DETECTING OUTLIERS IN THE ESTIMATOR BANK-BASED DIRECTION FINDING TECHNIQUES USING THE LIKELIHOOD RATIO QUALITY ASSESSMENT

Pouyan Parvazi *Alex B. Gershman*

Communication Systems Group
Technical University of Darmstadt
64283 Darmstadt, Germany

Yuri I. Abramovich

Defence Science and Technology Organization
PO Box 1500, Edinburgh
SA 5111, Australia

ABSTRACT

Estimator bank-based direction finding techniques make use of a number of parallel randomly weighted MUSIC direction-of-arrival (DOA) estimates and obtain the final estimate by keeping only the "successful" candidates while sorting out the outlying estimates. In this paper, we develop a powerful approach to detect the outliers in the estimator bank-based direction finders using the likelihood ratio quality assessment. Computer simulations show substantial improvements of the proposed approach as compared to the earlier techniques used to sort out the outliers in the estimator bank-based direction finding methods.

Index Terms— Array signal processing, direction finding, estimator banks

1. INTRODUCTION

One of main problems that limit the performance of subspace-based direction finding methods at low signal-to-noise ratios (SNRs) or small number of snapshots is the so-called threshold effect caused by outliers in DOA estimates [1], [2]. To improve the threshold performance of subspace-based techniques, several methods have been proposed [3–7]. The approaches of [3–5] use a set of different parallel DOA estimates to form the so-called *estimator bank*. Then, only the "successful" estimates are chosen, while the outlying estimates are dropped. To detect the "successful" estimates among the estimates available in the estimator bank, it is proposed in [3] to use a preliminary information about the angular sectors where the sources are located. Unfortunately, the information about source angular sectors is not always available or may be imprecise. In [4] and [5], it is proposed to substitute the estimated DOAs resulting from multiple parallel estimates into the likelihood function and then pick the estimate with the highest value of this function as the final DOA estimate.

In this paper, we develop an alternative approach to detect and select the "successful" estimates in the estimator bank techniques using the likelihood ratio quality assessment [8]. Our computer simulations validate substantial improvements of the proposed approach as compared to the earlier outlier detection techniques used in estimator banks.

2. PROBLEM FORMULATION

Consider an array of M sensors receiving L ($L < M$) narrowband signals from mutually uncorrelated far-field sources. The number of sources is assumed to be known throughout the paper. The array outputs can be modeled as

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t) \tag{1}$$

where

$$\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \cdots, \mathbf{a}(\theta_L)] \tag{2}$$

is the $M \times L$ direction matrix, $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_L]^T$ is the $L \times 1$ vector of source DOAs, $\mathbf{s}(t)$ is the $L \times 1$ vector of random source waveforms, $\mathbf{n}(t)$ is the $M \times 1$ vector of zero-mean Gaussian sensor noise with the variance σ^2 in each sensor, and $(\cdot)^T$ denotes the transpose.

Using (1), the covariance matrix of the array outputs can be written as

$$\mathbf{R} = E[\mathbf{y}(t)\mathbf{y}^H(t)] = \mathbf{A}\mathbf{S}\mathbf{A}^H + \sigma^2\mathbf{I}_M \tag{3}$$

where $\mathbf{S} = E[\mathbf{s}(t)\mathbf{s}^H(t)]$ is the $L \times L$ source covariance matrix.

The sample covariance matrix can be expressed as

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}(t)\mathbf{y}^H(t)\,. \tag{4}$$

The eigendecompostion of $\hat{\mathbf{R}}$ can be written in the following form

$$\hat{\mathbf{R}} = \hat{\mathbf{U}}_{\rm S}\hat{\mathbf{\Lambda}}_{\rm S}\hat{\mathbf{U}}_{\rm S}^H + \hat{\mathbf{U}}_{\rm N}\hat{\mathbf{\Lambda}}_{\rm N}\hat{\mathbf{U}}_{\rm N}^H \tag{5}$$

where the $L \times L$ and $(M-L) \times (M-L)$ diagonal matrices $\hat{\mathbf{\Lambda}}_{\rm S}$ and $\hat{\mathbf{\Lambda}}_{\rm N}$ contain, respectively, the L and $(M-L)$ signal- and noise-subspace eigenvalues of $\hat{\mathbf{R}}$, and the columns of the $M \times L$ and $M \times (M-L)$ matrices $\hat{\mathbf{U}}_{\rm S}$ and $\hat{\mathbf{U}}_{\rm N}$ contain the corresponding eigenvectors.

3. ESTIMATOR BANK-BASED DIRECTION FINDING

The concept of estimator banks [3] is, given the sample covariance matrix $\hat{\mathbf{R}}$, to generate a set of K parallel candidate DOA estimates

$$\mathcal{F} = \{f_k(\theta)\}_{k=1}^{K} \tag{6}$$

and then to pass each of these estimates through some outlier detection procedure. As a result, those DOA estimates that are detected to contain outliers are dropped, whereas the "successful" estimates (that pass the test) are combined to obtain the final DOA estimate.

It has been proposed in [3] to generate the candidate estimates in (6) using the weighted MUSIC function

$$f(\theta) = \frac{1}{\mathbf{a}^H(\theta)\hat{\mathbf{U}}_{\mathrm{N}}\mathbf{W}\hat{\mathbf{U}}_{\mathrm{N}}^H\mathbf{a}(\theta)} \tag{7}$$

with rank-one weighting matrices $\mathbf{W} = \mathbf{w}_k\mathbf{w}_k^H$ where $\mathbf{w}_k$ is an $L \times 1$ random vector drawn from the complex Gaussian distribution. Then, the underlying estimates in (6) can be expressed as

$$f_k(\theta) = \frac{1}{|\mathbf{a}^H(\theta)\hat{\mathbf{U}}_{\mathrm{N}}\mathbf{w}_k|^2}\,. \tag{8}$$

There have been several approaches to formulate the test to detect the outlying estimates. In [3], it has been proposed to preestimate the source angular sectors (for example, using conventional beamforming methods) and then to reject or approve each particular candidate estimate $f_k(\theta)$ based on the number of peaks in the aforementioned angular sectors. Unfortunately, preliminary estimates of the angular sectors are not always available in practice, and they may be erroneous in the case when the sources have significantly different powers.

Another popular approach to detect outliers in multiple parallel estimates is to obtain the source DOAs from each candidate function $f_k(\theta)$ and then to substitute them into the likelihood function, always picking the DOAs that yield the highest value of this function, see [4–6].

In the next section, we propose a new competitive approach to detect outliers in estimator banks that offers significant advantages as compared to the existing outlier detection techniques. In particular, it does not require any preliminary estimates of the source angular sectors as the approach of [3], and provides better threshold performance than the approach of [4]. Moreover, in contrast to the methods of [4] and [5], it enables to predict how reliable is the final DOA estimate (i.e., to decide whether the final DOA estimate itself is an outlier).

4. DETECTING OUTLIERS USING LIKELIHOOD RATIO TEST

The proposed approach is based on the computation of the likelihood ratio for each estimate in (6) using the sphericity test [8]. The key idea behind this test is, first of all, to estimate the source DOAs and the source and noise powers, and then to compare an artificially built covariance matrix $\tilde{\mathbf{R}}$ (which is recovered from the preliminary estimated DOAs and signal/noise powers, as well as known array manifold) and the sample covariance matrix $\hat{\mathbf{R}}$.

To calculate $\tilde{\mathbf{R}}$, one can start with the DOA estimates $\hat{\boldsymbol{\theta}}$ obtained from each $f_k(\theta)$ in the standard manner. Using these DOA estimates, we obtain the estimate of the direction matrix as $\hat{\mathbf{A}} = \mathbf{A}(\hat{\boldsymbol{\theta}})$. The noise power is estimated by averaging the noise-subspace eigenvalues, and then the source covariance matrix is estimated as

$$\hat{\mathbf{S}} = \hat{\mathbf{A}}^{\dagger}(\hat{\mathbf{R}} - \hat{\sigma^2}\mathbf{I}_M)\hat{\mathbf{A}}^{\dagger H} \tag{9}$$

where $(\cdot)^{\dagger}$ stands for the pseudo-inverse.

With these estimates, $\tilde{\mathbf{R}}$ can be computed as

$$\tilde{\mathbf{R}} = \hat{\mathbf{A}}\hat{\mathbf{S}}\hat{\mathbf{A}}^H + \hat{\sigma^2}\mathbf{I}_M. \tag{10}$$

Applying the sphericity test with the following hypothesis [8]

$$\begin{aligned} H_0: \quad & \mathrm{E}\{\tilde{\mathbf{R}}^{-1/2}\hat{\mathbf{R}}\tilde{\mathbf{R}}^{-1/2}\} = c\mathbf{I}_M \quad \text{against} \\ H_1: \quad & \mathrm{E}\{\tilde{\mathbf{R}}^{-1/2}\hat{\mathbf{R}}\tilde{\mathbf{R}}^{-1/2}\} \neq c\mathbf{I}_M \end{aligned} \tag{11}$$

for $c \geq 0$, we have that the likelihood ratio is given by

$$\gamma(\tilde{\mathbf{R}}) = \left(\frac{\det(\tilde{\mathbf{R}}^{-1}\hat{\mathbf{R}})}{\left[\frac{1}{M}\mathrm{tr}(\tilde{\mathbf{R}}^{-1}\hat{\mathbf{R}})\right]^M}\right)^N \triangleq \gamma_0^N(\tilde{\mathbf{R}}). \tag{12}$$

The quality of each particular DOA estimate in (6) can be accessed by examining the value of γ_0 that shows the reliability of the estimate tested.

As pointed out in [7] and [8], the values of γ_0 corresponding to non-outlying DOA estimates should exceed $\gamma_0(\mathbf{R})$ (that is, the value of γ_0 taken at the true covariance matrix). It means that for a proper $\tilde{\mathbf{R}}$ we should have

$$\gamma_0(\tilde{\mathbf{R}}) \geq \gamma_0(\mathbf{R}). \tag{13}$$

Unfortunately, in practical cases we do not have access to the true covariance matrix $\mathbf{R}$, i.e., the value of $\gamma_0(\mathbf{R})$ is unknown. However, we can use the fact that the statistical distribution of $\gamma_0(\mathbf{R})$ is not scenario-dependent. Indeed it only depends on the parameters N and M because [8]

$$\gamma_0(\mathbf{R}) = \frac{\det\hat{\mathbf{C}}}{\left[\frac{1}{M}\mathrm{tr}\hat{\mathbf{C}}\right]^M} \tag{14}$$

where $\hat{\mathbf{C}} \triangleq \mathbf{R}^{-1/2}\hat{\mathbf{R}}\mathbf{R}^{-1/2} \sim \mathcal{CW}(M, N; \mathbf{I}_M)$ and $\mathcal{CW}$ stands for the central complex Wishart distribution. Using this fact, the probability density function (pdf) $p(\gamma_0(\mathbf{R}))$ can

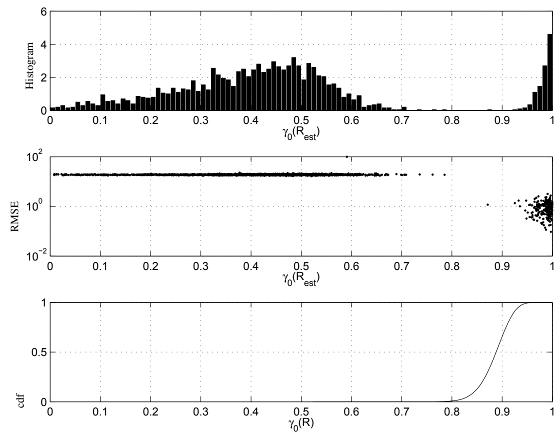

Fig. 1. Illustration of the likelihood ratio-based test (16) for $\mathrm{SNR} = 10$ dB. The upper subplot shows the histogram of $\gamma_0(\tilde{\mathbf{R}})$. The middle subplot shows the scatter plot of the RMSEs of the MUSIC DOA estimates versus $\gamma_0(\tilde{\mathbf{R}})$. The lower subplot displays the cdf of $\gamma_0(\mathbf{R})$.

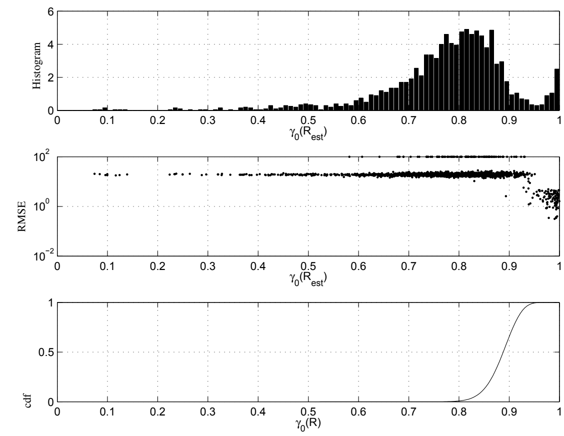

Fig. 2. Illustration of the likelihood ratio-based test (16) for $\mathrm{SNR} = 5$ dB. The upper subplot shows the histogram of $\gamma_0(\tilde{\mathbf{R}})$. The middle subplot shows the scatter plot of the RMSEs of the MUSIC DOA estimates versus $\gamma_0(\tilde{\mathbf{R}})$. The lower subplot displays the cdf of $\gamma_0(\mathbf{R})$.

be computed [8]. In particular, we can compute the following confidence interval:

$$P(\gamma_0(\mathbf{R}) \leq \alpha) = \int_0^{\alpha} p(\gamma_0(\mathbf{R}))\, d\gamma_0 = P_\alpha \qquad (15)$$

where $P(\cdot)$ stands for the cumulative distribution function (cdf) and P_α is the probability to be selected to obtain the threshold α.

After computing the threshold α, this value can be used in lieu of $\gamma_0(\mathbf{R})$ in (13). Thus, we have the following likelihood ratio-based quality assessment test that amounts to checking whether the inequality

$$\gamma_0(\tilde{\mathbf{R}}) \geq \alpha \qquad (16)$$

is satisfied. If the inequality is not satisfied for any particular DOA estimate from the estimator bank, then this estimate is identified as outlier and is dropped. The final estimate is then formed in a regular way using only the "successful" estimates that have passed the test [3].

5. SIMULATIONS

In our first example, we assume a uniform linear array (ULA) of $M = 5$ omnidirectional sensors spaced half a wavelength apart, and $L = 3$ uncorrelated equipower sources located at $-25°$, $0°$ and $4°$ relative to the array broadside direction. We also assume that $N = 100$. For each simulation run (with a total number of runs being equal to 2000), the DOAs are estimated based on the conventional spectral MUSIC algorithm and then the value of $\gamma_0(\tilde{\mathbf{R}})$ is calculated. The histograms of this value are shown for $\mathrm{SNR} = 10$ dB and $\mathrm{SNR} = 5$ dB in the top subplots of Figs. 1 and 2, respectively. The middle subplots of Figs. 1 and 2 display the scatter plots of root-mean-square errors (RMSEs) of the MUSIC DOA estimates versus γ_0 for $\mathrm{SNR} = 10$ dB and $\mathrm{SNR} = 5$ dB, respectively. The bottom subplots of Figs. 1 and 2 display the cdf's of $\gamma_0(\mathbf{R})$ (which are the same for both values of SNR as they are scenario-independent).

Figs. 1 and 2 clearly verify that the developed outlier test is adequate because the outlying estimates (that correspond to the high RMSE values in the middle subplots) can be sorted out by setting a proper threshold α. For example, from the bottom subplots of Figs. 1 and 2 we can see that the choice of $P_\alpha = 0.5$ (which corresponds to the threshold value $\alpha = 0.88$) is quite appropriate.

In our second example, we assume a ULA of $M = 10$ omnidirectional sensors spaced half a wavelength apart, and $L = 2$ uncorrelated equipower sources located at $10°$ and $15°$ relative to the array broadside direction. We also assume that the dimension of estimator bank is $K = 20$. Fig. 3 displays the stochastic CRB along with the DOA estimation RMSEs versus SNR for $N = 100$. Fig. 4 displays the similar values versus N for $\mathrm{SNR} = 0$ dB. The following methods are tested in these two figures:

- the conventional spectral MUSIC estimator;
- the estimator bank method that uses the multiple candidate estimators of (8) and selects the best single estimate yielding the largest stochastic likelihood function value (referred as "ML-EB" in Figs. 3 and 4);
- the estimator bank method that uses the multiple candidate estimators of (8) and sorts out the outlying estimates using the proposed quality assessment method

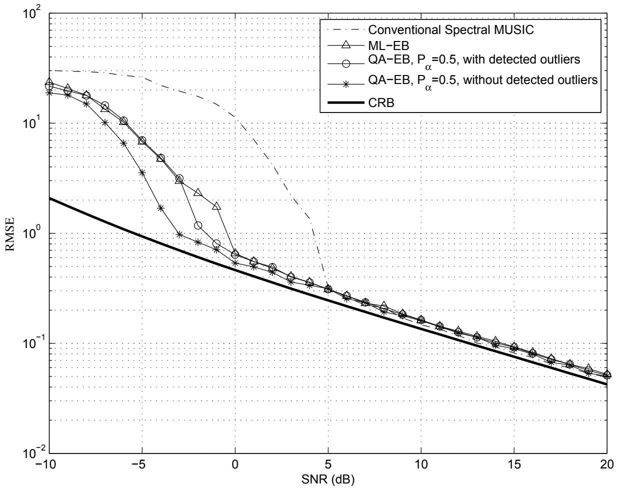

Fig. 3. DOA estimation RMSE's versus SNR.

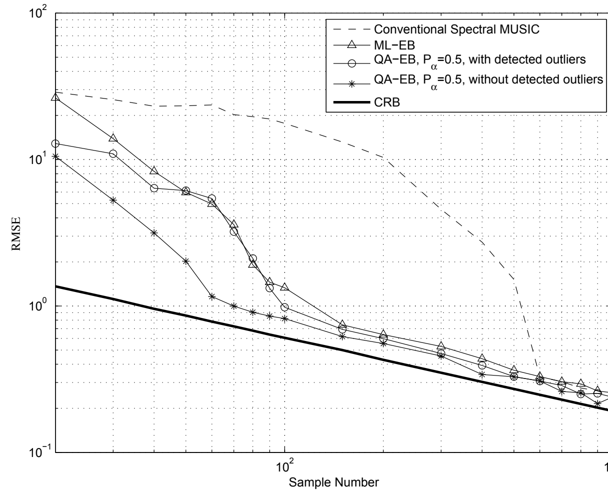

Fig. 4. DOA estimation RMSE's versus N.

with $P_\alpha = 0.5$ (referred as "QA-EB, $P_\alpha = 0.5$" in Figs. 3 and 4). When computing the experimental RMSEs for this method, two cases have been considered. In the first case, all simulation runs have been used, whereas in the second case, only those runs have been taken into account that correspond to the final DOA estimates classified as non-outlying estimates. The corresponding curves are labelled as "with detected outliers" and "without detected outliers", respectively.

All the curves in Figs. 3 and 4 are averaged over 300 simulation runs.

It can be observed from these figures that the proposed modification of the estimator bank approach provides substantially lower SNR and number of snapshot thresholds as compared to the MUSIC estimator and the earlier estimator bank approach (ML-EB). Another advantage of the proposed approach is that, according to the figures, it also predicts well how reliable is the final estimate.

6. CONCLUSIONS

A likelihood ratio-based quality assessment method has been introduced to detect the outlying estimates in estimator bank-based direction finding methods. Our simulations have demonstrated significant threshold performance improvements of the proposed approach as compared to the earlier estimator bank techniques.

7. REFERENCES

[1] Y. I. Abramovich *et al.*, "Positive-definite Toeplitz completion in DOA estimation for nonuniform linear antenna arrays — Part I: Fully augmentable arrays," *IEEE Trans. Signal Processing*, vol. 46, pp. 2458-2471, Sept. 1998.

[2] J. K. Thomas, L. L. Scharf, and D. E. Tufts, "The probability of a subspace swap in the SVD," *IEEE Trans. Signal Processing*, vol. 43, pp. 730-734, March 1995.

[3] A. B. Gershman, "Pseudo-randomly generated estimator banks: A new tool for improving the threshold performance of direction finding," *IEEE Trans. Signal Processing*, vol. 46, pp. 1351-1364, May 1998.

[4] P. Stoica and A. B. Gershman, "Maximum likelihood DOA estimation by data-supported grid search," *IEEE Signal Processing Letters*, vol. 6, pp. 273-275, Oct. 1999.

[5] A. B. Gershman and P. Stoica, "New MODE-based techniques for direction finding with an improved threshold performance," *Signal Processing*, vol. 76, 221-235, no. 3, Aug. 1999.

[6] M. Hawkes, A. Nehorai, and P. Stoica, "Performance breakdown of subspace-based methods: Prediction and cure," *ICASSP'01*, May 2001, Salt Lake City, UT, USA, vol. 6, pp. 4005-4008.

[7] Y. I. Abramovich and N. K. Spencer, "Performance breakdown of subspaced-based methods in arbitrary antenna arrays: GLRT-based prediction and cure," *ICASSP'04*, May 2004, Montreal, Canada, vol. 2, pp. 117-120.

[8] Y. I. Abramovich, N. K. Spencer, and A. Y. Gorokhov, "Bounds on maximum likelihood ratio — Part I: Application to antenna array detection-estimation with perfect wavefront coherence," *IEEE Trans. Signal Processing,* pp. 1524-1536, Dec. 2004.