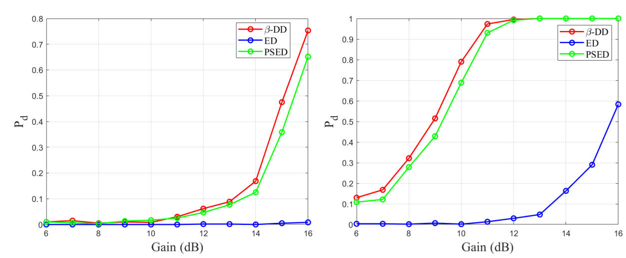
Fig. 4: Illustration of experimental equipment.

interference, the collected data has many spikes, and the signal level fluctuates greatly under the condition of the same gain and SNR, the data processing method mentioned in the third part is not ideal. In order to eliminate the influence of large fluctuation of measured data, we normalize the obtained power spectrum and do the detection.

Fig. 5 shows that the performance of ED method is poor and the signal can hardly be detected. The detection performance of PSED method is not good under the condition of low gain (SNR), while β-DD method is improved under the same gain and has the best performance in most cases.



(a) 8PSK. (b) 16QAM.



(c) BPSK. (d) QPSK.

Fig. 5: Detection performance on experimental data.

V. Conclusions

In this paper, we have revisited information entropy based detection method and shown that it can be regarded as a KL divergence based detection method. To understand the potential performance improvement of using other metrics, β-divergence has been introduced. Both simulation and real experiments have demonstrated the superiority of our proposed β-divergence detector.

References

[1] V. Kostylev, "Energy detection of a signal with random amplitude," in *2002 IEEE International Conference on Communications. Conference Proceedings. ICC 2002*, vol. 3, 2002, pp. 1606–1610 vol.3.

[2] J. Du, H. Huang, X. J. Jing, and X. Chen, "Cyclostationary feature based spectrum sensing via low-rank and sparse decomposition in cognitive radio networks," in *2016 16th International Symposium on Communications and Information Technologies (ISCIT)*, 2016, pp. 615–619.

[3] J. Chen, A. Gibson, and J. Zafar, "Cyclostationary spectrum detection in cognitive radios," in *2008 IET Seminar on Cognitive Radio and Software Defined Radios: Technologies and Techniques*, 2008, pp. 1–5.

[4] Y. Shin, S.-W. Nam, C.-K. An, and E. Powers, "Design of a time-frequency domain matched filter for detection of non-stationary signals," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings*, vol. 6, 2001, pp. 3585–3588 vol.6.

[5] Y. Lu, P. Zhu, D. Wang, and M. Fattouche, "Machine learning techniques with probability vector for cooperative spectrum sensing in cognitive radio networks," in *2016 IEEE Wireless Communications and Networking Conference*, 2016, pp. 1–6.

[6] C. An, D. Zhang, C. Yuan, L. Li, X. Zhao, S. Geng, W. Zheng, and W. Shao, "Spectrum sensing based on knn algorithm for 230 mhz power private networks," in *2018 12th International Symposium on Antennas, Propagation and EM Theory (ISAPE)*, 2018, pp. 1–4.

[7] D. Zhang and X. Zhai, "Svm-based spectrum sensing in cognitive radio," in *2011 7th International Conference on Wireless Communications, Networking and Mobile Computing*, 2011, pp. 1–4.

[8] Z. Hong and Y. Xu, "A novel weak signal detector based on power spectrum entropy under low snr," in *2021 28th International Conference on Telecommunications (ICT)*, 2021, pp. 1–6.

[9] Y. L. Zhang, Q. Y. Zhang, and T. Melodia, "A frequency-domain entropy-based detector for robust spectrum sensing in cognitive radio networks," *IEEE Communications Letters*, vol. 14, no. 6, pp. 533–535, 2010.

[10] S. Haykin, D. J. Thomson, and J. H. Reed, "Spectrum sensing for cognitive radio," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 849–877, 2009.

[11] A. Cichocki and A.-H. Phan, "Fast local algorithms for large scale nonnegative matrix and tensor factorizations," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 92, no. 3, pp. 708–721, 2009.

[12] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, 2009.

[13] W. Pu, S. Ibrahim, X. Fu, and M. Hong, "Stochastic mirror descent for low-rank tensor decomposition under non-euclidean losses," *IEEE Transactions on Signal Processing*, vol. 70, pp. 1803–1818, 2022.