A FULL GENERALIZED LIKELIHOOD RATIO TEST FOR SOURCE DETECTION

Pei-Jung Chung[1)], *Kon Max Wong*[2)]

[1)]Institute for Digital Communications, The University of Edinburgh, UK
[2)] Dept. of Electrical and Computer Engineering, McMaster University, Canada
p.chung@ed.ac.uk, wong@mail.ece.mcmaster.ca

ABSTRACT

This work presents a novel full generalized likelihood ratio test (GLRT) for signal detection in a sensor array environment. The multiple hypothesis test approach is well known to have excellent detection performance among several popular methods. Existing multiple test procedures consider the relation between two adjacent models. When the number of signals or the assumed number of signals is large, it tends to overestimate the number of signals. The proposed full GLRT procedure overcomes this disadvantage by employing complete information between candidate models and leads to gain in test power. A further advantage is that a confidence interval for the true number of signals can be constructed based on the outcome of the GLRT procedure. Numerical results show that the full GLRT procedure improves detection performance significantly in comparison with existing multiple test based approaches in challenging scenarios.

***Index Terms*—** array processing, full generalized likelihood ratio test, signal detection, model order selection, confidence interval

1. INTRODUCTION

This work discusses signal detection using a full generalized likelihood ratio test. Estimating the number of signals embedded in noisy observations is a key issue in array processing, harmonic retrieval, wireless communication and many other applications. It has attracted lots of research interest over last decades [2] [9] [10]. In early works [2] [7], a multiple testing procedure was suggested to determine the number of signals sequentially. In more recent works [3] [4] along this line, the false discovery rate (FDR) was introduced to improve the test power and probability of correct detection. Compared to another popular information theoretic criterion based approach [9] [10], the multiple testing procedure has better performance under critical conditions such as low SNR and small samples.

A common feature of the afore mentioned multiple testing procedures is that each test considers only adjacent models in the hierarchy of candidate models and detect signals in a stepwise manner. In this paper, we suggest a full generalized likelihood ratio test that exploit more information between models and further improve detection performance. The full GLRT was first proposed by Hosoya for model identification [5] and was shown to be more favorable than the standard χ^2-test. Here, we adopt the idea of comparing the candidate model with all models of higher order and construct a new test procedure for signal detection. The proposed procedure is not only more powerful than existing stepwise approaches, it also provides means of constructing confidence interval for the true number of signals.

In the following, we will give a brief description of data model. Then the full GLRT procedure for signal detection is developed in Section 3. Section 4 is devoted to derivation of test statistics and p-values. Simulation results are presented in Section 5. Finally, concluding remarks are given in Section 6.

2. PROBLEM FORMULATION

Consider an array of n sensors receiving m narrow band signals emitted by far-field sources located at $\boldsymbol{\theta}_m = [\,\theta_1,\ldots,\theta_m]^T$. The array output $\boldsymbol{x}(t) \in \mathbb{C}^{n\times 1}$ is expressed as

$$\boldsymbol{x}(t) = \boldsymbol{H}_m(\boldsymbol{\theta}_m)\boldsymbol{s}_m(t) + \boldsymbol{n}(t), \quad t = 1,\ldots,T \tag{1}$$

where the ith column of the array response matrix

$$\boldsymbol{H}_m(\boldsymbol{\theta}_m) = [\boldsymbol{d}(\theta_1)\cdots\boldsymbol{d}(\theta_i)\cdots\boldsymbol{d}(\theta_m)] \tag{2}$$

$\boldsymbol{d}(\theta_i) \in \mathbb{C}^{n\times 1}$ is the steering vector associated with the signal arriving from the direction θ_i. The signal waveform $\boldsymbol{s}_m(t) = [s_1(t),\ldots,s_m(t)]^T \in \mathbb{C}^{m\times 1}$ is considered as deterministic and unknown. The noise $\boldsymbol{n}(t) \in \mathbb{C}^{n\times 1}$ is independent, identically complex normally distributed with zero mean and covariance matrix $\sigma^2\boldsymbol{I}$, where σ^2 is the unknown noise spectral parameter and $\boldsymbol{I}$ is an identity matrix of corresponding dimension. Given the observations $\{\boldsymbol{x}(t)\}_{t=1}^T$, the problem of central interest is to determine the number of signals, m.

3. SIGNAL DETECTION USING A FULL GENERALIZED LIKELIHOOD RATIO TEST

We formulate source detection as a model order selection problem. Suppose M is the maximal number of signals. The following hierarchical structure represents models corresponding to increasing numbers of signals $m = 0, \cdots, M$:

$$\mathcal{M}_0 \subset \mathcal{M}_1 \cdots \subset \mathcal{M}_m \cdots \subset \mathcal{M}_M, \tag{3}$$

where $\mathcal{M}_0$ denotes the noise only case. Each model class $\mathcal{M}_i$ is associated with the null hypothesis:

$$H_i : \boldsymbol{x}(t) = \boldsymbol{H}_i(\boldsymbol{\theta}_i)\boldsymbol{s}_i(t) + \boldsymbol{n}(t). \tag{4}$$

To validate $\mathcal{M}_i$ $(i < M)$, H_i is tested against all models of higher orders, H_j, $i < j \leq M$. If H_i is retained for one of these tests, $\mathcal{M}_i$ becomes candidate for the best model. In the sequential test, whether $\mathcal{M}_i$ remains as candidate model solely depends on the outcome of testing against $\mathcal{M}_{i+1}$. Apparently, the proposed full test increases the probability for $\mathcal{M}_i$ to be considered as candidate. In particular, when $\mathcal{M}_i$ is the true model, the probability of correct detection becomes larger. Finally, after the set of candidate models is determined, the model order is estimated by the one with the lowest complexity.

Mathematically, this procedure can be described as follows. Let T_{ij} and p_{ij} denote the test statistic and the corresponding p-value (observed significance value) when testing H_i against H_j, respectively. By definition, the p-value is computed as $p_{ij} = 1 - P_{H_i}(T_{ij})$ where $P_{H_i}(\cdot)$ is the cumulative distribution function under the null hypothesis H_i. H_i is rejected if $p_{ij} < \alpha$ where α is a pre-specified significance level. In the procedure described above, H_i is tested against H_j, $j = i+1, \ldots, M$. It is considered as a candidate model if H_i is not rejected in one of the tests. This implies that the minimum p-value

$$p_i = \min\{p_{ij} \,|j = i+1, i+2, \ldots, M\} \tag{5}$$

must be equal to or larger than α for $\mathcal{M}_i$ to become a candidate model, i.e. $p_i \geq \alpha$.

Let $\{H_{i_1}, H_{i_2}, \cdots, H_{i_r}\}$ denote the retained hypotheses where $I = \{i_1, i_2, \cdots, i_r\}$ is a subset of $\{1, 2, \cdots, M\}$. The number of signals (or model order) corresponds to the one with the smallest index:

$$\hat{m} := \min\{i_1, i_2, \cdots, i_r\}. \tag{6}$$

As pointed out in [5], the above procedure also provides means of constructing confidence interval of model order. In other words, if i_0 is the index corresponding to the true model $\mathcal{M}_{i_0}$, then

$$\Pr\{i_0 \in I | i_0\} = 1 - \alpha. \tag{7}$$

The above property follows directly from the construction of the procedure. It was shown for model identification problem that the smallest index $\hat{m}$ is a consistent estimator for i_0 under mild conditions on significance level α.

4. TEST STATISTICS

The test statistics for the proposed procedure are derived from the generalized likelihood ratio principle. Let $\hat{\boldsymbol{\theta}}_m$ denote the ML estimate obtained from minimizing the negative concentrated likelihood function:

$$\hat{\boldsymbol{\theta}}_m = \arg\min_{\boldsymbol{\theta}_m} \mathrm{tr}[(\boldsymbol{I} - \boldsymbol{P}_m(\boldsymbol{\theta}_m))\hat{\boldsymbol{C}}_x], \tag{8}$$

where $\boldsymbol{P}(\boldsymbol{\theta}_m)$ represents the projection matrix onto the subspace spanned by the columns of $\boldsymbol{H}_m(\boldsymbol{\theta}_m)$ and $\hat{\boldsymbol{C}}_x = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}(t)\boldsymbol{x}(t)^H$ is the sample covariance matrix.

We apply the likelihood ratio (LR) principle to obtain the test statistic T_{ij} $(i = 1, \cdots, (M-1), j = i+1, \cdots, M)$:

$$T_{ij} = \log\left(\frac{\mathrm{tr}[(\boldsymbol{I} - \boldsymbol{P}_i(\hat{\boldsymbol{\theta}}_i))\hat{\boldsymbol{C}}_x]}{\mathrm{tr}[(\boldsymbol{I} - \boldsymbol{P}_j(\hat{\boldsymbol{\theta}}_j))\hat{\boldsymbol{C}}_x]}\right) \tag{9}$$

$$= \log\left(1 + \frac{n_1}{n_2}F_{ij}\right), \tag{10}$$

$$F_{ij} = \frac{n_2}{n_1}\frac{\mathrm{tr}[(\boldsymbol{P}_j(\hat{\boldsymbol{\theta}}_j) - \boldsymbol{P}_i(\hat{\boldsymbol{\theta}}_i))\hat{\boldsymbol{C}}_x]}{\mathrm{tr}[(\boldsymbol{I} - \boldsymbol{P}_j(\hat{\boldsymbol{\theta}}_j))\hat{\boldsymbol{C}}_x]}. \tag{11}$$

When $i = 1$, we define $\boldsymbol{P}_0(\cdot) = \boldsymbol{0}$.

Since T_{ij} is a monotonic increasing function of F_{ij}, the test can be equivalently conducted by using F_{ij}. Under the null hypothesis H_i, the statistic F_{ij} is asymptotically F_{n_1,n_2}-distributed. The degrees of freedom n_1, n_2 are given by

$$n_1 = (2(j-i) + (r_{i+1} + \cdots + r_j))\,T, \tag{12}$$

$$n_2 = (2n - 2j - (r_{i+1} + \cdots + r_j))\,T, \tag{13}$$

where $r_i = \dim(\theta_i)$ denote the number of the nonlinear parameters associated with the ith signal. This can be verified by taking Taylor expansion around the true nonlinear parameters [7]. For the underlying DOA parameter, $r_i = 1$, we have

$$n_1 = 3(j-i)T, \quad n_2 = (2n - 2j - (j-i))T. \tag{14}$$

More details about the F_{n_1,n_2}-distribution can be found in [6]. Let $\mathcal{F}_i$ represent the cumulative distribution function of F_{n_1,n_2}, then p-value for testing H_i against H_j is given by

$$p_{ij} = 1 - \mathcal{F}_i(F_{ij}). \tag{15}$$

Replacing the p-values obtained from (15) in (5), the number of signals can be estimated by (6). To summarize, all steps of the proposed full GLRT procedure are listed in **Algorithm 1**.

4.1. Comparison with Existing Tests

In the existing hypothesis tests for signal detection such as those discussed in [3],[4], the candidate model $\mathcal{M}_i$ is only tested against the next model in the hierarchy $\mathcal{M}_{i+1}$ (see (3)). In other words, H_i is considered as a candidate model if it passes this single test. On the contrary, the proposed full GLRT procedure utilizes information collected from comparing $\mathcal{M}_i$ with all model classes of higher orders. It is validated if it is retained in any of these tests. Therefore, it can achieve powerful results than existing approaches. Moreover, the set of hypotheses that are not rejected constitute a confidence interval for the true number of signals. Despite of the gain in power, the proposed approach only requires a few extra steps in computing p-values since the main computational burden lies in finding the ML estimates.

5. SIMULATION RESULTS

The proposed full GLRT procedure is tested by numerical experiments. In the simulation, a uniform linear array of 10 sensors with inter-element spacings of half a wavelength is employed. The narrow band signals are generated by $m = 3$ uncorrelated signals located at $\boldsymbol{\theta}_3 = [-30° \; 20° \; 24°]$ relative to the broadside. Note that two signals are closely located. The signal-to-noise ratio (SNR) of the strongest signal runs from $[-8 : 1 : 6]$ dB in a 1 dB step. The maximal number of signals $M_{max} = 6$. Each experiment performs 300 trials. As comparison, the multiple test [3] is applied to the same data. The significance level α of the full GLRT and the FDR level q for the multiple test are both kept at 0.1.

In the first experiment, all signals are of equal strengths; namely, the difference in SNR is $[0 \; 0 \; 0]$ dB. We also test by various number of snapshots, $T = 30, 60$, respectively. The empirical probability of correct detection is illustrated in Fig. 1. By *correct detection*, it is meant that $\hat{m} = m$, the correct number of signals. For the same number of snapshots, the full GLRT always outperforms the multiple hypothesis test. In the threshold region from -4 to 0 dB, the gap in probability of detection can be as large as 14%. Both algorithms perform better with larger number of samples at $T = 60$. At SNR > 3 dB, they all achieve 100% of correct detection.

As mentioned previously, the proposed procedure also provides a confidence interval to the level of $(1 - \alpha)$ of the true number of signals. As an example, it was observed in one trial at $T = 30$, SNR $= -2$dB, it was observed that the hypotheses $\{H_3, H_4, H_5, H_6\}$ were not rejected. Then the index set $\{3, 4, 5, 6\}$ constitute of 90% confidence interval for the true number of signals.

In the second experiment, the signal strength of one signal is weaker than other two. The difference in SNR $[0 \; -1 \; 0]$ dB. This implies that it becomes more difficult to detect a weak signal close to a strong one. As shown in Fig. 2, both approaches show lower probability of correct detection. However, the full GLRT remains the superior one. The gain of using the proposed procedure is as high as 12% at SNR $= -2$dB for $T = 30$. For $T = 60$, similar performance can be observed from Fig. 2.

In summary, the proposed full GLRT has an overall better performance than the multiple test procedure. In particular, the former leads to a significant gain in the threshold region. As simulation results shown in [3], the multiple test outperforms the information theoretic criterion based MDL approach in various scenarios. One could expect that the full GLRT procedure would provide higher detection capability than the MDL approach.

Input: array observations $\{\boldsymbol{x}(t), t = 1, \ldots, T\}$
an upper bound on the number of signals M,
significance level α.

1. for $i = 1, \ldots, M$
   Find the ML estimate $\hat{\boldsymbol{\theta}}_i$
   end;
2. for $i = 1, \ldots, M - 1$
   (a) for $j = i + 1, \cdots, M$
   compute test statistic F_{ij} and p_{ij}
   end;
   (b) Find the minimum
   $p_i = \min\{p_{ij} \, | j = i + 1, i + 2, \ldots, M\}$.
   (c) Validate H_i
   H_i is retained if $p_i \geq \alpha$
   H_i is rejected if $p_i < \alpha$.
   end;
3. Let $\{H_{i_1}, H_{i_2}, \cdots, H_{i_r}\}$ be hypotheses that are not rejected. The number of signals is estimated by
   $\hat{m} = \min\{i_1, i_2, \ldots, i_r\}$

Output: estimated number of signals $\hat{m}$,
$(1 - \alpha)$ -confidence interval: $\{i_1, i_2, \ldots, i_r\}$

Algorithm 1: A Full GLRT for Source Detection

6. CONCLUSION

In this paper, we have considered a full generalized ratio test for detecting signals embedded in noisy sensor array data. The proposed approach tests each of the nested models against all models of higher order. Among the models that pass the tests, the one with lowest complexity determines the

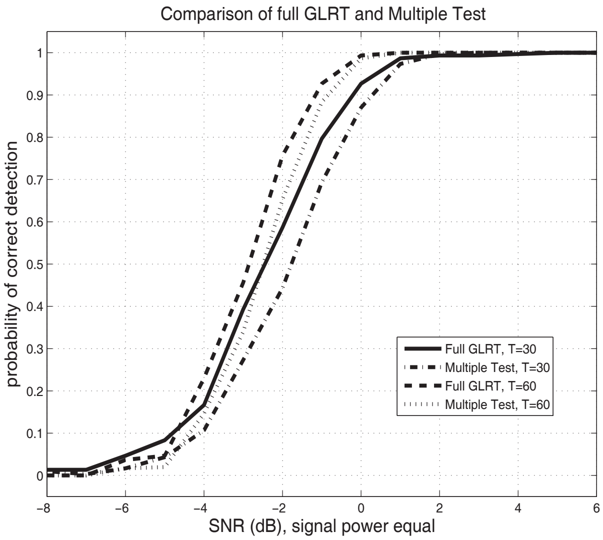

Fig. 1. Probability of correct detection. $m = 3$, $M_{max} = 10$. Reference DOA parameter $\boldsymbol{\theta}_3 = [-30° \; 20° \; 24°]$, SNR $= [-8 : 1 : 6]$ dB, SNR difference $= [0 \; 0 \; 0]$ dB.

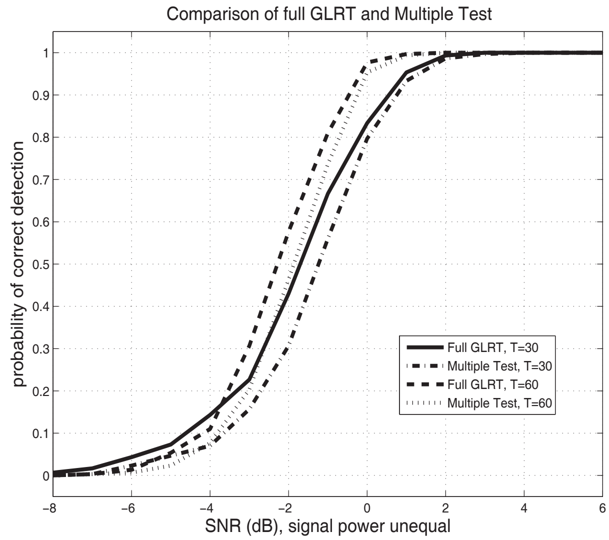

Fig. 2. Probability of correct detection. $m = 3$, $M_{max} = 10$. Reference DOA parameter $\boldsymbol{\theta}_3 = [-30° \; 20° \; 24°]$, SNR $= [-8 : 1 : 6]$ dB, SNR difference $= [0 \; -1 \; 0]$ dB.

number of signals. Compared to the multiple test based approach in [3] which considers the relation between adjacent models, more information between candidate models is employed in the current procedure. It is no surprise that the full GLRT approach outperforms the multiple test significantly when the latter fails to provide reliable results. Furthermore, the proposed test provides meas of constructing confidence interval of the true number of signals. Simulation shows that in the critical case involving closely located source and various strengths, the full GLRT the probability of correct detection can be improved by 12% by the full GLRT approach. Given the fact that the computational complexity of both algorithms is of the same order, we believe that the full GLRT is a promising alternative to the multiple test for source enumeration of sensor array signals.

7. REFERENCES

[1] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, (57):289-300, 1995.

[2] J. F. Böhme. Statistical array signal processing of measured sonar and seismic data. In *Proc. SPIE 2563 Advanced Signal Processing Algorithms*, pp. 2-20, San Diego, USA, Jul 1995.

[3] P.-J. Chung and J. F. Böhme and C. F. Mecklenbräuker and A. O. Hero. Detection of the number of sigansl using the Benjamini-Hochberg procedure. *IEEE Trans. Signal Processing*, 55(6), pp. 2497-2508, June 2007.

[4] Pei-Jung Chung, Dirk Maiwald, Nicolai Czink, Christoph F. Mecklenbräuker, Bernard H. Fleury: *Determining the Number of Propagation Paths from Broadband MIMO Measurements via Bootstrapped Likelihood Statistics and the False Discovery Criterion - Part I Methodology*, Proc. IEEE CAMSAP 2007.

[5] Y. Hosoya. Hierarchical Statistical models and generalized likelihood ratio test. *J.R Statist. Soc. B*, (51), No.3, pp. 435-447, 1989.

[6] N. L. Johnson and S. Kotz and N. Balakrishnan. *Continuous univariate distributions 2nd ed.*, New York, London : Wiley & Sons, 1994.

[7] D. Maiwald, *Breitbandverfahren zur Signalentdeckung und -ortung mit Sensorgruppen in Seismik- und Sonaranwendungen*. Dr.-Ing. Dissertation, Dept. of Electrical Engineering, Ruhr-Universität Bochum, Germany. Shaker Verlag, 1995.

[8] R. H. Shumway. Replicated time-series regression: an approach to signal estimation and detection. In *Handbook of Statistics, Vol. 3*, Editor: D. R. Brillinger and P. R. Krishnaiah, Elsevier Science Publishers B.V., pp. 383-408, 1983.

[9] M. Wax and T. Kailath, *Detection of signals by information theoretic criteria,* IEEE Trans. ASSP, vol. 33, no. 2, pp. 387-392, April 1985.

[10] M. Wax, *Detection and localization of multiple sources via the stochastic signals model,* IEEE Trans. Signal Processing, vol. 39, no. 11, pp. 2450-2456, Nov. 1991.