Spectrum Opportunity Detection with Weak and Correlated Signals

Yao Xie
Department of Electrical and Computer Engineering
Duke University, North Carolina, 27705
Email: yao.xie@duke.edu

David Siegmund
Department of Statistics
Stanford University, California, 94305
Email: siegmund@stanford.edu

***Abstract*—We present a novel score detector for temporal spectrum opportunity detection in cognitive radio by exploiting the differences in both energy and correlation of the empty band and the occupied band. Motivated by the challenge of detecting a weak primary user's signal without precise knowledge of the signal, where the conventional energy detector faces the limit of "SNR wall", we assume a simple model which captures a key difference between noise and primary user's signal – their correlation structures. Besides the merit of incorporating signal correlation, our score detector also avoids the computational complexity of covariance matrix inversion incurred by the corresponding maximum likelihood statistic assuming signal correlation. We provide a theoretical approximation to the false-alarm-rate of the score detector, which can be used to determine the threshold efficiently. We demonstrate that our approximation is quite accurate, and that our score detector has an advantage when the signal is weak and correlated.**

I. Introduction

Cognitive radio is an emerging technology which has a great potential to increase spectrum efficiency. A key challenge in cognitive radio is spectrum opportunity detection. Recently, temporal spectrum opportunity detection via change-point detection approach has attracted much interest [1], because the change-point formulation fits well with the temporal opportunity detection objective. Also, there is a wealth of tools for change-point detection from the statistics literature. Existing methods for spectrum opportunity detection primarily use three signal processing techniques: matched filtering of the spectrum of the received signal with that of the primary user, cyclostationary feature detection, and energy detection exploiting the difference between received signal power and noise power. The matched filtering and feature detector require detailed knowledge of primary user's signal. When such knowledge is not available, energy detector is a popular choice. However, there is a fundamental limit with the energy detector – the *SNR wall* – when the signal-to-ratio (SNR) is low, error for energy detector is very high. In wireless environment, however, due to fading, SNR of the primary user's signal can be quite low. Hence, in cognitive radio, there is a need for developing weak signal detector, which also requires little priori knowledge about the primary user's signal.

Motivated by the challenge of detecting weak signals, in this paper we propose a novel score detector by exploiting differences between signal and noise in both energy and correlation. We assume a simple model for the primary user's signal, which captures a key difference between signal and noise: their correlation structure. Based on this model, we develop a detector using the score statistic, which avoids the computationally intensive inversion of covariance matrix in the corresponding maximum likelihood statistic assuming signal correlation. In contrast to the matched filtering detector, our model does not require precise knowledge about primary user's signal. We derive a theoretical approximation to the false-alarm-rate of our score detector, which can be used to determine the threshold efficiently. We demonstrate using numerical simulation that our approximation is quite accurate. We also show that the score detector has better performance than the energy detector and the maximum likelihood detector [1] that only assumes power change.

Sections are organized as follows. Section II presents formulation. Section III derives the score detector. Section IV contains a theorem that approximates the probability-of-false-alarm for the score detector. Section V demonstrates that our approximation is accurate and the score detector has good performance. Finally, VI concludes the paper.

II. Formulation

Consider a cognitive radio system with a primary user and a secondary user. A secondary user tunes to a frequency band and starts to take samples y_n, $n = 1, \cdots, N$, where N is the number of samples. If the band is not occupied, the samples are white noise. If the band is occupied, the samples consist white noise and a primary user's signal. Assume that the primary user's signal emerges at an unknown time k, $1 \leq k \leq N$. Using these samples, our goal is to detect whether or not the band is occupied. This problem can be formulated as the following hypothesis test

$$\mathsf{H}_0: \quad y_n = w_n, \quad n = 1, \cdots, N, \tag{1}$$

$$\mathsf{H}_1: \quad \begin{cases} y_n = w_n, & n = 1, \cdots, k; \\ y_n = x_n + w_n, & n = k+1, \cdots, N, \end{cases} \tag{2}$$

where $w_n \sim \mathcal{N}(0, \sigma^2)$, i.e., it is i.i.d Gaussian white noise with zero mean and variance σ_0^2, and x_n is the primary user's signal seen by the secondary user. Typically x_n is temporally correlated, due to dispersion of wireless channel and modulation scheme used by the transmitter. Let $^\top$ denote the transpose of a vector or matrix. We assume the covariance matrix of the signal vector $\boldsymbol{x}_k \triangleq [x_{k+1}, \cdots, x_N]^\top$ is given by

$$\mathbb{E}\{\boldsymbol{x}_k \boldsymbol{x}_k^\top\} \triangleq P\boldsymbol{V}_{N-k,\theta},$$

where P is the signal power, and

$$\boldsymbol{V}_{N-k,\theta} \in \mathbb{R}^{(N-k)\times(N-k)}$$

is the normalized covariance matrix parameterized by

$$\theta \in \Theta \subset \mathbb{R}^d.$$

For example, when x_n has weak temporal correlation, we can model it using the first-order autoregressive process (AR-1), which corresponds to $d = 1$. Hence, under H_1, the covariance matrix of the received signal vector $\boldsymbol{y}_k \triangleq [y_{k+1}, \cdots, y_N]^\top$ is given by

$$\mathbb{E}_{\mathsf{H}_1}\{\boldsymbol{y}_k\boldsymbol{y}_k^\top\} \triangleq \boldsymbol{\Sigma} = \sigma_0^2 \boldsymbol{I}_{N-k} + P\boldsymbol{V}_{N-k,\theta}. \tag{3}$$

We assume the signal correlation model (e.g., AR-1) is known, but parameters P, k and θ are unknown.

The energy detector detects the primary user by comparing the receive energy with a threshold $b > 0$

$$\sum_{n=1}^{N} y_n^2 \geq b.$$

In [1], the maximum likelihood (ML) detector exploiting power difference without assuming signal correlation structure is studied. For fixed sample size N, it corresponds to the following detector:

$$\max_{\substack{0\leq k\leq N-N_0\\ P\in[P_1,P_2]}} \frac{P}{(P+\sigma_0^2)\sigma_0^2}\boldsymbol{y}_k^T\boldsymbol{y}_k - (N-k)\log\left(\frac{P+\sigma_0^2}{\sigma_0^2}\right) \geq b$$

for some threshold b, where the unknown P is assumed to be within the range of $[P_1, P_2]$, and N_0 is a parameter that specifies the minimum number of samples needed detection.

The log likelihood ratio of hypothesis H_1 versus H_0 for given P, k and θ is given by:

$$\mathcal{L}(P,k,\theta) = \frac{1}{2}\mathbf{y}_k^\top\left(\frac{\boldsymbol{I}}{\sigma_0^2} - \boldsymbol{\Sigma}^{-1}\right)\mathbf{y}_k - \frac{1}{2}\log\left(\frac{|\boldsymbol{\Sigma}|}{\sigma_0^{2(N-k)}}\right). \tag{4}$$

The maximum likelihood detector assuming signal correlation structure detects when:

$$\max_{0\leq k\leq N-N_0, P\geq 0, \theta\in\Theta} \mathcal{L}(P,k,\theta) \geq b \tag{5}$$

for some threshold $b > 0$. It has a major drawback that it requires computing a matrix inversion $\boldsymbol{\Sigma}^{-1}$ for each and all possible values of k, θ, and P.

III. Score Detector

We present a maximum score detector that can avoid the covariance matrix inversion in the maximum likelihood formulation (5). Maximum score is an alternative to maximum likelihood that is frequently used in deriving efficient test, especially when the parameter for H_0 only contains one single point. In our case, H_0 corresponds to $P = 0$ in (3), and hence we can derive an efficient score detector.

The score detector is derived as follows. Consider the derivative of $\mathcal{L}$ in (4) with respect to P and evaluated at $P = 0$:

$$\left.\frac{\partial\mathcal{L}}{\partial P}\right|_{P=0} = \frac{1}{2\sigma_0^4}\text{tr}[\boldsymbol{V}_{N-k,\theta}(\mathbf{y}_k\mathbf{y}_k^\top - \sigma_0^2\boldsymbol{I})]. \tag{6}$$

where tr$(\boldsymbol{A})$ denotes the trace of a square matrix $\boldsymbol{A}$. In the following $\mathbb{P}\{\cdot\}$, $\mathbb{E}\{\cdot\}$ and $\mathsf{var}\{\cdot\}$ denote the probability, mean and variance under the null hypothesis. We can verify the mean of the derivative under the null hypothesis is zero: $\mathbb{E}\left\{\left.\frac{\partial\mathcal{L}}{\partial P}\right|_{P=0}\right\} = 0$. The covariance of (6) under null hypothesis can be shown to be

$$\mathsf{var}\left\{\left.\frac{\partial\mathcal{L}}{\partial P}\right|_{P=0}\right\} = \frac{1}{2\sigma_0^4}\text{tr}(\boldsymbol{V}_{N-k,\theta}\boldsymbol{V}_{N-k,\theta}^\top). \tag{7}$$

The score statistics is given by $\left.\frac{\partial\mathcal{L}}{\partial P}\right|_{P=0}$ normalized by its mean and variance:

$$Z(k,\theta) = \text{tr}\left[\frac{\boldsymbol{V}_{N-k,\theta}}{\sqrt{2\text{tr}(\boldsymbol{V}_{N-k,\theta}\boldsymbol{V}_{N-k,\theta}^\top)}}\left(\frac{\mathbf{y}_{k+1}\mathbf{y}_{k+1}^\top}{\sigma_0^2} - \boldsymbol{I}_{N-k}\right)\right]. \tag{8}$$

Note that $Z(k,\theta)$ is a two dimensional random field in $k \in \{1, \cdots N - N_0\}$ and $\theta \in \Theta$, with zero mean and unit variance. The second expression shows that $Z(k,\theta)$ can be interpreted as a "matched-filter": matching the theoretical covariance matrix with the sample covariance matrix. The score detector detects a signal when:

$$\max_{0\leq k\leq N-N_0, \theta\in\Theta} Z(k,\theta) \geq b, \tag{9}$$

for some threshold $b > 0$.

We consider two performance metrics for the detectors presented above: the probability-of-false-detection

$$P_e \triangleq \mathbb{P}_{\mathsf{H}_0}\{T \geq b\},$$

and the probability-of-detection under H_1:

$$P_d \triangleq \mathbb{P}_{\mathsf{H}_1}\left\{T \geq f^{-1}(\alpha)\right\},$$

where T is the statistic of the corresponding detector. Assume that P_e is related to the threshold by some function f: $P_e = f(b)$. We will obtain this function in Theorem 1.

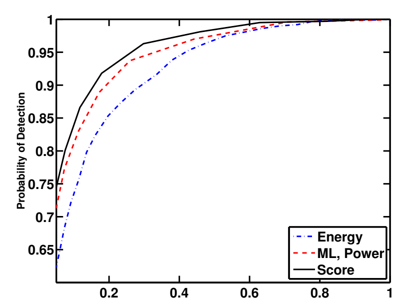

Fig. 1: Probability-of-detection versus probability-of-false-detection for the energy detector, ML power detector, and the score detector, for $\theta_0 = 0.7$ and $\sigma_s^2 = 0.07$. The noise variance is $\sigma_0^2 = 1$.

IV. Approximation of Probability-of-False-Detection

We prove the following theorem that approximates the probability-of-false-detection P_e of the score detector for a given threshold b, which can be used to determine the threshold efficiently.

Theorem 1: When $b \to \infty$, under H_0, the probability-of-false-detection of the score detector (9) is given by

$$P_e \triangleq \mathbb{P}_{\mathsf{H}_0}\left\{\max_{0\le k\le N-N_0,\theta\in\Theta} Z(k,\theta)\ge b\right\}$$
$$= \frac{1}{(2\pi)^{\frac{d}{2}}}\sum_{k=0}^{N-N_0}\int_{\theta\in\Theta}\frac{[b\xi_0(k,\theta)]^{\frac{d}{2}}}{\xi_0(k,\theta)}\cdot$$
$$g(k,\theta)|\boldsymbol{H}(k,\theta)|^{\frac{1}{2}}\frac{b^2\mu(k,\theta)}{2(N-k)}\nu\left(\sqrt{\frac{b^2\mu(k,\theta)}{N-k}}\right)d\theta+o(1).$$

where

$$\tilde{\boldsymbol{V}} = \frac{\boldsymbol{V}_{N-k,\theta}}{\sqrt{2\mathrm{tr}(\boldsymbol{V}_{N-k,\theta}\boldsymbol{V}_{N-k,\theta}^\top)}},$$

$\xi_0(k,\theta)>0$ is the solution to

$$h(\xi)\triangleq \mathrm{tr}\left[(\boldsymbol{I}_{N-k}-2\xi_0\tilde{\boldsymbol{V}})^{-1}\tilde{\boldsymbol{V}}\right]-\mathrm{tr}(\tilde{\boldsymbol{V}})=b, \tag{10}$$

$$\mathsf{var}_{\xi_0}\{Z\} = 2\mathrm{tr}\left[(\boldsymbol{I}_{N-k}-2\xi_0\tilde{\boldsymbol{V}})^{-1}\tilde{\boldsymbol{V}}(\boldsymbol{I}-2\xi_0\tilde{\boldsymbol{V}})^{-1}\tilde{\boldsymbol{V}}\right],$$

$$\psi(\xi) = -\mathrm{tr}(\xi\tilde{\boldsymbol{V}})-\frac{1}{2}\log|\boldsymbol{I}-2\xi\tilde{\boldsymbol{V}}|,$$

$$g(k,\theta)=\frac{\exp\{-\xi_0(k,\theta)b+\psi(\xi_0(k,\theta))\}}{\sqrt{2\pi\mathsf{var}_{\xi_0}\{Z\}}}, \tag{11}$$

the Hessian of the covariance matrix (Fisher information matrix) is given by

$$\boldsymbol{H}(k,\theta) = -\left.\frac{\partial^2\mathbb{E}\{Z(k,\theta)Z(k,s)\}}{\partial^2 s}\right|_{s=\theta}, \tag{12}$$

$$\mu(k,\theta)=(N-k)\left[\frac{\mathrm{tr}\left(\boldsymbol{V}_{N-k+1,\theta}\boldsymbol{V}_{N-k+1,\theta}^\top\right)}{\mathrm{tr}\left(\boldsymbol{V}_{N-k,\theta}\boldsymbol{V}_{N-k,\theta}^T\right)}-1\right],$$

and special function (given by Corollary 8.44 of [2] or in [3]):

$$\nu(x)\approx\frac{\frac{2}{x}\left[\Phi\left(\frac{x}{2}\right)-\frac{1}{2}\right]}{\frac{x}{2}\Phi(\frac{x}{2})+\phi\left(\frac{x}{2}\right)}, \tag{13}$$

where $\phi(x)$ and $\Phi(x)$ are the density and distribution function of $\mathcal{N}(0,1)$, respectively.

The proof of this theorem is provided in Appendix.

V. Numerical Examples

In this section, we first verify that our approximation in Theorem 1 is very accurate, and then demonstrate that the score detect compares favorably with other detectors that do not exploit signal correlation.

A. Approximation accuracy of Theorem 1 for P_e

Consider the first-order autoregressive process AR(1), with $d=1$, $\theta\in[0.1,0.5]$, $N=100$, $N_0=3$. The AR(1) process evolves according to

$$x_{n+1}=\theta x_n+\varepsilon_n, \tag{14}$$

where $|\theta|<1$, and the process noise ε_n are i.i.d. $\mathcal{N}(0,1)$. The covariance matrix for AR(1) is given by

$$\boldsymbol{V}_{N-k,\theta}=\begin{pmatrix} 1 & \theta & \theta^2 & \cdots & \theta^{N-k-1}\\ \theta & 1 & \theta & \cdots & \theta^{N-k-2}\\ & & \ddots & & \end{pmatrix}. \tag{15}$$

To demonstrate that our approximation in Theorem 1 is also accurate for $d>1$, we also consider a first-order autoregressive moving-average process ARMA(1,1), with $d=2$, parameters θ and ϕ. The AR(1, 1) process evolves according

$$x_{n+1}+\phi x_n=\theta\varepsilon_n+\varepsilon_{n+1},$$

where $\varepsilon_n\sim\mathcal{N}(0,1)$. The corresponding covariance matrix is given by

$$\boldsymbol{V}_{N-k,\theta}=$$
$$\begin{pmatrix} 1+\theta^2-2\phi\theta & (\phi-\theta)(1-\phi\theta) & \phi(\phi-\theta)(1-\phi\theta) & \cdots\\ (\phi-\theta)(1-\phi\theta) & 1+\theta^2-2\phi\theta & (\phi-\theta)(1-\phi\theta) & \cdots\\ \phi(\phi-\theta)(1-\phi\theta) & \ddots & \ddots & \ddots \end{pmatrix}.$$

In Table I (for AR(1)) and Table II (for AR(1, 1)), the Monte Carlo results are all obtained from 1000 trials. Note that our approximation in Theorem 1 is very close to the corresponding Monte Carlo result in both examples.

TABLE I: Approximate of P_e for AR(1)

b	Monte Carlo	Theorem 1
3.5000	0.1170	0.1045
4.5000	0.0560	0.0518
5.5000	0.0250	0.0242
6.5000	0.0090	0.0112

TABLE II: Approximate P_e for ARMA(1, 1)

b	Monte Carlo	Theorem 1
3.5000	0.0780	0.0667
4.5000	0.0340	0.0341
5.0000	0.0260	0.0235
6.0000	0.0120	0.0109

B. Comparison of probability-of-detection P_d

We compare our score detector with the energy detector and the maximum likelihood (ML) power detector [1]. We simulate the signal as an AR(1) process (14), with $\theta=\theta_0$, and process noise with variance σ_s^2. The signal power of the AR(1) process can be estimated using the variance of the steady state $P=\sigma_s^2/(1-\theta_0^2)$. Assume $N=100$, the primary user emerges at $k=50$ with $P=1$, and consider various true signal correlation θ_0. The probability-of-detection is simulated via 500 Monte Carlo trials. For the score statistic, we set $\theta\in[0.3,0.8]$. Fig. 1 demonstrates that the score detector has the highest probability-of-detection among the three detectors, over a wide range of probability-of-false-detection (false-alarm rate), when $\theta_0=0.7$ and $\sigma_s^2=0.07$. Table III further demonstrates that the score detector has the best performance among the three detectors different θ_0 and σ_s^2 values.

TABLE III: P_d for a fixed $P_e = \alpha = 0.1$. Each cell of the table corresponds to "score detector / ML power detector / energy detector". The bold numbers correspond to score detector.

σ_s^2 \ θ_0	0.5	0.7	0.8
0.05	**0.79**/0.68/0.60	**0.90**/0.76/0.62	**0.92**/0.75/0.68
0.07	**0.89**/0.79/0.70	**0.95**/0.83/0.73	**0.98**/0.86/0.79

VI. Conclusions

We have present a novel score detector to detect weak signal exploiting both the signal temporal correlation and power. Here we consider cognitive radio spectrum detection. Nevertheless, our formulation of the score detector is general and it can be used for similar problems. We provide a theoretical approximation to the false-alarm-rate of the score detector, which can be used to determine the threshold efficiently. We demonstrate that our approximation is quite accurate, and that the score detector has an advantage when signal is weak and correlated.

Appendix

In the following we prove Theorem 1 for $d = 1$.

1) Last Hitting Time Formula: We discretize the parameter $\theta \in (\theta_1, \theta_2)$, $k \in \{1, \cdots N - N_0\}$ by rectangular mesh grid of size of $\frac{\Delta}{\sqrt{N}}$ times 1, where $\Delta > 0$ is a small number. The size of the mesh is chosen to balance the difference in the order of the variance in these two coordinates. Then P_e of the score detector can be approximated as

$$\mathbb{P}\left\{\max_{(i,j)\in D} Z\left(i, j\frac{\Delta}{\sqrt{N}}\right) \geq b\right\}, \tag{16}$$

where the index set

$$D = \left\{(i,j) : 0 \leq i \leq N, \theta_1 \leq j\frac{\Delta}{\sqrt{N}} \leq \theta_2\right\} \tag{17}$$

covers the parameter space. Let the index set $J(i_0, j_0)$ denotes everything to the "future" of the current index (i_0, j_0) (upper and to the right of (i_0, j_0)) in the random field.

$$J(i_0, j_0) = \{(i,j) \in D : j \geq j_0, \quad \text{or} \quad j = j_0 \quad \text{and} \quad i \geq i_0.\}$$

Using the "last hitting time" decomposition [5], we can rewrite (16) as

$$\begin{aligned}
&\mathbb{P}\left\{\max_{(i,j)\in D} Z\left(i, j\frac{\Delta}{\sqrt{N}}\right) \geq b\right\} \\
&\sim \sum_{(i_0,j_0)\in D} P\left\{Z\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right) \geq b, \max_{(i,j)\in J(i_0,j_0)} Z\left(i, j\frac{\Delta}{\sqrt{N}}\right) \leq b\right\} \\
&= \sum_{(i_0,j_0)\in D} \int_0^\infty P\left\{Z\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right\}\frac{dx}{b}\cdot \\
&\quad \mathbb{P}\left\{\max_{(i,j)\in J(i_0,j_0)} Z\left(i, j\frac{\Delta}{\sqrt{N}}\right) \leq b \middle| Z\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right\}.
\end{aligned} \tag{18}$$

Next we will find approximations for the probability

$$\mathbb{P}\left\{Z\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right\}\frac{dx}{b}, \tag{19}$$

and the conditional probability

$$\mathbb{P}\left\{\max_{(i,j)\in J(i_0,j_0)} Z\left(i, j\frac{\Delta}{\sqrt{N}}\right) \leq b \middle| Z\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right\}, \tag{20}$$

respectively.

2) Skewness Correction: Note that $Z(k, \theta)$ is a quadratic function in a Gaussian random vector, so Gaussian distribution is not likely to be a good approximation for $Z(k, \theta)$. Moreover, since (20) is smaller than (19) in magnitude, getting an accurate approximate for (20) is important. On the other hand, Gaussian approximation is better for the mean than for the tail of the distribution. To obtain a better approximation, we use the change-of-measure technique to shift the mean of the random field to the threshold, and use complete cumulative generating function. Note that (19) can be written as

$$\begin{aligned}
&\mathbb{P}\left\{Z\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right) = b + \frac{x}{b}\right\}\frac{dx}{b} \\
&\approx g\left(i_0, j_0\frac{\Delta}{\sqrt{N}}\right)\exp\left(-\frac{\xi_0}{b}x\right)\frac{dx}{b}.
\end{aligned} \tag{21}$$

where g and ξ_0 are defined in (12) and (10), respectively.

3) Local Analysis of Covariance: Consider the covariance $\mathbb{E}\{Z(n, \theta_1)Z(m, \theta_2)\}$ for scores at change-points n and m, and with parameters θ_1 and θ_2, respectively. Assume the covariance matrix associated with $\mathbf{y}_n$ is $\boldsymbol{V}_{N-n,\theta_1}$, and that associated with $\mathbf{y}_m$ is $\boldsymbol{V}_{N-m,\theta_2}$, respectively. Also assume $n > m$, so the dimension of the covariance matrix for $\mathbf{y}_m$ is larger than that for $\mathbf{y}_n$. Also note that $\mathbf{y}_n$ and $\mathbf{y}_m$ have overlapping samples, so $\boldsymbol{V}_{N-n,\theta_1}$ is a sub-block matrix of $\boldsymbol{V}_{N-m,\theta_2}$. Based on these observations, after some derivations, we find the covariance of $Z(k, \theta)$ under H_0 to be

$$\begin{aligned}
&\mathbb{E}\{Z(n, \theta_1)Z(m, \theta_2)\} \\
&= \frac{\text{tr}\left(\boldsymbol{V}_{N-n,\theta_1}\boldsymbol{V}_{N-n,\theta_2}^\top\right)}{\left[\text{tr}\left(\boldsymbol{V}_{N-n,\theta_1}\boldsymbol{V}_{N-n,\theta_1}^\top\right)\text{tr}\left(\boldsymbol{V}_{N-m,\theta_2}\boldsymbol{V}_{N-m,\theta_2}^\top\right)\right]^{1/2}}.
\end{aligned} \tag{22}$$

A special case is when $\theta_1 = \theta_2$, $n = m$, then (22) becomes $\mathbb{E}\left\{Z(n, \theta_1)^2\right\} = 1$, which is consistent with the unit variance of $Z(k, \theta)$.

To study the local covariance of the random field, set $\theta_1 = \theta$, $\theta_2 = \theta + \delta$, $n = k$ and $m = k - i$, $i = 1, \cdots, k - 1$ in (22). We have to run the index k for the change point backwards, because the smaller the k, the more post-change samples we have, and hence the larger the dimension of the covariance matrix $\boldsymbol{V}_{N-n,\theta}$. Assume δ and i are small relative to θ and k, respectively. We will expand everything in terms of θ, k, δ and i. Using the first order approximation (keeping only the first order terms), the local covariance is give by

$$\begin{aligned}
&\mathbb{E}\{Z(k, \theta)Z(k - i, \theta + \delta)\} \\
&\approx 1 - \gamma^2(k, \theta)\delta^2 - \frac{\mu(k, \theta)}{2(N - k)}i + o(\delta^2) + o(i).
\end{aligned} \tag{23}$$

When $d = 1$, the Hessian matrix $\boldsymbol{H}(k, \theta)$ is a scalar and is given by

$$\gamma(k,\theta)=\frac{\operatorname{tr}\left(\boldsymbol{V}'_{N-k,\theta}\boldsymbol{V}^{\top}_{N-k,\theta}\right)}{\operatorname{tr}\left(\boldsymbol{V}_{N-k,\theta}\boldsymbol{V}^{\top}_{N-k,\theta}\right)}, \tag{24}$$

where the prime f' denotes $\frac{\partial f(\theta)}{\partial\theta}$. Note that $\gamma(k,\theta)$ is independent of i. The $\mu(\theta_0,k_0)$ defined in (12), is the ratio of the average sum-of-squares of the "additional" terms when we decrease k by a small amount, over the average sum-of-squares of the "original" terms, in the covariance matrix $V_{N-k,\theta}$.

4) Local Random Field Decomposition: Note that the first order expansion of the covariance does not have cross product terms. This implies that if we assume $Z(k,\theta)$ to be Gaussian, it can be decomposed as a sum of two independent one dimensional random processes. Using the local covariance we just found, we have the following Lemma:

Lemma 2: Assume $\xi\to\infty$, $b\to\infty$, $n\to\infty$, with $\frac{\xi}{b}\sim 1$ and $\frac{b}{\sqrt{N}}\sim d$ where $d>0$ is some constant. The discretized process $b\left[Z\left(k-i,\theta+\frac{\Delta}{\sqrt{N}}j\right)-\xi\right]$, $i\in\mathbb{Z}, j\geq 0$, conditioned on $Z(k,\theta)=\xi$ can be written as sum of two independent processes:

$$\left\{b\left[Z\left(k-i,\theta+\frac{\Delta}{\sqrt{N}}j\right)-\xi\right]\middle| Z(k,\theta)=\xi\right\}=S_i+V_j,$$

where $S_i=\sum_{l=1}^{i}a_l$, with

$$a_l\sim\mathcal{N}\left(-\frac{\mu(k,\theta)}{2(N-k)}b^2,\frac{\mu(k,\theta)}{N-k}b^2\right),$$

and

$$V_j=\sqrt{2}\gamma\frac{b}{\sqrt{N}}\Delta jV-\gamma^2\frac{b^2}{N}\Delta^2 j^2$$

with $V\sim\mathcal{N}(0,1)$.

By Lemma 2, using the techniques in [5] and [4], we have the conditional probability can be written in terms of decomposed random processes:

$$\begin{aligned}&\mathbb{P}\left\{\max_{(i,j)\in J(i_0,j_0)} b\left[Z\left(i,j\frac{\Delta}{\sqrt{N}}\right)-Z\left(i_0,j_0\frac{\Delta}{\sqrt{N}}\right)\right]\leq -x\middle|\right.\\&\qquad\left.Z\left(i_0,j_0\frac{\Delta}{\sqrt{N}}\right)=b+\frac{x}{b}\right\}\\&\approx\mathbb{P}\left\{\max_{i\geq 1}S_i\leq -x\right\}\mathbb{P}\left\{\max_{i\leq 0}S_i+\max_{j\geq 1}V_j\leq -x\right\}\end{aligned} \tag{25}$$

A similar argument for (25) can be found in [5] and [4].

5) Limit by Shrinking Δ: Put this together with approximation for (19) in (21), the approximate significant level becomes

$$\begin{aligned}&\mathbb{P}\left\{\max_{(i,j)\in D} Z\left(i,j\frac{\Delta}{\sqrt{N}}\right)\geq b\right\}\\&\approx\sum_{(i_0,j_0)\in D} g\left(i_0,j_0\frac{\Delta}{\sqrt{N}}\right)\frac{\Delta}{\sqrt{N}}\cdot\\&\quad\frac{\sqrt{N}}{\Delta b}\int_0^\infty\exp\left(-\frac{\xi_0}{b}x\right)\mathbb{P}\left\{\max_{i\geq 1}S_i\leq -x\right\}\\&\quad\mathbb{P}\left\{\max_{i\leq 0}S_i+\max_{j\geq 1}V_j\leq -x\right\}dx.\end{aligned} \tag{26}$$

The following Lemma, which is an extension of Lemma 2 in [4], enables us to find an expression for integration over x in (18):

Lemma 3: Assume $x_1,x_2,\cdots$ i.i.d. $\mathcal{N}\left(-\mu_1,\sigma_1^2\right)$, with $\mu_1>0$. Define the random walk $S_0=0$, $S_i=\sum_{l=1}^{i}x_l$, $i=1,2,\cdots$, and the smooth varying random process $V_j=\beta\Delta jV-\frac{\beta^2}{2}\Delta^2 j^2$, for some constants $\Delta>0$, $\beta>0$. As $\Delta\to 0$, for some constant $\alpha>0$, we have

$$\begin{aligned}&\frac{1}{\Delta}\int_0^\infty e^{-\alpha x}\mathbb{P}\left\{\max_{i\geq 1}S_i\leq -x\right\}\\&=\mathbb{P}\left\{\max_{i\leq 0}S_i+\max_{j\geq 1}V_j\leq -x\right\}dx\\&\xrightarrow{\Delta\to 0}\frac{|\beta|}{\sqrt{2\pi}}\left(\frac{2\mu_1^2}{\sigma_1^2}\right)\nu\left(\frac{2\mu_1}{\sigma_1}\right).\end{aligned} \tag{27}$$

where $\nu(x)$ is defined in (13).

Finally, using Lemma 3 for (26) with $\alpha=\frac{\xi_0}{b}$, $\beta=\sqrt{2}\gamma\frac{b}{\sqrt{N}}$, $\mu_1=\frac{\mu(k,\theta)}{2(N-k)}b^2$ and $\sigma_1^2=\frac{\mu(k,\theta)}{N-k}b^2$ we have the approximate significance level

$$\begin{aligned}&\frac{1}{2\sqrt{\pi}}\sum_{(i_0,j_0)\in D} g\left(i_0,j_0\frac{\Delta}{\sqrt{N}}\right)\cdot\frac{b^2\mu(i_0,j_0\frac{\Delta}{\sqrt{N}})}{N-i_0}\\&\nu\left(\sqrt{\frac{b^2\mu\left(i_0,j_0\frac{\Delta}{\sqrt{N}}\right)}{N-i_0}}\right)\gamma\left(i_0,j_0\frac{\Delta}{\sqrt{N}}\right)\frac{\Delta}{\sqrt{N}}.\end{aligned} \tag{28}$$

As $\Delta\to 0$, the Riemann sum (28) converges to the approximation in Theorem 1.

The proof is more complex when the number of parameters $d>1$, and we have to use different techniques, such as computing a Mill's ratio type of expression by decomposing the process as a random walk plus a smooth Gaussian field. The complete proof can be found in [6].

REFERENCES

[1] L. Lai, Y. Fan, and H. V. Poor, "Quickest detection in cognitive radio: a sequential change detection framework," *IEEE GLOBECOM*, 2008.

[2] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*. Springer Series in Statistics, Springer, 1985.

[3] D. Siegmund and B. Yakir, *The statistics of gene mapping*. Springer, 2007.

[4] H.-J. Kim and D. Siegmund, "The likelihood ratio test for a change-point in simple linear regression," *Biometrica*, vol. 76, no. 3, pp. 409–423, 1989.

[5] D. Siegmund, "Approximate tail probabilities for the maxima of some random fields," *The Annuals of Probability*, vol. 16, no. 2, pp. 487–501, 1988.

[6] Y. Xie and D. Siegmund, "Change-point detection of weak signals: How to use signal correlation?," tech. rep., 2010.