Kernelized Generalized Likelihood Ratio Test for Spectrum Sensing in Cognitive Radio

Lily Li, *Student Member, IEEE*, Shujie Hou, *Student Member, IEEE*, and Adam Lane Anderson, *Senior Member, IEEE*

***Abstract*—Spectrum sensing in next-generation wireless radio networks is considered a key technology to overcome the problem of spectrum scarcity. Unfortunately, many approaches to spectrum sensing do not work well in low signal-to-noise ratio (SNR) environments. This paper proposes and analyzes a new algorithm named kernelized generalized likelihood ratio test (KGLRT) for spectrum sensing in cognitive radio systems to overcome this problem. Effectively, KGLRT uses a nonlinear kernel to map input data onto a high-dimensional feature space; then, the widely accepted (linear) generalized likelihood ratio test is used for hypothesis testing. This new algorithm gives a gain of 4 dB in SNR over its linear counterpart. A theoretical analysis for this algorithm is given for the first time and is shown analogous to algorithms used in image signal processing. The detection metrics are found to be concentrated random variables; furthermore, the probability distributions of the detection metrics are proved to follow the *F*-distributions, which agree with the results obtained using the concentration inequality. The analytical thresholds are derived for target false-alarm probabilities. The thresholds are independent of noise power; thus, the proposed algorithm can overcome noise uncertainty issues at very low SNR levels. Simulations validate the theoretical results.**

***Index Terms*—Spectrum sensing, GLRT, KGLRT.**

I. Introduction

SPECTRUM sensing is a core function required in cognitive radio [1] or any spectral approach to adaptive communications. In general, this sensing is used to detect the presence of the primary signal; however, spectrum sensing is more challenging in an extremely low signal-to-noise ratio (SNR) environment, e.g., -20 dB, which forces researchers to revisit classical detection problems for these modalities. Due to high rate hardware, for even short durations of interest (on the order of one second), we often have at our disposal a large number of samples. For example, with digital television (DTV) signals, it is feasible for there to be 10,000 samples that could fit as input to the available algorithms. To take advantage of this large number of samples in sensing, we formulate the problem as a high-dimensional probability space through the use of kernel functions.

Associated with high dimensional control, the concentration of measure phenomenon [2] means that certain random fluctuations become tractable. With the success of asymptotic methods in random matrix theory, we can exploit the high-dimensionality of data in the feature space. In fact, one essential feature of Big Data research topics is related to high-dimensional statistics. High dimensionality enables us to obtain central limit theorems for some advanced statistics—detection metrics in our case. In other words, the power of our proposed algorithm arises from the high dimensionality of the feature space and the novel use of kernel mappings.

A. Related Works

Different kinds of techniques have been developed for spectrum sensing. Classic methods include energy detection (ED), matched filter detection, cyclostationary feature detection, wavelet-based detection and cooperative detection. See e.g., [1], [3] for details. The line of research relevant to this paper starts with eigenvalue-based detection [4]–[7]. The idea is to form a sample covariance matrix (SCM) from the measured data samples; then, a specific statistic only using eigenvalues of the SCM is used for the detection metric in a hypothesis test. This is a counterpoint to when the eigenvectors of the SCM are used [1], [8], [9].

Using the leading eigenvector of the SCM, a detection algorithm named feature template matching (FTM) is first proposed in [10] and presented with more details in [8]. The kernel version of feature template matching (KFTM) is considered in [9]. Compared with FTM, KFTM deals with high-dimensional data and applies a kernel trick that not only avoids the limitations of linear problems but also decreases calculation complexities, resulting in significantly improved detection performance. Only a leading eigenvector (rank-one matrix approximation) is considered in the nonlinear kernel version [9]. This paper, on the other hand, considers the subspace counterpart (rank-k matrix approximation) of both linear and nonlinear (kernel) versions. As in [11], [12], linear subspace detection methods have been used for spectrum sensing to improve detection performance in low SNR environments.

A critical advancement is made in the context of hyperspectral target detection for pattern recognition in [13], where it is

Manuscript received September 10, 2017; revised January 27, 2018 and March 19, 2018; accepted March 19, 2018. Date of publication April 6, 2018; date of current version August 13, 2018. This work was supported by the National Science Foundation under Grant CNS-1738034. The review of this paper was coordinated by Prof. D. B. da Costa. *(Corresponding author: Adam Lane Anderson.)*

L. Li and S. Hou are with the Department of Electrical and Computer Engineering, Tennessee Technological University, Cookeville, TN 38505 USA (e-mail: lli21@students.tntech.edu; shou42@students.tntech.edu).

A. L. Anderson is with Oak Ridge National Laboratory and the Department of Electrical and Computer Engineering, Tennessee Technological University, Cookeville, TN 38505 USA (e-mail: aanderson@tntech.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2018.2824023

necessary for the GLRTs in the feature space to be kernelized before they can be implemented due to the high dimensionality of the feature space. Their algorithm is derived but no analysis is attempted. The current work helps fill this gap by conducting an analysis for the algorithm used in the context of cognitive radio. The distribution of the decision threshold is derived to satisfy an F-distribution; numerical simulations agree with the derived result. Surprisingly, in unrelated work [14], F-distribution is also used to model composite fading channels accurately.

Other relevant works include: [15] which proposes a kernelized energy detection algorithm (KED) which is much more robust against impulsive noise and displays a considerably better detection performance than the conventional ED. [16] shows how a kernel matrix can asymptotically be well approximated by an analytically tractable random matrix.

B. Contributions

This paper devises a subspace matching method [17] combined with the so-called kernel trick that was first used in the context of images by [13] (which lacked analysis for the algorithm). This paper's main contributions are as follows:

1) Based on [17], [18], a linear version (called linear GLRT here) algorithm has been generated using the linear subspace of the detected primary user. The subspace is built via the well-known Karhunen-Loeve transform (KLT). To the best of our knowledge, this is the first time the extended nonlinear version of GLRT, called KGLRT, is proposed for spectrum sensing with detailed analysis of the proposed algorithms. Through second order statistics of GLRT, higher-order statistics/correlation is investigated. By applying a kernel trick, the detection metric in high-dimensional space greatly simplifies calculations required for the nonlinear mapping. Simulations using measured DTV signals show that, compared with its linear counterpart, our proposed KGLRT yields a 4 dB gain in SNR for a fixed false-alarm rate. This significant gain justifies the approach of exploiting high-dimensionality for spectrum sensing. Simulation comparisons with other spectrum sensing methods [5], [19]–[22] show that KGLRT achieves the best performance.
2) The statistical distributions of proposed GLRT and KGLRT for different kernels are thoroughly analyzed. Under null hypothesis, the closed-form expressions with decision thresholds for different false-alarm probabilities are obtained. Simulations validate these closed-form expressions. Furthermore, the derived thresholds make no use of noise power, which avoids the noise uncertainty problem in spectrum sensing and encourages the proposed method in low SNR environments.
3) The analysis based on Tao's Lemma is accomplished. This Lemma shows the concentration inequality regarding distance between orthogonal projections [23]. Our results demonstrate the advantage of the proposed high-dimensional detection metrics for both linear and nonlinear cases.

This paper is organized as follows: In Section II, the problem model is built upon a binary hypothesis test, and the GLRT algorithm for subspace matching is derived step by step. In Section III, the kernel trick is introduced in detail; and then the KGLRT subspace detection algorithm is proposed as the nonlinear version of GLRT; furthermore, the computational complexities of both GLRT and KGLRT are compared in Section III-D. In Section IV, the statistical distribution of the detection metric is derived and the decision thresholds are obtained. For different spectrum sensing algorithms, comparisons of computational complexities and detection performances are presented in Section V. The paper is summarized and future challenges are proposed in Section VI.

Notation: In the following, we depict variables in lower case letters, vectors in lower case boldface letters and matrices in uppercase boldface letters. The transpose operator is given be $(\cdot)^T$ while $\langle \mathbf{x}, \mathbf{y} \rangle$ would be the inner product of two vectors $\mathbf{x}$ and $\mathbf{y}$, for example.

II. Linear Subspace Detection

A. Linear Subspace Model Statement

Consider a cognitive radio scenario with a pair of users: one designated as "primary" and the other as "secondary". Let $y(t)$ be the continuous-time signal received by the secondary user through some unknown channel; the signal $y(t)$ is sampled with period T_s to create the discrete-time received samples $y[l] = y(lT_s)$.

Spectrum sensing is performed between two hypotheses: the null hypothesis $\mathcal{H}_0$, where the primary user is absent; and the alternative hypothesis $\mathcal{H}_1$, where the primary user is present

$$\begin{aligned} \mathcal{H}_0 &: y[l] = w[l] \\ \mathcal{H}_1 &: y[l] = x[l] + w[l] \end{aligned} \tag{1}$$

where $w[l] \sim \mathcal{N}(0, \sigma^2)$ are samples of an independent and identically distributed $(i.i.d.)$ Gaussian noise source, and $x[l]$ are the received samples including the effects of path loss, channel fading and time dispersion [6]. $x[l]$ can be the superposition of signals from primary users after passing through an unknown channel with unknown distribution. Signal and noise are assumed to be independent of each other.

At some discrete moment in time, i, the sensing segment consisting of n consecutive observations is congregated into a received vector

$$\mathbf{y} = (y([i], y[i+1], ..., y[i+n-1]) \tag{2}$$

Spectrum sensing is performed based on the statistics of $\mathbf{y}$ by the secondary user. Parameters P_d and P_f are used to evaluate detection performance. Our goal is to maximize detection probability P_d at a target false alarm rate P_f.

B. Generalized Likelihood Ratio Test

It is assumed that the primary user's signal $\mathbf{x} \in \mathbb{R}^{n\times 1}$ lies in a K-dimensional subspace which is spanned by the columns of an orthonormal matrix $\mathbf{T} \in \mathbb{R}^{n\times K}$, so $\mathbf{x} = \mathbf{T}\boldsymbol{\theta}$, where $\boldsymbol{\theta} \in \mathbb{R}^{K\times 1}$ is an unknown coefficient vector [18], and $\mathbf{T}$ represents a full

rank matrix whose column vectors span the original signal subspace. Since $\mathbf{T}$ is orthonormal, it has the well-known property that $\mathbf{T}^T\mathbf{T} = \mathbf{I}$, where $\mathbf{I}$ stands for the identity matrix while matrix $\mathbf{T}$ and its transpose $\mathbf{T}^T$ are not commutative. Additive noise is represented by vector $\mathbf{w} \in \mathbb{R}^{n\times 1}$ with $\mathbf{w} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_n)$. We concatenate N successive samples of vectors given by (2). The spectrum sensing detection problem presented in (1) can be formulated as

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{y} = \mathbf{w} \\ \mathcal{H}_1 &: \mathbf{y} = \mathbf{T}\boldsymbol{\theta} + \mathbf{w} \end{aligned} \tag{3}$$

Since cognitive radios have the ability to learn [24], part of the received primary user signals can be saved and utilized to learn the matrix $\mathbf{T}$. $\mathbf{T}$ can be approximated by matrix $\hat{\mathbf{T}} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_K)$, with $\{\mathbf{v}_i, i = 1, .., K\}$ are dominant eigenvectors of the sample covariance matrix for the zero mean or centered samples [13]. $K \leq \min(N, n)$ is the dimension of the subspace which best preserves the primary user's power in the least mean squared error sense.

According to Neyman-Pearson criterion, likelihood ratio test (LRT) which is defined as

$$\rho_{lrt} = \frac{f_1(\mathbf{y}|\mathcal{H}_1)}{f_0(\mathbf{y}|\mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} T_{lrt} \tag{4}$$

is the optimal detection rule where T_{lrt} is the threshold value, $f_1(\mathbf{y}|\mathcal{H}_1)$ and $f_0(\mathbf{y}|\mathcal{H}_0)$ stand for the conditional probability density function of $\mathbf{y}$ under the hypotheses $\mathcal{H}_1$ and $\mathcal{H}_0$. Their distributions can be represented as conditional Gaussian probability densities $\mathcal{N}(\mathbf{T}\boldsymbol{\theta}, {\sigma_1}^2\mathbf{I})$ and $\mathcal{N}(0, {\sigma_0}^2\mathbf{I})$.

However, in general while exploring LRT approach, the parameters $\boldsymbol{\theta}, \sigma_0, \sigma_1$ are unknown. An alternative way is to replace $\boldsymbol{\theta}, \sigma_0, \sigma_1$ by their maximum likelihood estimators $\widehat{\boldsymbol{\theta}}, \widehat{\sigma}_0, \widehat{\sigma}_1$ [17], [18]. The general likelihood ration test (GLRT) can then be obtained in a straightforward manner as

$$\begin{aligned} \widehat{\rho}_{lrt} &= \frac{\max_{\theta} f_1(\mathbf{y}|\mathcal{H}_1)}{\max_{\theta} f_0(\mathbf{y}|\mathcal{H}_0)} \\ &= \left(\frac{\widehat{\sigma}_1^2}{\widehat{\sigma}_0^2}\right)^{-n/2} \exp\left\{-\frac{1}{2\widehat{\sigma}_1^2}\|\widehat{\mathbf{w}_1}\|^2 + \frac{1}{2\widehat{\sigma}_0^2}\|\widehat{\mathbf{w}}_0\|^2\right\} \\ &\quad \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} T_{lrt} \end{aligned} \tag{5}$$

The least squares solutions to the model presented in (3) are produced by projection of the input vector onto a subspace. $\widehat{\mathbf{w}}_i$ are the MLE estimators of $\mathbf{w}_i$ for both the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$.

$$\begin{aligned} \widehat{\mathbf{w}}_0 &= \mathbf{y} \\ \widehat{\mathbf{w}}_1 &= (\mathbf{y} - \mathbf{T}\boldsymbol{\theta}) = (\mathbf{I} - \mathbf{P}_T)\mathbf{y} \end{aligned} \tag{6}$$

where $\mathbf{P}_T = \mathbf{T}(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T = \mathbf{T}\mathbf{T}^T$ is a projection matrix associated with the K dimensional subspace $\mathbf{T}$. We then obtain:

$$\mathbf{P}_T^T = \mathbf{P}_T, \mathbf{P}_T^2 = \mathbf{P}_T \tag{7}$$

Afterwards, the signal energy can be derived as:

$$\begin{aligned} \|\widehat{\mathbf{w}}_1\|^2 &= (\mathbf{y} - \mathbf{P}_T\mathbf{y})^T(\mathbf{y} - \mathbf{P}_T\mathbf{y}) \\ &= \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{P}_T\mathbf{y} - \mathbf{y}^T\mathbf{P}_T^T\mathbf{y} + \mathbf{y}^T\mathbf{P}_T^T\mathbf{P}_T\mathbf{y} \\ &= \mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\mathbf{P}_T\mathbf{y} + \mathbf{y}^T\mathbf{P}_T\mathbf{y} \\ &= \mathbf{y}^T(\mathbf{I} - \mathbf{T}\mathbf{T}^T)\mathbf{y} \end{aligned} \tag{8}$$

We also know that $\|\widehat{\mathbf{w}}_0\|^2 = \mathbf{y}^T\mathbf{y}$. Since $\mathbf{w}_i$ has normal distribution as $N(0, {\sigma_i}^2\mathbf{I})$, therefore

$$\widehat{\sigma}_0^2 = \frac{1}{n}\|\mathbf{y}\|^2 = \frac{1}{n}\|\widehat{\mathbf{w}}_0\|^2 \tag{9}$$

$$\widehat{\sigma}_1^2 = \frac{1}{n}\left\|\mathbf{y} - \mathbf{T}\widehat{\boldsymbol{\theta}}\right\|^2 = \frac{1}{n}\|\widehat{\mathbf{w}}_1\|^2 \tag{10}$$

$$\widehat{\boldsymbol{\theta}} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{y} = \mathbf{T}^T\mathbf{y} \tag{11}$$

Substituting the maximum likelihood estimators $\widehat{\sigma}_0^2$, $\widehat{\sigma}_1^2$ and $\widehat{\boldsymbol{\theta}}$ into (5) and taking $n/2$ root, GLRT rule is

$$\begin{aligned} \rho_{glrt} &= (\widehat{\rho}_{lrt})^{2/n} \\ &= \frac{\|\widehat{\mathbf{w}}_0\|^2}{\|\widehat{\mathbf{w}}_1\|^2} \\ &= \frac{\widehat{\sigma}_0^2}{\widehat{\sigma}_1^2} \\ &= \frac{\mathbf{y}^T\mathbf{y}}{\mathbf{y}^T(\mathbf{I} - \mathbf{T}\mathbf{T}^T)\mathbf{y}} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} T_{glrt} \end{aligned} \tag{12}$$

III. Kernel-Based Subspace Detection

A. Kernel Method

Kernel methods have been generally employed in machine learning where data is implicitly mapped to feature space via the so-called "kernel trick". Let $\mathcal{X}$ be an n-dimensional input space ($\mathcal{X} \in \mathbb{R}^n$). With a mapping φ, $\mathbf{x}$ is mapped to $\varphi(\mathbf{x})$ in a nonlinear feature space $\mathcal{F}$, where $\mathbf{x}$ is an input vector in $\mathcal{X}$ and φ is an implicitly defined mapping. After mapping data via a map φ into the feature space $\mathcal{F}$ [25], the kernel function k is defined as the inner-product of data, i.e.,

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle\varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j)\rangle \tag{13}$$

where $< \cdot >$ stands for the inner product. The feature space $\mathcal{F}$, also named as kernel space, is a space of functions mapping $\mathcal{X}$ into $\mathbb{R}$, denoted as $\mathcal{F} := \{f : \mathcal{X} \to \mathbb{R}\}$. The map φ is defined via

$$\begin{aligned} \varphi &: \mathcal{X} \to \mathcal{F} \\ \mathbf{x} &\mapsto \varphi(\mathbf{x}) \end{aligned} \tag{14}$$

Thus a feature space $\mathcal{F}$ associated with φ can be constructed. Such a space is defined by the selected kernel function. The mapping φ can be so ordinary that various types of data can be operated upon; thus, a large number of affiliations between data can be detected [25]. Typically, projective kernels (functions of inner-product) and radial kernels (functions of distance)

are the two most popular groups of kernels. In this paper, polynomial kernel (a projective kernel) and Gaussian radial basis function (RBF) kernel (a radial kernel) are chosen for use and comparison.

The Gaussian RBF kernel is defined as

$$k(\mathbf{x}_i,\mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i-\mathbf{x}_j\|^2}{2\upsilon^2}\right) \tag{15}$$

where $\upsilon > 0$. The polynomial kernel with order d is defined as

$$k(\mathbf{x}_i,\mathbf{x}_j) = (\langle \mathbf{x}_i,\mathbf{x}_j\rangle + c)^d \tag{16}$$

and $c = 0$ for homogeneous kernels and $c = 1$ for inhomogeneous kernels.

After Gaussian kernel mapping, the dimension of the feature space (kernel space) can be huge or even infinite [26]. For polynomial kernel with order d, the dimension of the feature space, which is the number of different monomials, becomes $(d+P-1)!/d!(P-1)!$; To explicitly avert direct operation on $\varphi(\mathbf{x})$, the kernel function of (13) is selected to exploit only the inner product of data in the feature space. This trick adequately bypasses calculating the large high-dimensional data set generated by different nonlinear mappings. Without knowing precisely what φ stands for, we can develop a kernel-based algorithm in kernel space using only kernel functions. Thus the proposed nonlinear algorithm can not only be attained without causing too much computational load, but also take higher order statistics/ correlation into account.

Accordingly, when data from the original input space exhibits a nonlinear formation, applying the kernel trick leads to less computational load and achieves better performance than linear methods. On the other hand, not all functions defined in the input space can be used as a kernel function. Only when a mapping φ exists and (13) is credible can the kernel function k be well-defined. Equivalently, if a function meets Mercer's condition [27], it is a well-defined kernel.

B. Linear Subspace Models Defined in Kernel Space

Recent works in random matrix theory, as presented in [28], consider the spectrum of certain kernel matrices in high-dimensional space. Under various technical assumptions, kernel random matrices, i.e. $n \times n$ matrices whose (i,j)th entry is $f(\langle \mathbf{x}_i,\mathbf{x}_j\rangle/p)$ (polynomial kernel belongs to this group) or $f(\|\mathbf{x}_i-\mathbf{x}_j\|_2^2/p)$ (Gaussian kernel belongs to this group), where p is the dimension of the data and $\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_n$ are independent vectors, behave essentially like matrices closely connected to sample covariance matrices. Here f is assumed to be a local smooth function and p can approach infinity but is comparable to the size of n, i.e. $p/n \in (0,1)$. As mentioned in [28], in high-dimensional space, the nonlinear methods that rely on kernel metrics may behave like their linear counterparts. Furthermore, for a more specific case, following the procedure in [13], we make the essential assumption that the spectrum sensing in the feature space can be formulated in hypotheses

$$\begin{aligned}\mathcal{H}_{0_\varphi} &: \varphi(\mathbf{y}) = \mathbf{w}_\varphi \\ \mathcal{H}_{1_\varphi} &: \varphi(\mathbf{y}) = \mathbf{T}_\varphi\boldsymbol{\theta}_\varphi + \mathbf{w}_\varphi\end{aligned} \tag{17}$$

where the columns of $\mathbf{T}_\varphi$ are the basis of the primary user's subspace in $\mathcal{F}$, $\boldsymbol{\theta}_\varphi$ are unknown vectors whose entries are coefficients that account for the abundances of the corresponding columns in $\mathbf{T}_\varphi$. Although the assumption of (17) appears very restrictive, practical examples show that this linear model is approximatively valid in the feature space, as confirmed by the superior performance of the algorithm.

We can consider that the linear subspace model (3) in the feature space $\mathcal{F}$ approximately has a corresponding nonlinear subspace model in the original input space, denoted as (17). Therefore, the GLRT of the linear subspace in $\mathcal{F}$ is equivalent to the GLRT of the corresponding nonlinear subspace model in the input space. When the two hypotheses $\mathcal{H}_{0_\varphi}$ and $\mathcal{H}_{1_\varphi}$ in feature space $\mathcal{F}$ still follow Gaussian distributions [13], based on (12), GLRT for $\varphi(y)$ can be transformed into

$$\rho_{kglrt} = \frac{\varphi(\mathbf{y})^T\varphi(\mathbf{y})}{\varphi(\mathbf{y})^T(\mathbf{I}_\varphi - \mathbf{T}_\varphi\mathbf{T}_\varphi{}^T)\varphi(\mathbf{y})} \tag{18}$$

where $\mathbf{I}_\varphi$ is an identity matrix in $\mathcal{F}$. Apparently, (18) cannot be implemented straightaway because after mapping φ the dimension might become infinite. It should be kernelized to obtain an expression in terms of the kernel function k. Hence we name GLRT of the nonlinear subspace of the original input space as KGLRT. KGLRT is the nonlinear version of GLRT in an arbitrary dimension space. Since Gaussian distribution transformed into the feature space exhibits a more complex structure than in input space, KGLRT can accomplish a better detection job than traditional GLRT. As mentioned above, our choice of kernel functions are Gaussian RBF kernel and polynomial kernel.

C. Kernelized Generalized Likelihood Ration Test

The *priori* knowledge of primary user's signal is accessible in specific situations, which can be used for spectrum sensing. As in matched filter detection [7], [29], [30], the receiver needs to know the primary user's signal in advance. For linear feature template matching in [8], [10], feature template is generated by primary user's signal. In [31], *prior* information of signal is considered. The kernel feature template matching method proposed in [9] requires a historical database of signal samples in order to apply the kernel trick. Since a cognitive radio system has the ability to learn [24], the historical information of primary users, such as DTV or 4G signal can be saved and utilized for subspace learning.

Similar to linear GLRT, for the zero mean or centering data $\varphi(\mathbf{x}_i)$, $i = 1,2,...,M$, each column of $\mathbf{T}_\varphi$ is the eigenvector corresponding to the non zero eigenvalue of

$$\mathbf{R}_{\varphi(\mathbf{x})} = \frac{1}{M}\sum_{i=1}^{M}\varphi(\mathbf{x}_i)\varphi(\mathbf{x}_i)^T \tag{19}$$

If matrix $\mathbf{R}_{\varphi(\mathbf{x})}$ has K_f non zero eigenvalues, to avoid heavy computational load of calculating $\mathbf{R}_{\varphi(\mathbf{x})}$, each column of $\mathbf{T}_\varphi$ can be represented as a linear combination of $\varphi(\mathbf{x}_1), \varphi(\mathbf{x}_2), ...\varphi(\mathbf{x}_M)$ [32]. e.g.,

$$\mathbf{v}_i^f = \sum_{j=1}^{M} \boldsymbol{\beta}_i \varphi(\mathbf{x}_j)$$

where $\boldsymbol{\beta}_i = (\beta_{1i}, \beta_{2i}, ..., \beta_{Mi})^T, i = 1, 2, ..., K_f$ are the eigenvectors corresponding to all non-zero eigenvalues of the kernel matrix defined as

$$\mathbf{K} = (k(\mathbf{x}_i, \mathbf{x}_j))_{M\times M} = (\langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j)\rangle)_{M\times M} \tag{20}$$

It is clear that every kernel matrix is a positive semi-definite matrix. Let $\mathbf{B} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_{K_f})$. Hence $\mathbf{T}_\varphi$ can be expressed implicitly as [32]

$$\mathbf{T}_\varphi = (\mathbf{v}_1{}^f, \mathbf{v}_2{}^f, ..., \mathbf{v}_{k_f}{}^f) = (\varphi(\mathbf{x}_1), ..., \varphi(\mathbf{x}_M))\mathbf{B} \tag{21}$$

Accordingly, $\varphi(\mathbf{y})^T \mathbf{T}_\varphi \mathbf{T}_\varphi{}^T \varphi(\mathbf{y})$ can be represented as

$$\varphi(\mathbf{y})^T \mathbf{T}_\varphi \mathbf{T}_\varphi^T \varphi(\mathbf{y}) = \varphi(\mathbf{y})^T (\varphi(\mathbf{x}_1), \varphi(\mathbf{x}_2), ..., \varphi(\mathbf{x}_M))$$

$$(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_{K_f})(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_{K_f})^T \begin{pmatrix} \varphi(\mathbf{x}_1)^T \\ \varphi(\mathbf{x}_2)^T \\ \vdots \\ \varphi(\mathbf{x}_M)^T \end{pmatrix} \varphi(\mathbf{y})$$

$$= (k(\mathbf{y}, \mathbf{x}_1), k(\mathbf{y}, \mathbf{x}_2), ..., k(\mathbf{y}, \mathbf{x}_M)(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_{K_f})$$

$$(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_{K_f})^T \begin{pmatrix} k(\mathbf{y}, \mathbf{x}_1) \\ k(\mathbf{y}, \mathbf{x}_2) \\ \vdots \\ k(\mathbf{y}, \mathbf{x}_M) \end{pmatrix}$$

$$= \mathbf{k}^T \mathbf{B}\mathbf{B}^T \mathbf{k}$$

where $\mathbf{k} = (k(\mathbf{y}, \mathbf{x}_1), k(\mathbf{y}, \mathbf{x}_2), ..., k(\mathbf{y}, \mathbf{x}_M))^T$. Substitute the last equation into (18),

$$\rho_{kglrt} = \frac{k(\mathbf{y}, \mathbf{y})}{k(\mathbf{y}, \mathbf{y}) - \mathbf{k}^T \mathbf{B}\mathbf{B}^T \mathbf{k}} \tag{22}$$

Thus, the KGLRT detection rule is

$$\rho_{kglrt} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} T_{kglrt} \tag{23}$$

In general, we should center the values of $\varphi(\mathbf{x}_i)$, $i = 1, 2, ..., M$ and $\varphi(\mathbf{y}_i)$, $i = 1, 2, ..., N$ first since usually we consider the zero mean case. However, even though the data set of $\varphi(\mathbf{x})$ and $\varphi(\mathbf{y})$ are not centered, the sample covariance matrix still embodies correlation between data sets. Therefore, the above derivation procedures can be used for both centering and non-centering situations.

Algorithm 1: KGLRT – For Centering and Non-Centering.

1: Retrieve training samples of primary user from a given database.
2: Choose appropriate kernel function k. Kernel matrix $\mathbf{K} = (k(\mathbf{x}_i, \mathbf{x}_j))_{M\times M}$ based on the given samples $\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_M$. This matrix is updated when different features are chosen.
3: Use secondary user signal to build the received vectors: $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N$. Calculate $\widetilde{\mathbf{K}} = (k(\mathbf{y}_i, \mathbf{y}_j))_{N\times N}$.
4: Compute the inner product of the received N-dimensional vector $k(\mathbf{y}, \mathbf{y})$.
5: $\mathbf{k} = (k(\mathbf{y}, \mathbf{x}_1), k(\mathbf{y}, \mathbf{x}_2), ..., k(\mathbf{y}, \mathbf{x}_M))$, the kernel vector is computed.
6: **if** centering **then**
7: Calculate the centering matrix $\mathbf{K} = \mathbf{K}_c$.
8: The centered inner product of received vector $k(\mathbf{y}, \mathbf{y}) = k_c(\mathbf{y}, \mathbf{y})$ is calculated.
9: Calculate the cross kernel matrix $\mathbf{K}^t = (k(\mathbf{x}_i, \mathbf{y}_j))_{M\times N}$.
10: The centered kernel vector $\mathbf{k}_c$ of $\mathbf{k}$ is calculated. $\mathbf{k} = \mathbf{k}_c$.
11: **end if**
12: Eigen decomposition of the kernel matrix $\mathbf{K}$ to obtain the eigenvectors $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_{K_f}$ which corresponding to all non zero eigenvalues.
13: Compute the value of ρ_{kglrt} according to (22).
14: Determine a threshold T_{kglrt}.
15: **if** $|\rho_{kglrt}| > T_{kglrt}$ **then**
16: Primary exists.
17: **else**
18: Primary does not exist.
19: **end if**
20: Hence , $P_d = P(|\rho_{kglrt}| > T_{kglrt}|\mathbf{y} = \mathbf{x} + \mathbf{w})$, and $P_f = P(|\rho_{kglrt}| > T_{kglrt}|\mathbf{y} = \mathbf{w})$.

Similarly, the centering of ρ_{ckglrt} in feature space can be cast as

$$\rho_{ckglrt} = \frac{k_c(\mathbf{y}, \mathbf{y})}{k_c(\mathbf{y}, \mathbf{y}) - \mathbf{k}_c{}^T \mathbf{B}\mathbf{B}^T \mathbf{k}_c} \tag{24}$$

detailed derivation see Appendix A

Following the above derivation, the kernel GLRT algorithm is illustrated as in **Algorithm 1**.

D. Complexity Analysis

Either the subspace of linear space or feature space needs to be learned *a priori* and then used as a template. For linear GLRT we do not need to know exactly the PU signal; we only need to know (or learn) the subspace where the signal lies in. For KGLRT, we need storage of historical data to build the kernel matrix.

The main computational load for liner GLRT includes two parts: generate sample covariance matrix and its eigen-decomposition. Similarly, KGLRT algorithm needs to calculate

the kernel matrix and eigen-decomposition of the kernel matrix. Assume we have N segments of observations where each observation is represented by a $n \times 1$ vector. To calculate the sample covariance matrix of a signal, $n(N-1)(n+1)/2$ additions and $nN(n+1)/2$ multiplications are needed, thus the computational order is $O(n^2N)$. To generate the kernel matrix, one needs to compute $N(N+1)$ kernel functions. Either Gaussian or polynomial kernel function operate on two n dimension vectors and have computational order n, thus generating the kernel matrix the computational complexity is of order $O(nN^2)$. With the use of fast component analysis algorithm [33], [34], the computational complexity of eigen-decomposition of sample covariance matrix has order $O(n^2)$, whereas for the kernel matrix the order is $O(N^2)$. Since N is greater than n, the computation complexity of KGLRT is around N/n times its linear version.

IV. Statistics Analysis Under $\mathcal{H}_0$

A. Linear GLRT Case

Overall, for a fixed false-alarm probability P_f, the target of a well-performed detection algorithm is to maximize the detection probability P_d with reasonable computational complexity. Setting the threshold is the core issue for any likelihood ratio testing algorithm. If the closed-form of probability under $\mathcal{H}_0$ can be formulated, the theoretical value of the threshold can be derived for some target P_f, which is a goal of this work.

For simplicity, the analysis, which will be extended to KGLRT, is performed based on the GLRT algorithm. The detection metric in (12) can be written as

$$\begin{aligned}\rho_{glrt} &= \frac{\mathbf{y}^T\mathbf{y}}{\mathbf{y}^T(\mathbf{I}-\mathbf{T}\mathbf{T}^T)\mathbf{y}} \\ &= \frac{\mathbf{y}^T(\mathbf{I}-\mathbf{T}\mathbf{T}^T)\mathbf{y}+\mathbf{y}^T\mathbf{T}\mathbf{T}^T\mathbf{y}}{\mathbf{y}^T(\mathbf{I}-\mathbf{T}\mathbf{T}^T)\mathbf{y}} \\ &= 1+\gamma\cdot\frac{K}{n-K}\end{aligned}$$

where

$$\gamma = \frac{\mathbf{y}^T\mathbf{T}\mathbf{T}^T\mathbf{y}}{\mathbf{y}^T(\mathbf{I}-\mathbf{T}\mathbf{T}^T)\mathbf{y}}\cdot\frac{n-K}{K}$$

and γ is an alternative metric of the GLRT algorithm. We use γ for GLRT detection.

Theorem 1: The probability distribution of γ is $\mathcal{F}$ distribution, $\gamma \sim \mathcal{F}(K, n-K)$ when the covariance matrix of noise is $\sigma^2\mathbf{I}$.

Proof: See Appendix B. ■

Given a threshold T_{glrt} , it yields

$$P_f = P(\gamma > T_{glrt}|\mathcal{H}_0) = 1 - F(T_{glrt}; K, n-K) \quad (25)$$

where $F(\cdot; K, n-K)$ is the cumulative distribution function (CDF) of F distribution with parameters $K, n-K$. Thus for a target P_f

$$T_{glrt} = F^{-1}(1-P_f; K, n-k) \quad (26)$$

while $F^{-1}(\cdot, K, n-K)$ being the inverse function of $F(\cdot, K, n-K)$. (26) shows that the threshold is noise power independent. Intuitively, the noise power is eliminated since the detection metric (35) is in a ratio form.

B. Statistics Analysis in Kernelized GLRT

Theorem 1 can be extended to a nonlinear case in order to find the threshold for the KGLRT algorithm. From (35), we obtain

$$\rho_{kglrt} = 1 + \gamma_\varphi \frac{K_f}{n_f - K_f}$$

where

$$\gamma_\varphi = \frac{\mathbf{y}_\varphi^T\mathbf{T}_\varphi\mathbf{T}_\varphi^T\mathbf{y}_\varphi}{\mathbf{y}_\varphi^T(\mathbf{I}_\varphi - \mathbf{T}_\varphi\mathbf{T}_\varphi^T)\mathbf{y}_\varphi}\cdot\frac{n_f-K_f}{K_f}$$

where K_f is the dimension of $\mathbf{T}_\varphi$ and n_f is the dimension of $\varphi(\mathbf{y})$. For a kernel function k_φ, if it is smooth and continuous as well as data being band-limited [13], the topographic ordering of the data in the input space is also preserved in the feature space after the nonlinear mapping. Hence we can assume the Gaussian distribution of the two hypotheses $\mathcal{H}_{0_\varphi}$ and $\mathcal{H}_{1_\varphi}$ is valid, which can be found in [26], [35]. The model presented by (18) becomes a linear model in feature space. We use γ_φ as the detection metric. Similarly, as for GLRT, we define

$$\mathbf{P}_{T_\varphi} = \mathbf{T}_\varphi\mathbf{T}_\varphi^T, \mathbf{P}_{T_\varphi^\perp} = \mathbf{I}_\varphi - \mathbf{T}_\varphi\mathbf{T}_\varphi^T$$

It is straight forward that $\mathbf{P}_{\mathbf{T}_\varphi}$ and $\mathbf{P}_{\mathbf{T}_\varphi^\perp}$ both are orthogonal projections, following the steps in the proof of *Theorem 1*, we have $\gamma_\varphi \sim \mathcal{F}(K_f, n_f - K_f)$.

For convenience, let d_1 be equal to K_f and d_2 be equal to $n_f - K_f$. The distribution of γ_φ for null hypotheses for different kernels are discussed below:

1) *Gaussian kernel:* In theory, the dimension of feature space generated by Gaussian kernel (denoted as n_f) is infinite [26], thus d_2 can not be calculated. However, only a small compact subspace of $\mathcal{F}$ is occupied by mapped data [13], thus the dimension used is actually finite. From simulations an unbiased estimator of d_2 can be obtained. Since $\mathbf{T}_\varphi$ can be estimated, we have $d_1 = K_f$. After empirical simulations to a great extent, $\gamma_\varphi \sim \mathcal{F}(d_1, \hat{d}_2)$ where $\hat{d}_2$ is an estimator of d_2.
   Let $E(\cdot)$ denote expectation and $\mu = E(\rho_{ckglrt} - 1)$, then

$$E(\gamma_\varphi) = \frac{d_2}{K_f}\cdot E(\rho_{ckglrt} - 1) = \mu\cdot\frac{d_2}{K_f}$$

   μ can be estimated from simulation samples by

$$\hat{\mu} = \frac{1}{SIM}\sum_{i=1}^{SIM}(\rho_{ckglrt} - 1)$$

   where ρ_{ckglrt} is the sample from the ith simulation and SIM is used to denote the total number of Monte-Carlo simulations. With the assumption that $\gamma_\varphi \sim \mathcal{F}(d_1, \hat{d}_2)$, we know $E(\gamma_\varphi) = \frac{\hat{d}_2}{\hat{d}_2 - 2}$, for $d_2 \geq 2$. Thus, the estimation of

d_2 can be derived as

$$\hat{d}_2 = \frac{K_f}{\hat{\mu}} + 2$$

Hence the analytical threshold for a given P_f can be derived as

$$T_{ckglrt} = \frac{K_f + 2\hat{\mu}}{K_f} \tag{27}$$

2) *Polynomial kernel:* Theoretically analyzing the distribution of γ_φ under $\mathcal{H}_0$ is difficult. Alternatively, through numerous simulations, the distribution can be approximately considered as $\gamma_\varphi \sim \mathcal{F}(\hat{d}_1, \hat{d}_2)$ for second and third order polynomial kernels, where $\hat{d}_1$ and $\hat{d}_2$ are estimators of d_1 and d_2.

Let var denote variance and $S^2 = var(\rho_{ckglrt} - 1)$, then

$$\mu \cdot \frac{d_2}{d-1} = E(\gamma_\varphi) = \frac{d_2}{d_2 - 2}$$

$$S^2 \cdot (\frac{d_1}{d_2})^2 = var(\gamma_\varphi) = \frac{2{d_2}^2(d_1 + d_2 - 2)}{d_1(d_2-2)^2(d_2-4)}$$

The estimators of d_1 and d_2 can be derived as

$$\hat{d}_1 = \hat{\mu}(\hat{d}_2 - 2), \quad \hat{d}_2 = 4 + \frac{2(\hat{\mu}+1)\hat{\mu}}{\hat{S}^2}$$

where

$$\hat{S}^2 = \frac{1}{SIM - 1} \sum_{i=1}^{SIM} \left((\rho_{ckglrt} - 1)^2 - \hat{\mu}\right)^2 \tag{28}$$

is an unbiased estimator of S^2. Therefore, after applying second or third order polynomial kernel, the threshold for a target false alarm probability is

$$T_{ckglrt} = 1 + \frac{\hat{S}^2}{\hat{S}^2 + \hat{\mu}(\hat{\mu}+1)} \tag{29}$$

C. Concentration Inequalities of Generalized Likelihood Ratio Test

The concentration inequalities of statistics for the proposed algorithm will be analyzed in this section. Based on the well-known Tao's Lemma presented as below:

Lemma 1. [36]: Let $X = (\xi_1, ..., \xi_n) \in C^n$ be a random vector whose entries are independent with zero mean, variance 1, and are bounded in magnitude by K almost surely for some $K \geq 10(E\|\xi^4\| + 1)$. Let H be a subspace of dimension d and π_H be the orthogonal projection onto H. Then $P\left(\left|\|\pi_H(X)\| - \sqrt{d}\right| \geq t\right) \leq 10\exp\left(-\frac{t^2}{10K^2}\right)$.

In particular, one has

$$\|\pi_H(X)\| = \sqrt{d} + O(Klogn)$$

with overwhelming probability.

The same conclusion holds (with 10 being replaced by another explicit constant) if one of the entries ξ_i of X is assumed to have variance c instead of 1, for some absolutely constant $c > 0$.

It is straightforward to show that $E\|\pi_H(X)\|^2 = d$ [23]. We study the linear case first. Since any Gaussian signal can be centered and normalized, we can assume entries of vector $\mathbf{y}$ in (35) are mean 0, variance 1 independent Gaussian variables. [37] shows for sub-Gaussian case the magnitudes of entries are bounded almost surely, then this result also works for Gaussian case, while all Gaussian variables belong to sub-Gaussian family. Hence the conditions in *Lemma 1* are satisfied. Projection $\mathbf{P}_T$ is the orthogonal projection of $\mathbf{y}$ onto subspace $\mathbf{T}$, and $\mathbf{P}_{T^\perp}$ is the orthogonal projection of $\mathbf{y}$ onto orthogonal complement of $\mathbf{T}$. From (35) we can derive expectation of the detection metric γ as below:

$$\begin{aligned} E(\gamma) &= E\left(\frac{\mathbf{y}^T\mathbf{P}_T^T\mathbf{P}_T\mathbf{y}}{\mathbf{y}^T\mathbf{P}_{T^\perp}^T\mathbf{P}_{T^\perp}\mathbf{y}}\right) \cdot \frac{n-K}{K} \\ &= \frac{E\left(\mathbf{y}^T\mathbf{P}_T^T\mathbf{P}_T\mathbf{y}\right)}{E\left(\mathbf{y}^T\mathbf{P}_{T^\perp}^T\mathbf{P}_{T^\perp}\mathbf{y}\right)} \cdot \frac{n-K}{K} \\ &= \frac{E\|\mathbf{P}_T\mathbf{y}\|^2}{E\|\mathbf{P}_{T^\perp}\mathbf{y}\|^2} \cdot \frac{n-K}{K} \\ &= \frac{K}{n-K} \cdot \frac{n-K}{K} \\ &= 1 \end{aligned} \tag{30}$$

The second equality comes from the fact that the numerator and denominator of γ are independent (proved in Appendix B). From the above derivation, we know that $\mathbf{E}(\gamma) = 1$ with overwhelming probability. This result agrees with the analysis that $\gamma \sim \mathcal{F}(d_1, d_2)$, $\mathbf{E}(\gamma) = \frac{d_2}{d_2-2} \approx 1$ while d_2 is large enough. Equivalently, this conclusion is valid only in high-dimensional space. As we mentioned before, a nonlinear model in an original linear space becomes a linear model if it is presented in a feature space. When vector $\mathbf{y}$ in linear space satisfies Gaussian distribution, as well as in feature space $\mathbf{y}_\varphi$ itself can be approximately considered as Gaussian distribution too [13]. Thus the result in (30) can be extended to Gaussian kernel GLRT case.

V. Numerical Results

In this section, we first give a brief review for other spectrum sensing algorithms from theoretical perspective. Then give simulation results for different algorithms using a digital TV (DTV) signal, which was captured (field measurements) in Washington DC [38].

A. Comparison With Other Algorithms

In the following, we will discuss other algorithms for comparison purpose. Energy detection [19] is a commonly employed method for spectrum sensing and will be used as a benchmark to the proposed detection algorithms. We compare GLRT, KGLRT with ED and four other algorithms, Arithmetic-to-geometric (AGM) [20] and Maximum-minimum (MME) [5] are eigenvalues based detection techniques, while Cooperative Power

spectral density Cancellation (CPSC) [21] and Circular-folding CPSC (CF-CPSC) [22] are FFT based methods.

1) ED: ED is easily implemented and optimal when signals are uncorrelated. However, it suffers from noise uncertainty since its threshold is noise power dependent and detection requires accurate knowledge of the noise power.
2) AGM: AGM finds an unstructured estimation of the sample covariance matrix $\mathbf{R}_y$ can be represented as $\mathbf{R}_x + \sigma_n^2$ under $\mathcal{H}_1$ and σ_n^2 under $\mathcal{H}_0$. The dimension of $\mathbf{R}_y$ is M and $\lambda_i, i = 1, 2, \ldots, M$ are ordered eigenvalues of $\mathbf{R}_y$. The arithmetic mean and geometric mean of all eigenvalues can be calculated, compare their ratio with the threshold,

$$T_{AGM} = \frac{\frac{1}{M}\sum_i \lambda_i}{(\prod_i \lambda_i)^{\frac{1}{M}}} > \gamma_{AGM}$$

AGM is able to sense the spectrum without prior knowledge and overcome the noise uncertainty.
3) MME: MME is also a eigenvalue-based detection method, it shows the ratio of the maximum eigenvalue and minimum eigenvalue of the sample covariance matrix $\mathbf{R}_y$ can be used to detect the signal existence. By using results from random matrix theory, the ratio can be approximated and the threshold is found. MME can be used without prior knowledge of the signal and noise but has difficult to obtain a precisely closed form formula for the threshold. [39], [40] have made progress but not resolved yet.
4) CPSC: Cooperative Power spectral density Split Cancellation (CPSC) method takes advantages of the asymptotic normality and independence of Fourier transform, thus the stochastic properties of the power spectral density can be achieved. CPSC algorithm divides the power spectrum into L subbands, calculates the ratio of the each subband to the full band, after comparing with the threshold, the decision for each subband can be made. Making the final decision based on the results from all the subbands. The test statistics is expressed as

$$r_{avg} = \frac{1}{U}\sum_{u=1}^{U} \frac{F_{l,u}}{\sum_{l=1}^{L} F_{l,u}}$$

There exist U second users and $F_{l,u}$ is the spectrum power of the subband l of each second user u. The threshold is not related to the noise power.
5) CF-CPSC: Based on CPSC, [22] proposes a circular folding CPSC (CF-CPSC). The fundamental difference between CF-CPSC and CPSC is that CF-CPSC applied a circular-even component [41] that compute $F_u(k)$ as

$$F_u(k) = \begin{cases} \dfrac{F_u^{'}(1) + F_u^{'}(M/2+1)}{2}, & k = 1 \\ \dfrac{F_u^{'}(k) + F_u^{'}(M-k+2)}{2}, & k = 2, 3, ..., M \end{cases}$$

where $F_u{}'(k)$ is the power spectral density and M is the total size of the sampling set. The cumulative distribution functions of the main random variables that form the decision statistic is derived, yields a impressively precise closed-form expression of the threshold.
6) Complexity Comparison: For N segments of data vectors with dimension n, N is larger than n, the calculation complexity for ED is the simplest as it only needs nN multiplications and $(nN-1)$ additions. Linear GLRT, MME and AGM have the same order since they all need to calculate the sample covariance matrix (order $O(n^2N)$) and do eigen-decomposition (order $O(n^2)$). The main computational load of KGLRT is generating the kernel matrix (order $O(nN^2)$) and the kernel matrix decomposition (order $O(N^2)$). Both CPSC and CF-CPSC need Nn-point FFTs, which have order $O(nN\log(Nn))$. The comparison of these algorithms is illustrated in Table I.

TABLE I
COMPLEXITY COMPARISON OF DIFFERENT ALGORITHMS

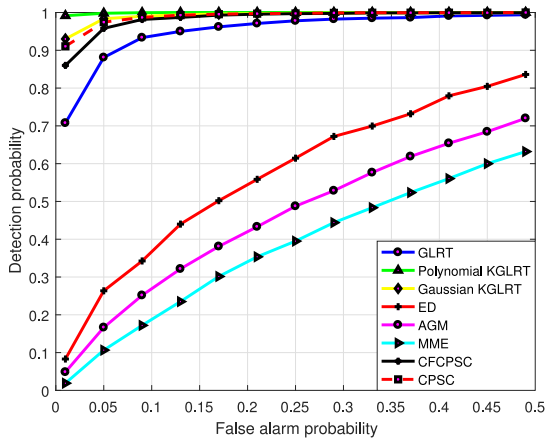
Algorithm	Complexity Order
ED	$O(nN)$
AGM	$O(n^2N)$
MME	$O(n^2N)$
GLRT	$O(n^2N)$
KGLRT	$O(nN^2)$
CPSC	$O(nN\log(nN))$
CF-CPSC	$O(nN\log(nN))$

B. Simulation Results

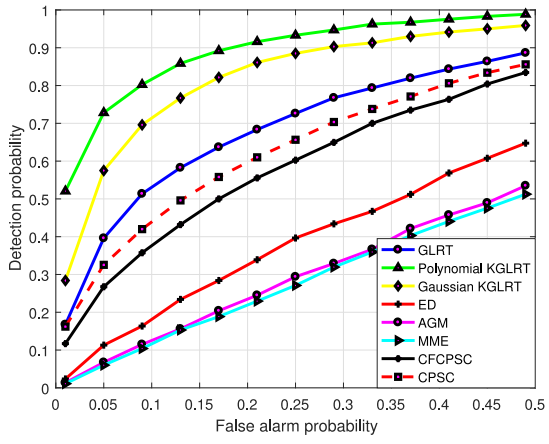
For simulations using DTV signal, field measurements were recorded in Washington D.C. from [38]. The data rate of the vestigial sideband (VSB) DTV signal is 10.76 MSps. The captured DTV signals were sampled at a receiver using 21.52 MSps and down modified to a low central intermediate frequency (IF) of 5.38 MHZ. The SNR level of receiver signal and the communication channel between the transmitter and receiver are unknown. White Gaussian noise is added to obtain various SNR levels for simulation purposes. The performance of the proposed nonlinear algorithm (KGLRT), its linear version (GLRT), ideal ED, AGM, MME, CPSC and CF-CPSC are provided. The process of choosing an appropriate kernel is still an open problem as in almost every kernel-based algorithm [26]. The most widely used - polynomial kernel and the Gaussian RBF kernel - are chosen in this paper to demonstrate advances of the proposed approaches to spectrum sensing. Parameters are empirically chosen for different kernel functions.

1) Detection Performance on Proposed Algorithms: A size of $2e4$ data set is generated and is formed into a 400 by 50 matrix. The signal is corrupted by white Gaussian noise. 1000 Monte Carlo simulations are performed for centering KGLRT in **Algorithm 1**, linear GLRT, ED, MME, AGM, MME, CPSC and CF-CPSC. The false-alarm is preset as 1% and 10%. The proposed nonlinear method performs best among all the methods and outperform its linear version at around 4 dB gain.

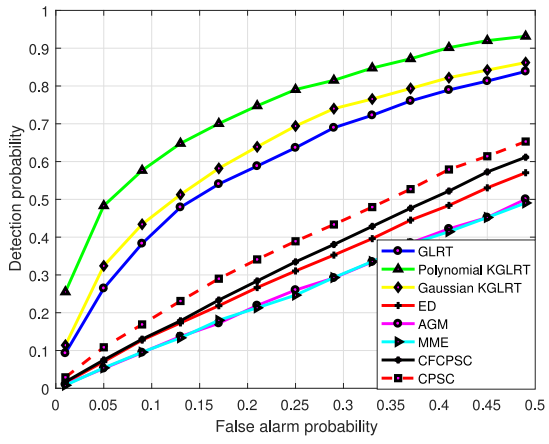
The receiver operating characteristic (ROC) curves generated by different algorithms is illustrated in Fig. 1 at SNR level of -12 dB, -16 dB and -20 dB using DTV signal. Fig. 2 shows the plots of P_d vs SNR at $P_f = 1\%$ and P_d vs SNR at $P_f = 10\%$.



(a)



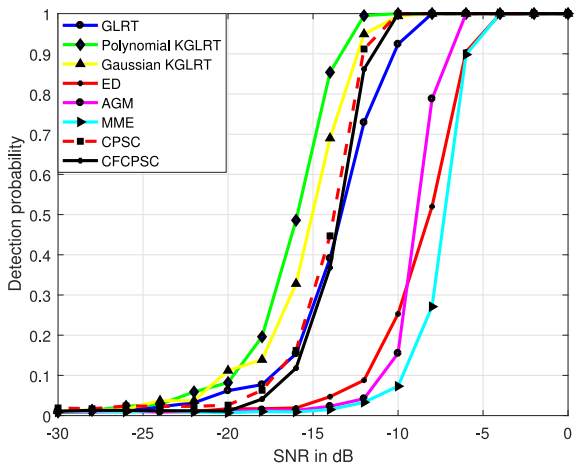
(b)



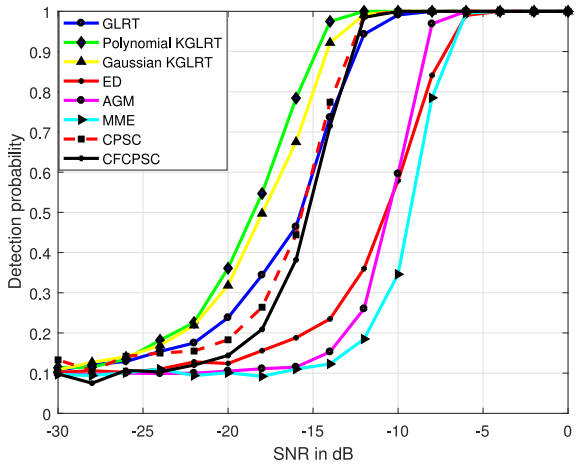
(c)

Fig. 1. ROC curves at different SNR levels, considering signals with unknown distribution through unknown channel; the subspace of signals learned as a priori knowledge. (a) SNR equals -12 dB. (b) SNR equals -16 dB. (c) SNR equals -20 dB.

The thresholds are set up numerically to satisfy the target false alarm probabilities. Polynomial kernel with $d = 2$, $c = 1$ and Gaussian kernel with $\upsilon^2 = 40^2/2$ are exploited. It can be seen that KGLRT, CPSC, CF-CPSC and GLRT all outperform ideal ED, AGM and MME. Ideal ED is comparable to AGM and



(a)



(b)

Fig. 2. Detection probabilities at different preset false alarm probabilities, demonstrating Polynomial KGLRT and Gaussian KGLRT outperform other algorithms. (a) False alarm probability equals 1%. (b) False alarm probability equals 10%.

MME since noise power is assumed be accurately known. Performances of CPSC and CF-CPSC are between GLRT and Gaussian GLRT. However, from Fig. 1 we find that the performance of both CPSC and CF-CPSC drop rapidly when SNR gets lower and perform worse than linear GLRT.

In cognitive radio, P_f relates to spectral utilization since false detections lead to timid secondary users. Fig. 3 shows P_f vs. SNR at $P_d = 90\%$. The thresholds are set numerically to satisfy the target detection probability. The analytical P_f derived in Section IV are also plotted for comparison. Analytical results match the numerical results from simulations. Different KGLRT approaches outperform GLRT approaches as expected. Furthermore, polynomial KGLRT can achieve the lowest false-alarm probability among the three proposed methods.

2) Statistics Under $\mathcal{H}_0$: Simulation results verify the theoretical analysis for PDF and thresholds. The signal is corrupted by White Gaussian noise at different SNR levels. For SNR levels of -8, -12, -16 and -20 dB, one thousand Monte Carlo simulations are done for GLRT, centering Gaussian kernel GLRT and centering polynomial kernel GLRT. The total size of re-

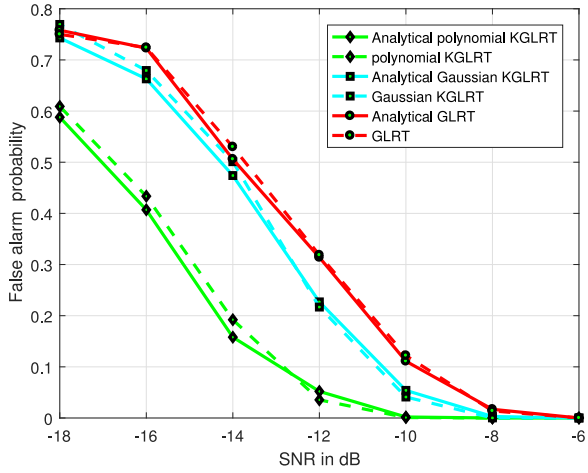


Fig. 3. For a preset target detection probability at 90%, simulated and analytical false alarm probabilities for GLRT, Gaussian KGLRT, and Polynomial KGLRT, showing Polynomial KGLRT performs the best, the second is Gaussian KGLRT, and then GLRT.

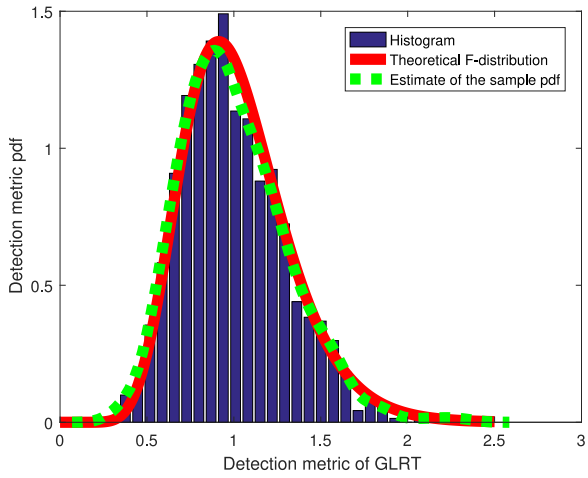


Fig. 4. Comparison of histogram, simulated and theoretical pdfs of linear GLRT detection metric ρ_{glrt} under $\mathcal{H}_0$.

ceived data set is 10^4. The two selected kernels are second-order non homogeneous polynomial kernel and Gaussian kernel with $\hat{\upsilon}^2 = 40^2/2$.

The theoretical results of $\mathcal{F}(K, n-K)$ and $\mathcal{F}(K_f, n_f - K_f)$ are generated for comparison in Figs. 4 and 5. Accordingly, the histograms and pdf approximations for ρ_{glrt} and ρ_{ckglrt} of Gaussian kernel are plotted too. The pdf plots are generated from simulation samplings using Matlab function $ksdensity$. The theoretical and simulated results exhibit similar distributions as expected.

Figs. 6–8 corresponds to GLRT, Gaussian RBF kernel GLRT, and polynomial kernel GLRT at different SNR level. These plots show that theoretical and simulated thresholds are extraordinarily similar to each other, confirm the correctness of analytical results of this paper [see (26), (27) and (29)].

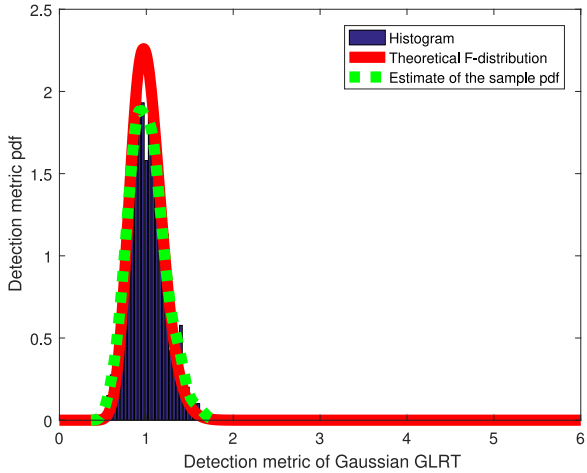


Fig. 5. Comparison of histogram, simulated and theoretical pdfs of Gaussian KGLRT detection metric ρ_{ckglrt} under $\mathcal{H}_0$.

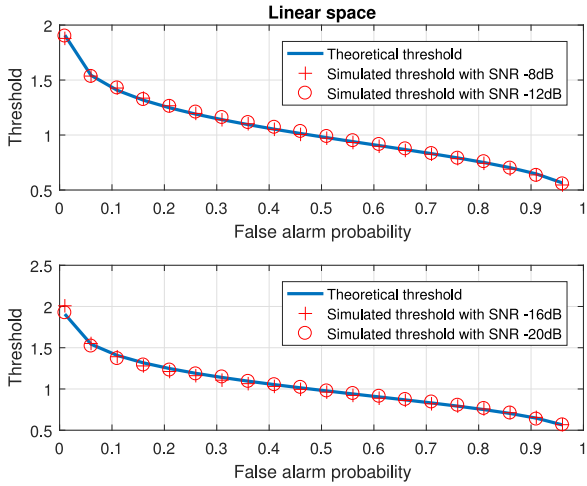


Fig. 6. Comparison of theoretical and simulated thresholds for GLRT, agree with each other at different SNR levels.

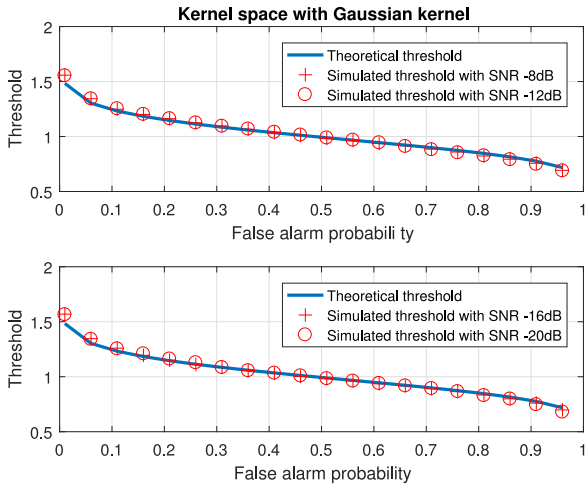


Fig. 7. Comparison of theoretical and simulated thresholds for Gaussian KGLRT, agree with each other at different SNR levels.

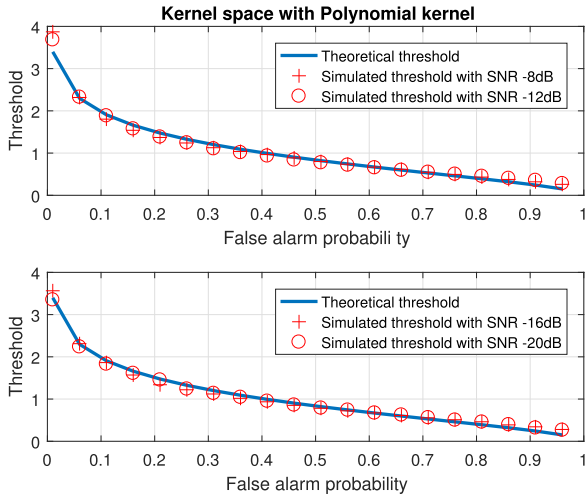


Fig. 8. Comparison of theoretical and simulated thresholds for Polynomial KGLRT, agree with each other at different SNR levels.

VI. Conclusion

Through effectively applying the kernel trick, for the first time a nonlinear algorithm is introduced for spectrum sensing by taking advantage of high dimensionality associated with a nonlinear mapping. Data in high-dimensional feature space exhibits more complex structure which helps to improve detection performance. However, the price of such an improved performance is that computational complexity may rise to extreme heights. To avoid this fatal disadvantage, kernel trick is the solution to bypass direct computing in feature space. After performing mathematical calculations of inner products to find the detection metric, our proposed algorithm achieves impressive performance progress with an acceptable computational load.

Detailed derivations of theoretical statements are given for the proposed algorithm (i.e., KGLRT), which are also valid for its linear counterpart (i.e., GLRT). From Tao's Lemma of distance between orthogonal projections, the concentration property of the analytical detection metric is revealed and the detection metric is found to be a concentrated random variable that gives guidance to set a threshold. The probability distributions for both GLRT and KGLRT algorithms under null hypothesis are formulated in closed-form. Therefore, given a target false alarm rate, the closed-form of detection threshold can be obtained. Closed-form formulations are verified by simulations. Analytical formulation of detection threshold clearly shows that threshold is irrelevant to noise power, which addresses the challenging issue of noise uncertainty that most of the traditional methods such as ED cannot handle.

From this papers work there are abundant future research directions. For example, we can apply the proposed algorithms in cognitive radio vehicular ad hoc networks (CR-VANETs) for reliable spectrum sensing. Different vehicles share the spectrum in an intelligent way. Spectrum sensing is the key step for spectrum sharing and spectrum management. We can also extend our work from the setting of single primary user and single secondary user to the setting of multiple primary users and secondary users with consideration of multiple spectrum bands. Cooperative spectrum sensing exploits spatial and temporal diversities for fast and accurate detection of primary users. When the signals from primary users are extremely weak compared with highway ambient noises (a low-SNR environment), it is necessary to optimize/derive the detection performance and/or to select the detection threshold for any secondary user over any spectrum band efficiently. This task can increase safety, comfort and information agility for vehicles on the road [42]–[44]. Our work also might be applied to other applications such as a real-time Smart Grid system where the smart meters are modeled as secondary users [45].

Appendix A
Derivation for Centered Threshold

We can center data in feature space $\mathcal{F}$ as

$$\varphi_c(\mathbf{x}_i) = \varphi(\mathbf{x}_i) - \frac{1}{M}\sum_{i=1}^{M}\varphi(\mathbf{x}_i), i = 1, 2, ..., M$$

$$\varphi_c(\mathbf{y}_i) = \varphi(\mathbf{y}_i) - \frac{1}{N}\sum_{i=1}^{N}\varphi(\mathbf{y}_i), i = 1, 2, ..., N$$

Consequently, the centering of ρ_{ckglrt} in feature space can be cast as

$$\rho_{ckglrt} = \frac{k_c(\mathbf{y},\mathbf{y})}{k_c(\mathbf{y},\mathbf{y}) - \mathbf{k}_c{}^T\mathbf{B}\mathbf{B}^T\mathbf{k}_c} \tag{31}$$

where

$$\begin{aligned} k_c(\mathbf{y},\mathbf{y}) &= \varphi_c(\mathbf{y})^T\varphi_c(\mathbf{y}) \\ &= \left(\varphi(\mathbf{y}) - \frac{1}{N}\sum_{i=1}^{N}\varphi(\mathbf{y}_i)\right)^T\left(\varphi(\mathbf{y}) - \frac{1}{N}\sum_{i=1}^{N}\varphi(\mathbf{y}_i)\right) \\ &= \varphi(\mathbf{y})^T\varphi(\mathbf{y}) - \frac{1}{N}\sum_{i=1}^{N}\varphi(\mathbf{y})^T\varphi(\mathbf{y}_i) \\ &\quad - \frac{1}{N}\sum_{i=1}^{N}\varphi(\mathbf{y})\varphi(\mathbf{y}_i)^T + \frac{1}{N^2}\sum_{i=1}^{M}\varphi(\mathbf{y}_i)^T\sum_{i=1}^{N}\varphi(\mathbf{y}_i) \\ &= k(\mathbf{y},\mathbf{y}) - 2(k(\mathbf{y},\mathbf{y}_1),...,k(\mathbf{y},\mathbf{y}_N))\mathbf{u} + \mathbf{u}^T\widetilde{\mathbf{K}}\mathbf{u} \end{aligned} \tag{32}$$

and

$$\mathbf{k}_c = \mathbf{k} - \mathbf{K}^t\mathbf{u}$$

$\mathbf{u} = (1/N, 1/N, ..., 1/N)^T$ is an N-dimension unit column vector, and $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N$ are i.i.d copies of $\mathbf{y}$. Matrix $\widetilde{\mathbf{K}} = (k(\mathbf{y}_i, \mathbf{y}_j))_{N\times N}$ is a positive semi-definite matrix, and $\mathbf{K}^t = (k(\mathbf{x}_i, \mathbf{y}_j))_{M\times N}$ is the cross kernel matrix of $\mathbf{x}$ and $\mathbf{y}$

Appendix B
Proof of Theorem 1

Proof: As mentioned before, $\mathbf{P}_T = \mathbf{T}(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^\mathbf{T} = \mathbf{T}\mathbf{T}^T$ is a projection matrix associated with the K dimensional

subspace $\mathbf{T}$. It is easy to derive that

$$\mathbf{P}_T = \mathbf{P}_T^T, \mathbf{P}_T = \mathbf{P}_T^2. \tag{33}$$

Define another projection $\mathbf{P}_{T^\perp} = \mathbf{I} - \mathbf{T}\mathbf{T}^T$, clearly $\mathbf{P}_{T^\perp}$ is associated with the space $\mathbf{T}^\perp$; we call it the orthogonal complement of $\mathbf{T}$

$$\mathbf{T}^\perp = v \in \mathbf{V} : v^{'}t = 0, \forall t \in \mathbf{T}$$

The dimension of $\mathbf{T}^\perp$ is $n - K$ and $\mathbf{P}_{T^\perp} = \mathbf{T}^\perp \mathbf{T}^{\perp T}$. The following properties hold

$$\mathbf{P}_{T^\perp} = \mathbf{P}_{T^\perp}^T, \mathbf{P}_{T^\perp} = \mathbf{P}_{T^\perp}^2 \tag{34}$$

hence, $\mathbf{P}_T$ is the orthogonal projection matrix that projects vectors onto $\mathbf{T}$ and $\mathbf{P}_{T^\perp}$ is the orthogonal projection that projects vectors to subspace $\mathbf{T}^\perp$. This yields,

$$\gamma = \frac{\mathbf{y}^T \mathbf{P}_T^T \mathbf{P}_T \mathbf{y}}{\mathbf{y}^T \mathbf{P}_{T^\perp}^T \mathbf{P}_{T^\perp} \mathbf{y}} \cdot \frac{n-K}{K} \tag{35}$$

Under the null hypothesis $\mathbf{y} \sim N(0, \sigma^2 \mathbf{P}_T)$, hence we have

$$\mathbf{P}_T \mathbf{y} \sim \mathcal{N}(0, \sigma^2 \mathbf{P}_T \mathbf{P}_T^T) = \mathcal{N}(0, \sigma^2 \mathbf{P}_T) \tag{36}$$

and

$$\mathbf{P}_{T^\perp} \mathbf{y} \sim \mathcal{N}(0, \sigma^2 \mathbf{P}_{T^\perp} \mathbf{P}_{T^\perp}^T) = \mathcal{N}(0, \sigma^2 \mathbf{P}_{T^\perp}) \tag{37}$$

It can be easily seen that $\mathbf{P}_T \mathbf{y}$ and $\mathbf{P}_{T^\perp} \mathbf{y}$ are independent of each other, and their inner products $\mathbf{y}^T \mathbf{P}_T^T \mathbf{P}_T \mathbf{y}$ and $\mathbf{y}^T \mathbf{P}_{T^\perp}^T \mathbf{P}_{T^\perp} \mathbf{y}$ are also independent.

On the other hand,

$$(\mathbf{v}_1, ..., \mathbf{v}_K)^T \mathbf{y} = \mathbf{T}^T \mathbf{y} \sim \mathcal{N}(0, \sigma^2 \mathbf{T}^T \mathbf{T}) = \mathcal{N}(0, \sigma^2) \tag{38}$$

So it is straight forward that $\mathbf{v}_i \mathbf{y}$ is independent with $\mathbf{v}_j \mathbf{y}$ while $i \neq j$. Thus

$$\frac{1}{\sigma^2} \mathbf{y}^T \mathbf{P}_T^T \mathbf{P}_T \mathbf{y} = \frac{1}{\sigma^2} \mathbf{y}^T \mathbf{T}\mathbf{T}^T \mathbf{y} = \frac{1}{\sigma^2} (\mathbf{T}^T \mathbf{y})^T \mathbf{T}^T \mathbf{y}$$
$$= \sum_{i=1}^{K} \left(\frac{\mathbf{v}_i^T \mathbf{y}}{\sigma} \right)^2$$

since $\sum_{i=1}^{K} (\frac{\mathbf{v}_i{}^T \mathbf{y}}{\sigma}) \sim \mathcal{N}(0, \mathbf{I})$, we obtain

$$\frac{1}{\sigma^2} \mathbf{y}^T \mathbf{P}_T^T \mathbf{P}_T \mathbf{y} \sim \chi^2(K)$$

where $\chi^2(x)$ denotes the chi-square distribution with x degrees of freedom. Similarly,

$$\frac{1}{\sigma^2} \mathbf{y} \mathbf{P}_{T^\perp}^T \mathbf{P}_{T^\perp} \mathbf{y} \sim \chi^2(n-K)$$

Based on the fact that the numerator and denominator of γ are independent of each other, both of them follow the chi-square distribution of degrees of freedom of K and $n - K$, hence their ratio follows $\mathcal{F}$ distribution, $\gamma \sim \mathcal{F}(K, n-K)$. ■

ACKNOWLEDGMENT

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

The authors would like to thank Dr. Z. Hu for his insightful comments and suggestions.

REFERENCES

[1] R. C. Qiu, Z. Hu, H. Li, and M. C. Wicks, *Cognitive Radio Communication and Networking: Principles and Practice*. Hoboken, NJ, USA: Wiley, 2012.

[2] D. L. Donoho *et al.*, "High-dimensional data analysis: The curses and blessings of dimensionality," in *Proc. AMS Conf. Math Challenges Lecture*, vol. 1, 2000, pp. 1–32.

[3] L. Lu, X. Zhou, U. Onunkwo, and G. Y. Li, "Ten years of research in spectrum sensing and sharing in cognitive radio," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 1, pp. 28–43, 2012.

[4] Y. Zeng, C. L. Koh, and Y.-C. Liang, "Maximum eigenvalue detection: Theory and application," in *Proc. IEEE Int. Conf. Commun.*, 2008, pp. 4160–4164.

[5] Y. Zeng and Y.-C. Liang, "Maximum-minimum eigenvalue detection for cognitive radio," in *Proc. IEEE 18th Int. Symp. Pers., Indoor Mobile Radio Commun.*, 2007, pp. 1–5.

[6] 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.

[7] J. Ma, G. Y. Li, and B. H. Juang, "Signal processing in cognitive radio," *Proc. IEEE*, vol. 97, no. 5, pp. 805–823, May 2009.

[8] P. Zhang and R. Qiu, "GLRT-based spectrum sensing with blindly learned feature under rank-1 assumption," *IEEE Trans. Commun.*, vol. 61, no. 1, pp. 87–96, Jan. 2013.

[9] S. Hou and R. C. Qiu, "Kernel feature template matching for spectrum sensing," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2258–2271, Jun. 2014.

[10] P. Zhang, R. Qiu, and N. Guo, "Demonstration of spectrum sensing with blindly learned features," *IEEE Commun. Lett.*, vol. 15, no. 5, pp. 548–550, May 2011.

[11] M. Rashidi, K. Haghighi, A. Owrang, and M. Viberg, "A wideband spectrum sensing method for cognitive radio using sub-Nyquist sampling," in *Proc. IEEE Digit. Signal Process. Workshop IEEE Signal Process. Edu. Workshop*, 2011, pp. 30–35.

[12] R. Rao, Q. Cheng, and P. K. Varshney, "Subspace-based cooperative spectrum sensing for cognitive radios," *IEEE Sensors J.*, vol. 11, no. 3, pp. 611–620, Mar. 2011.

[13] H. Kwon and N. M. Nasrabadi, "Kernel matched subspace detectors for hyperspectral target detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 178–194, Feb. 2006.

[14] S. K. Yoo, S. Cotton, P. Sofotasios, M. Matthaiou, M. Valkama, and G. Karagiannidis, "The Fisher-Snedecor f distribution: A simple and accurate composite fading model," *IEEE Commun. Lett.*, vol. 21, no. 7, pp. 1661–1664, Jul. 2017.

[15] A. Margoosian, J. Abouei, and K. N. Plataniotis, "Accurate kernel-based spectrum sensing for Gaussian and non-Gaussian noise models," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2015, pp. 3152–3156.

[16] R. Couillet and F. Benaych-Georges, "Understanding big data spectral clustering," in *Proc. IEEE 6th Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process.*, 2015, pp. 29–32.

[17] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.

[18] M. N. Desai and R. S. Mangoubi, "Robust Gaussian and non-Gaussian matched subspace detection," *IEEE Trans. Signal Process.*, vol. 51, no. 12, pp. 3115–3127, Dec. 2003.

[19] M. A. Abdulsattar and Z. A. Hussein, "Energy detection technique for spectrum sensing in cognitive radio: A survey," *Int. J. Comput. Commun.*, vol. 4, pp. 223–242, Sep. 2012.

[20] Y. L. T. Lim, R. Zhang, and Y. Zeng, "GLRT-based spectrum sensing for cognitive radio," in *Proc. IEEE Global Telcommun. Conf.*, 2008, pp. 1–5.
[21] R. Gao, Z. Li, P. Qi, and H. Li, "A robust cooperative spectrum sensing method in cognitive radio networks," *IEEE Commun. Lett.*, vol. 18, no. 11, pp. 1987–1990, Nov. 2014.
[22] R. C. Bomfin, R. A. de Souza, and D. A. Guimarães, "Circular folding cooperative power spectral density split cancellation algorithm for spectrum sensing," *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 250–253, Feb. 2017.
[23] T. Tao and V. Vu, "Random covariance matrices: Universality of local statistics of eigenvalues," *Ann. Probab.*, vol. 40, no. 3, pp. 1285–1315, 2012.
[24] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 1, pp. 116–130, Jan.–Mar. 2009.
[25] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[26] B. Schölkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2002.
[27] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge: U.K.: Cambridge Univ. Press, 2000.
[28] N. El Karoui *et al.*, "The spectrum of kernel random matrices," *Ann. Statist.*, vol. 38, no. 1, pp. 1–50, 2010.
[29] S. Haykin, D. J. Thomson, and J. H. Reed, "Spectrum sensing for cognitive radio," *Proc. IEEE*, vol. 97, no. 5, pp. 849–877, May 2009.
[30] 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, 2010, Art. no. 381465.
[31] J. Font-Segura and X. Wang, "GLRT-based spectrum sensing for cognitive radio with prior information," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 2137–2146, Jul. 2010.
[32] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, 1998.
[33] M. Agrawal, S. Vidyashankar, and K. Huang, "On-chip implementation of ECoG signal data decoding in brain-computer interface," in *Proc. IEEE 21st Int. Mixed-Signal Test. Workshop*, 2016, pp. 1–6.
[34] A. Sharma and K. K. Paliwal, "Fast principal component analysis using fixed-point algorithm," *Pattern Recognit. Lett.*, vol. 28, no. 10, pp. 1151–1155, 2007.
[35] M. Girolami, "Mercer kernel-based clustering in feature space," *IEEE Trans. Neural Netw.*, vol. 13, no. 3, pp. 780–784, May 2002.
[36] T. Tao and V. Vu, "Random matrices: Universality of local eigenvalue statistics," *Acta Math.*, vol. 206, no. 1, pp. 127–204, 2011.
[37] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," unpublished paper, 2010. [Online]. Available: https://arxiv.org/abs/1011.3027
[38] V. Tawil, "51 captured DTV signal," informal documents, May 2006. [Online]. Available: http://grouper.ieee.org/groups/802/22/Meeting_Documents
[39] A. Kortun, T. Ratnarajah, M. Sellathurai, C. Zhong, and C. B. Papadias, "On the performance of eigenvalue-based cooperative spectrum sensing for cognitive radio," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 49–55, Feb. 2011.
[40] F. Penna, R. Garello, and M. A. Spirito, "Cooperative spectrum sensing based on the limiting eigenvalue ratio distribution in wishart matrices," *IEEE Commun. lett.*, vol. 13, no. 7, pp. 507–509, Jul. 2009.
[41] V. K. Ingle and J. G. Proakis, *Digital Signal Processing Using MATLAB: A Problem Solving Companion*. Boston, MA, USA: Cengage Learning, 2016.
[42] M. Usha, B. Ramakrishnan, and J. Sathiamoorthy, "Performance analysis of spectrum sensing techniques in cognitive radio based vehicular ad hoc networks (Vanet)," in *Proc. IEEE 2nd Int. Conf. Comput. Commun. Technol.*, 2017, pp. 74–80.
[43] S. Atapattu, C. Tellambura, H. Jiang, and N. Rajatheva, "Unified analysis of low-SNR energy detection and threshold selection," *IEEE Trans. Veh. Technol.*, vol. 64, no. 11, pp. 5006–5019, Nov. 2015.
[44] A. A. Ahmed, A. A. Alkheir, D. Said, and H. T. Mouftah, "Cooperative spectrum sensing for cognitive radio vehicular ad hoc networks: An overview and open research issues," in *Proc. IEEE Can. Conf. Elect. Comput. Eng.*, 2016, pp. 1–4.
[45] R. C. Qiu and P. Antonik, *Smart Grid Using Big Data Analytics: A Random Matrix Theory Approach*. Hoboken, NJ, USA: Wiley, 2017.

Lily Li (S'11) received the B.S. degree in computer science from the University of Electronic and Science Technology of China, Chengdu, China, in 1992, the M.S. degree in computer science from New York University, New York, NY, USA, in 2000, and the M.S. degree in mathematics in 2010 from Tennessee Technological University, Cookeville, TN, USA, where she is currently working toward the Ph.D. degree in electrical and computer engineering.

Her research interests include statistics, machine learning, big data, and wireless communications.

Shujie Hou (S'13) received the B.S. degree in information and computational science from the Daqing Petroleum Institute, Daqing, China, in 2006, the M.S. degree in computational mathematics from the Dalian University of Technology, Dalian, China, in 2008, and the Ph.D. degree in electrical and computer engineering from Tennessee Technological University, Cookeville, TN, USA, in 2013.

Her current research interests include machine learning, signal processing, and cognitive radio.

Adam Lane Anderson (S'00–M'10–SM'15) received the B.S. and M.S. degrees from Brigham Young University, Provo, UT, USA, in 2002 and 2004, respectively, and the Ph.D. degree from the University of California at San Diego, La Jolla, CA, USA, in 2008, all in electrical engineering.

He is currently a Senior Embedded RF Systems Engineer with Oak Ridge National Laboratory, Oak Ridge, TN, USA, with a Joint Faculty Appointment with Tennessee Technological University, Cookeville, TN, USA. His research interests include atypical uses of spectrum for digital communications and other fields.

Dr. Anderson was the winner of the 2014 DARPA Spectrum Challenge. He was also the recipient of the 2014 Leighton E. Sissom Award for Creativity and Innovation, and is currently an Open Track Team lead in the 2017 DARPA Spectrum Collaboration Challenge.