GLRT-Based Spectrum Sensing for Cognitive Radio

Teng Joon Lim, Rui Zhang, Ying Chang Liang and Yonghong Zeng

Abstract—In this paper, we propose several spectrum sensing methods designed using the generalized likelihood ratio test (GLRT) paradigm, for application in a cognitive radio network. The proposed techniques utilize the eigenvalues of the sample covariance matrix of the received signal vector, taking advantage of the fact that in practice, the primary signal in a cognitive radio environment will either occupy a subspace of dimension strictly smaller than the dimension of the observation space, or have a spectrum that is non-white. We show that by making various assumptions on the availability of side information such as noise variance and signal space dimension, several feasible algorithms result which all outperform the standard energy detector.

Index Terms—Spectrum sensing, cognitive radio (CR), generalized likelihood ratio test (GLRT).

I. INTRODUCTION

It is now a well-established fact that current spectrum usage at any given location, frequency band and time is inefficient, with large fractions of time where the channel is not utilized. In fact, for certain location-specific services such as television broadcasting, the incumbent licensed user may not use a particular channel at a given location at all but transmissions in that channel by other entities are still illegal. In other instances, inefficient spectrum usage is naturally caused by the non-continuous nature of Internet and other sorts of traffic (e.g. voice with its silence intervals) on communication networks, and the simple random access protocols (e.g., CSMA/CA) adopted in many networks today, which dictate for instance that terminals sense the channel and then wait a random amount of time before transmitting even after detecting that it is unoccupied.

It is thus natural that for some time now, there has been interest in the concept of “spectrum sharing”, “opportunistic spectrum access” or “cognitive radio”, which can be defined as the co-existence of a secondary network with a primary one that is the legal owner of the spectral band of interest. Given its ownership of the frequency band, the primary network service provider has the right to expect that it can continue providing services to its (legal) users and finally, Section VI concludes this paper.

II. SIGNAL MODEL

Assuming a frequency non-selective (flat) fading channel between the primary transmitter and the secondary sensor, and M antennas at the latter\(^2\) with the i-th one observing \(x_i[n]\) at time \(n\), we define \(x[n] = [x_0[n],\ldots,x_M-1[n]]^T\) as the observed signal at time \(n\). A total of \(N\) time intervals are observed, and the aggregate observation matrix can be defined as \(X = [x[0],\ldots,x[N-1]]\). With a complex baseband signal representation, the hypothesis testing problem of interest is

\[
H_0 : x[n] = w[n], \quad n = 0,\ldots,N - 1 \tag{1}
\]

\[
H_1 : x[n] = s[n] + w[n], \quad n = 0,\ldots,N - 1 \tag{2}
\]

\(^1\)Spectrally white signals.

\(^2\)Algorithms developed in this paper can be readily extended to frequency selective fading channels with single or multiple receive antennas.
where \( w[n] \) is i.i.d. circularly symmetric complex Gaussian (CSCG) noise with zero mean and unknown variance \( \sigma^2 \) in each real dimension, while \( s[n] \) is the signal to be detected, belonging to the primary transmitter, which we assume to be CSCG with zero mean and a covariance matrix \( R_s = E[s[n]s^H[n]] \).

Assuming independence in the (primary) symbol sequence transmitted, \( s[n] \) and \( s[m] \) are independent for \( n \neq m \), so that
\[
p(X|H_1) = \prod_{n=0}^{N-1} p(x[n]|H_1) \tag{3}
\]

Similarly the joint distribution of the observations under \( H_0 \) can be written in a product form. We define the empirical or sample covariance matrix of \( x[n] \) to be
\[
\hat{R}_x = \frac{1}{N} \sum_{n=0}^{N-1} x[n]x^H[n] \tag{4}
\]
and assume that it is feasible to perform an eigen-decomposition so that \( U_x \) and \( \Lambda_x \) in
\[
\hat{R}_x = U_x \Lambda_x U_x^H \tag{5}
\]
are known in each block of \( N \) observations.

If the covariance matrix \( R_s \) is rank-deficient (for instance when the number of antennas at the primary transmitter is smaller than the number of antennas at the secondary receiver), then \( \text{rank}(R_s) = N_s < M \) and the smallest \( M - N_s \) eigenvalues of \( R_s \) will be approximately equal to the noise variance \( \sigma^2 \). The \( N_s \) largest eigenvalues of \( \hat{R}_x \) will be approximately the sum of an eigenvalue of \( R_s \) and \( \sigma^2 \). These approximations become exact in the limit \( N \to \infty \).

### III. Background

Many excellent textbooks (e.g. [3]) on statistical detection theory are available that describe the general theory in great depth. In this section, we only introduce the methods that form the basis for the algorithms to be proposed very briefly, for the sake of readability.

#### A. Neyman-Pearson Theorem

The NP theorem states that the binary hypothesis test that maximizes the probability of detection (defined as the probability of deciding \( H_1 \) when \( H_1 \) is true) for a given probability of false alarm (the probability of deciding \( H_1 \) when \( H_0 \) is true) uses the likelihood ratio
\[
L(x) = \frac{p(x|H_1)}{p(x|H_0)} \tag{6}
\]
as the test statistic, where \( x \) denotes the observations. Such a likelihood ratio test (LRT) decides \( H_1 \) when \( L(x) \) exceeds a threshold \( \gamma \), which is determined by the desired \( P_{FA} \) and the distribution of \( L(x) \) under \( H_0 \).

#### B. Estimator-Correlator Structure

Consider the binary hypothesis testing problem
\[
H_0 : x = w \quad H_1 : x = s + w
\]
where \( w \sim \mathcal{CN}(0, 2\sigma^2 I) \) and \( s \sim \mathcal{CN}(0, R_s) \), and \( \mathcal{CN}(\mu, \Sigma) \) stands for the CSCG distribution with mean \( \mu \) and covariance matrix \( \Sigma \). Both \( \sigma \) and \( R_s \) are known to the detector, and hence the likelihood ratio can be computed easily. After discarding terms that are independent of the hypothesis, the LRT decides \( H_1 \) if the test statistic
\[
T_{EC}(x) = x^H R_s (R_s + 2\sigma^2 I)^{-1} x > \gamma. \tag{7}
\]
The distribution of \( T_{EC}(x) \) when \( H_0 \) is true can be found [3], and therefore the value of \( \gamma \) leading to a required \( \gamma \) may be determined numerically.

\( T_{EC}(x) \) in (7) can be seen as the correlation of the observed signal vector with the MMSE (minimum mean squared error) estimate of \( s \), i.e. \( \hat{s}_0 = R_s(R_s + 2\sigma^2 I)^{-1} x \), and is thus known as an estimator-correlator (EC) method. This is the LRT when the signal \( s \) is Gaussian, zero-mean and has a known covariance matrix, and is observed in white Gaussian noise \( w \) with known variance.

#### C. Composite Hypothesis Testing with the GLRT

In most practical scenarios, it is not possible to know the likelihood functions exactly, because of uncertainty about one or more parameters in these functions. For instance, we may not know \( \sigma \) or \( R_s \) in the hypothesis testing problem just discussed. Hypothesis testing in the presence of uncertain parameters is known as “composite” hypothesis testing. One useful method for this type of problems is the generalized likelihood ratio test (GLRT) that has been used in many applications, e.g. in radar and sonar. This method first obtains the maximum likelihood estimate (MLE) of the unknown parameters under \( H_0 \) and \( H_1 \):
\[
\hat{\Theta}_0 = \arg\max_{\Theta_0} p(x|H_0, \Theta_0) \quad \hat{\Theta}_1 = \arg\max_{\Theta_1} p(x|H_1, \Theta_1)
\]
where \( \Theta_0 (\Theta_1) \) is the set of parameters unknown under \( H_0 (H_1) \), and then forms the GLRT statistic
\[
L_G(x) = \frac{p(x|\hat{\Theta}_1, H_1)}{p(x|\hat{\Theta}_0, H_0)}. \tag{8}
\]
The GLRT decides \( H_1 \) if \( L_G(x) > \gamma \). The threshold \( \gamma \) can be set by using either an empirical distribution formed from observations over a period when the signal is known to be absent, or we can make use of the asymptotic (\( N \to \infty \)) result that the log-likelihood ratio is central chi-square distributed. More precisely,
\[
2\ln L_G(x) \sim \chi^2_r \quad \text{under } H_0 \tag{9}
\]
where \( r \) is the number of scalar unknowns that take different values under \( H_0 \) and \( H_1 \). For instance, if \( \sigma \) is known while \( R_s \) is not, \( r \) will be equal to the number of unique real-valued scalar values in \( R_s \).
IV. DETECTION ALGORITHMS

A. White Gaussian Signal Assumption

If it is assumed that \( s[n] \sim CN(0, E_s I) \) in (2), and that \( \{s[n]\}_{n=0, \ldots, N-1} \) is an independent sequence, then the estimator-correlator structure in (7) yields the test statistic

\[
T_{ED} = \sum_{n=0}^{N-1} \|x[n]\|^2 > \gamma.
\]

which is of course equivalent to the energy detector (ED) to within an irrelevant scaling term. Thus the optimal detector in this scenario decides \( H_1 \) when

\[
T_{ED}(X) = \sum_{n=0}^{N-1} \|x[n]\|^2 > \gamma.
\]

B. All Parameters Unknown

While the energy detector is easy to implement and is the most powerful detector for white Gaussian signals, in our detection problem the signal is not white in general. If we recognize that the covariance matrix of a sequence of stationary observations can be estimated, then intuitively it is plausible that a blind method (assuming no prior knowledge of the primary signal) exists which performs better than the energy detector. The precise theoretical foundation for this intuition lies in the GLRT, which we now apply to the spectrum sensing problem.

Assume that we have no knowledge of the signal covariance matrices \( R_s \) or \( R_x \), or the thermal noise power \( \sigma^2 \). We do assume however that \( s[n] \) and \( w[n] \) are independent and jointly Gaussian. With unknown parameters in the distributions of \( x[n] \), the GLRT can be applied. The likelihood function under \( H_0 \), conditioned on the unknown parameter \( \sigma^2 \)

\[
p(X|H_0, \sigma^2) = \prod_{n=0}^{N-1} \frac{1}{(4\pi\sigma^2)^{M/2}} \exp \left[-\frac{1}{4\sigma^2} \|x[n]\|^2\right]
\]

and hence

\[
\ln p(X|H_0, \sigma^2) = -\frac{MN}{2} \ln(4\pi\sigma^2) - \frac{1}{4\sigma^2} \sum_{n=0}^{N-1} \|x[n]\|^2.
\]

The MLE of \( \sigma^2 \) under \( H_0 \) can be shown to be

\[
\hat{\sigma}^2 = \frac{1}{2MN} \sum_{n=0}^{N-1} \|x[n]\|^2,
\]

which upon substitution into (13) yields

\[
\ln p(X|H_0, \hat{\sigma}^2) = -\frac{MN}{2} \ln(2\pi) - \frac{MN}{2} \frac{1}{\hat{\sigma}^2} \sum_{n=0}^{N-1} \|x[n]\|^2.
\]

Assuming that \( x[p] \) is independent of \( x[q] \) for all \( p \neq q \), the likelihood function under \( H_1 \) conditioned on \( R_x \) is

\[
p(X|H_1, R_x) = \prod_{n=0}^{N-1} \frac{1}{(2\pi)^{M/2} \det^{1/2}(R_x)} \cdot \exp \left[-\frac{1}{2} x[n]^T R_x^{-1} x[n] \right].
\]

Notice that this is not a function of the unknown \( \sigma^2 \). Hence,

\[
\ln p(X|H_1, R_x) = -\frac{MN}{2} \ln(2\pi) - \frac{N}{2} \ln(\det(R_x)) - \frac{1}{2} \sum_{n=0}^{N-1} x[n]^T R_x^{-1} x[n].
\]

The MLE of \( R_x \) under \( H_1 \) can be derived as follows. First, for convenience, we define \( A = R_x^{-1} \), and hence (16) can be written as

\[
f(A) = -\frac{MN}{2} \ln(2\pi) - \frac{N}{2} \ln(\det(A)) - \frac{1}{2} \sum_{n=0}^{N-1} x[n]^T A x[n].
\]

The first derivative of \( f(A) \) can be obtained as

\[
\frac{\partial f(A)}{\partial A} = -\frac{N}{2} A^{-1} - \frac{1}{2} \sum_{n=0}^{N-1} x[n] x[n]^T.
\]

Since \( R_x \succeq 0 \) (i.e., \( R_x \) is positive semi-definite) it follows from (17) that \( f(A) \) is a concave function of \( A \). It then follows from (18) that the MLE of \( R_x \) that maximizes \( p(X|H_1, R_x) \) can be obtained as

\[
\hat{R}_x = \frac{1}{N} \sum_{n=0}^{N-1} x[n] x[n]^T.
\]

Note that the above ML estimate for \( R_x \) is identical to the sample covariance matrix defined in (4). Substituting (19) into (16) then yields

\[
\ln p(X|H_1, \hat{R}_x) = -\frac{MN}{2} \ln(2\pi) - \frac{N}{2} \ln(\det(\hat{R}_x)) - \frac{MN}{2} \frac{1}{\hat{\sigma}^2} \sum_{n=0}^{N-1} \|x[n]\|^2.
\]

In terms of the eigenvalues of \( R_x \),

\[
det(\hat{R}_x) = \prod_{m=0}^{M-1} \lambda_m(\hat{R}_x) = \prod_{m=0}^{M-1} \lambda_{m,x}
\]

where \( \lambda_{m,x} = \lambda_m(\hat{R}_x) \), and also,

\[
\frac{1}{MN} \sum_{n=0}^{N-1} \|x[n]\|^2 = \frac{1}{M} \text{tr}(\hat{R}_x) = \frac{1}{M} \sum_{m=0}^{M-1} \lambda_{m,x}.
\]

Hence, subtracting (14) from (21) gives the log-GLRT statistic

\[
\ln L_G(X) = \frac{MN}{2} \left\{ \ln \left( \frac{1}{M} \sum_{m=0}^{M-1} \lambda_{m,x} \right) - \frac{1}{M} \ln \left( \prod_{m=0}^{M-1} \lambda_{m,x} \right) \right\}.
\]

Finally, removing constant terms and using the monotonicity of the logarithm function gives the test

\[
T_{AGM}(\lambda_x) = \frac{1}{(T_{AGM}(\lambda_x))^{1/M}} \geq \frac{\gamma}{\gamma_0}.
\]

Note that the dependence of the test statistic on the eigenvalues of the sample covariance matrix \( \lambda_x \) has been made explicit, to
emphasize that the proposed test depends on the observations only through $R_x$. This test statistic is the ratio of the arithmetic mean to the geometric mean of the eigenvalues, which is well-known to be a good measure of the spread of a collection of values with a fixed sum because

$$\left( \prod_{m=0}^{M-1} \lambda_{m,x} \right)^{1/M} \leq \frac{1}{M} \sum_{m=0}^{M-1} \lambda_{m,x}$$  \hspace{1cm} (25)$$

with equality if and only if the values are all equal i.e. the geometric mean is upper bounded by the arithmetic mean – the more unequal the values, the further away the GM will be from its upper bound, the AM. We thus refer to this detection algorithm as the AGM (arithmetic to geometric mean) method.

C. Noise Variance Known, Signal Covariance Unknown

The energy detector is optimal when the received signal covariance matrix is a scaled identity matrix (a poor assumption), while the AGM technique works better by using the GLRT and considering both noise variance and signal covariance to be unknown (a good assumption). In this section, we introduce detectors that assume either that the noise variance or the signal-subspace dimension is known, but $R_x$ is unknown and thus has to be estimated in a GLRT.

The log-likelihood function under $H_0$ is given in (13) where $\sigma^2$ is now treated as a known parameter, and that under $H_1$, conditioned on the unknown parameter $R_x$, is

$$\ln p(X|H_1, R_x) = -\frac{MN}{2} \ln(2\pi) - \frac{N}{2} \ln(\det (R_x + 2\sigma^2 I))$$

$$\frac{1}{2} \sum_{n=0}^{N-1} x^T[n] (R_x + 2\sigma^2 I)^{-1} x[n].$$  \hspace{1cm} (26)$$

The MLE of $R_x$ under $H_1$ can be obtained as follows. First, like in the last section, we introduce $A = R_x^{-1} = (R_x + 2\sigma^2 I)^{-1}$ so that

$$\ln p(X|H_1, R_x) = \ln f(A)$$

in (17). Since $R_x \succeq 0$, it follows that $R_x \succeq 2\sigma^2 I$ and hence $A \succeq \frac{1}{2\sigma^2} I$. The MLE of $R_x$ can be obtained from the MLE of $A$ by solving the following constrained optimization problem:

Maximize $f(A)$ \hspace{1cm} (27)

Subject to $A \succeq 0$ \hspace{1cm} (28)

$$A \succeq \frac{1}{2\sigma^2} I.$$ \hspace{1cm} (29)$$

It was shown earlier that $f(A)$ is a concave function of $A$. Furthermore, since the constraints in (28) and (29) specify a convex set of $A$, it follows that the above optimization problem is convex. Notice how the assumption that $\sigma^2$ is known leads to a radically different MLE for $R_x$ compared to the unknown $\sigma^2$ case studied in the last section, where the MLE of $R_x$ was simply the sample covariance matrix.

From the Karush-Kuhn-Tucker (KKT) conditions, the optimal $A$ can be obtained as

$$A^* = U_x \text{Diag} \left( \min \left( \frac{1}{\lambda_{1,x}}, \frac{1}{2\sigma^2} \right), \ldots, \min \left( \frac{1}{\lambda_{M,x}}, \frac{1}{2\sigma^2} \right) \right) U_x^H,$$

where $\lambda_{m,x}$’s and $U_x$ are obtained from the eigen-decomposition of the sample covariance matrix as in (5). Without loss of generality, we also assume from this point that the eigenvalues are ordered from largest to smallest i.e.,

$$\lambda_{1,x} \geq \lambda_{2,x} \geq \cdots \geq \lambda_{M,x}.$$

The MLE of $R_x$ can now be obtained as

$$\hat{R}_x = U_x \text{Diag} \left( (\lambda_{1,x} - 2\sigma^2)^+ , \ldots , (\lambda_{M,x} - 2\sigma^2)^+ \right) U_x^H,$$  \hspace{1cm} (30)$$

where $(x)^+ = \max(x, 0)$. By substituting (26) and (13) into (8), with $\Theta_1 = R_x$ in (30) and $\Theta_0 = \emptyset$, the GLRT statistic can be obtained as

$$T_{\text{GLRT}}(\lambda_x) = N \sum_{m'=m}^{N} \ln \left( \frac{\lambda_{m,x}}{2\sigma^2} - \sum_{m''=m'}^{N} \frac{\lambda_{m,x}}{2\sigma^2} I \right)$$

$$+ 2 \sum_{m'=m}^{N} \ln \left( \frac{\lambda_{m,x}}{2\sigma^2} - 1 \right),$$  \hspace{1cm} (31)$$

where $m'$ corresponds to the largest $m$ such that $\lambda_{m,x} > 2\sigma^2$. Interestingly, this test statistic can be easily shown to be again a function of arithmetic geometric means, but this time of only the eigenvalues in the estimated signal subspace. We call this algorithm the SSE (signal-subspace eigenvalues) method. Some variations of the SSE method are presented as follows:

a) Rank of Signal Subspace Known, $\sigma^2$ Unknown: If rank($R_x$) = $N_x$ is known but the noise variance is unknown and estimated as the smallest eigenvalue of $R_x$ in (4), then (31) becomes

$$T_{\text{SSE}}(\lambda_x) = \ln \left( \prod_{m \leq N_x} \frac{\lambda_{m,x}}{\lambda_{M,x}} \right)^{1/N_x} - \frac{1}{N_x} \sum_{m \leq N_x} \frac{\lambda_{m,x}}{\lambda_{M,x}}.$$  \hspace{1cm} (32)$$

b) Signal Subspace Assumed to be Rank-One: Further simplification results when we can reasonably assume that $N_x = 1$. Then (32) becomes only a function of the eigenvalue spread or ratio of largest to smallest eigenvalues. Hence we have the MME (maximum to minimum eigenvalue) test statistic [10]:

$$T_{\text{MME}}(\lambda_x) = \frac{\lambda_{1,x}}{\lambda_{M,x}}.$$  \hspace{1cm} (33)$$

V. SIMULATION RESULTS

The tests proposed in this paper are summarized in Table V, which provides the equation number corresponding to the respective test statistic for convenience. In all the listed methods except the estimator-correlator, we assume no knowledge of the structure of the primary signal other than that it is Gaussian. For the eigenvalue-based methods (AGM, SSE1, SSE2 and MME), we also assume that the primary signal occupies a rank-deficient subspace of the observation signal space. Under this condition, the proposed algorithms are able to discriminate between the null and alternative hypotheses using the eigenvalues of the sample covariance matrix $R_x$.  

<table>
<thead>
<tr>
<th>Name</th>
<th>Test Statistic</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimator-Correlator (E-C)</td>
<td>$T_{EC}$</td>
<td>(7)</td>
</tr>
<tr>
<td>Energy Detector (ED)</td>
<td>$T_{ED}$</td>
<td>(11)</td>
</tr>
<tr>
<td>AM/AGM (AGM)</td>
<td>$T_{AGM}$</td>
<td>(24)</td>
</tr>
<tr>
<td>Signal-Subspace E-Values 1 (SSE1)</td>
<td>$T_{SSE}^1$</td>
<td>(31)</td>
</tr>
<tr>
<td>Signal-Subspace E-Values 2 (SSE2)</td>
<td>$T_{SSE}^2$</td>
<td>(32)</td>
</tr>
<tr>
<td>Maximum/Minimum E-Value (MME)</td>
<td>$T_{MME}$</td>
<td>(33)</td>
</tr>
</tbody>
</table>
Monte Carlo simulations are carried out with each simulation consisting of 10,000 independent samples. A secondary receiver with $M$ antennas was simulated with a primary transmitter having $Q$ transmit antennas, each carrying an independent data stream. Notice that $\text{rank}(R_s) = N_s = \min(M, Q)$. Independent flat-fading MIMO channels between each transmit-receiver antenna pair is assumed. For each detection algorithm, the threshold is chosen to achieve the target probability of false alarm $P_{FA}$ equal to 10%. 1000 Monte Carlo simulations are carried out with each simulation consisting of 10,000 independent samples.

We first choose $M = 4$ and $Q = 1$. The probability of detection $P_D$ is plotted against SNR in Fig. 1. As can be seen, the E-C always performs the best since it is the optimal detector assuming perfect knowledge of the received signal covariance and the noise variance. If only the noise variance is perfectly known, SSE1 performs 1dB better than perfect ED without noise uncertainty, denoted as ED(0dB), for the case of 90% probability of detection. However, if the noise variance is unknown, even with 0.5dB noise uncertainty [10], denoted as ED(0.5dB), the ED performs more than 11dB worse than the proposed eigenvalue-based methods (AGM, MME and SSE2). Notice that since $Q = 1$, SSE2 here is equivalent to MME.

We then choose $M = 8$ and $Q = 3$. The $P_D$ performance curves are given in Fig. 2. Again, when the noise variance is known, SSE1 is only slightly better than perfect ED at $P_D = 90\%$. However in the realistic case of unknown (or inaccurately known) noise variance, the proposed eigenvalue-based methods can achieve 90% detection probability at a SNR below -20dB. However, to achieve the same target probability of detection, the ED with just 0.5dB noise uncertainty needs a SNR of about -9dB.

VI. CONCLUSION

In this paper, we have applied the well-known GLRT principle to the problem of spectrum sensing in a cognitive radio network. Under mild assumptions on the primary signal (the one to be sensed), the proposed eigenvalue-based algorithms all perform better than the conventional energy detector, with and without the knowledge of noise variance. The price paid for this improvement is higher implementation complexity because of the need to estimate the signal covariance matrix, and perform an eigen-decomposition on this estimate. Practically, the algorithms enable shorter sensing intervals for given probabilities of detection and false alarm, or higher $P_D$/lower $P_{FA}$ for a given sensing interval – these features are valuable for the successful development of wireless networks based on opportunistic spectrum access.

REFERENCES