COMPARATIVE THRESHOLD PERFORMANCE STUDY FOR CONDITIONAL AND UNCONDITIONAL DIRECTION-OF-ARRIVAL ESTIMATION

Yuri I. Abramovich† and Ben A. Johnson*

†DSTO, ISR Division, Edinburgh, Australia, e-mail: yuri.abramovich@dsto.defence.gov.au
* Lockheed Martin Australia and ITR, UniSA, Mawson Lakes, SA, Australia, e-mail: ben.a.johnson@ieee.org

ABSTRACT

Comparative analysis of the threshold SNR and/or sample support values where genuine maximum likelihood DOA estimation starts to produce “outliers” is conducted for unconditional (stochastic) and conditional (deterministic) problem formulations. Theoretical predictions based on recent results from Random Matrix Theory (RMT) are provided and supported by simulation results.

Index Terms— Direction of arrival estimation, adaptive arrays, error analysis, maximum likelihood estimation.

1. INTRODUCTION

Currently, two major statistical models are applied to describe the problem of direction-of-arrival (DOA) estimation on an M-element array.

1. Unconditional (Stochastic) Model: This model is applied when an exhaustive (multi-variate) statistical description of a set of \( x(t) \in \mathbb{C}^{M \times 1}, t = 1, \ldots, N \) antenna training data is provided, with some unknown (estimated) parameters whose number does not depend on the amount of training data \( N \). Typically, a multi-variate Gaussian model is applied, accompanied by additional assumptions on the statistical independence of the training vectors \( x(t) \).

2. Conditional (Deterministic) Model: This model is applied when the emitted source waveforms, sampled at an antenna receiver sample rate, can not be described as i.i.d. (complex) Gaussian vectors. In this case, the sampled sequences are treated as a priori unknown deterministic parameters. Obviously, the number of these “parameters” grows with the number of observations \( N \), and therefore the consistent estimation of all parameters is impossible here.

Analysis conducted in [1] demonstrated that in the conditional case, DOA estimation accuracy improves as \( N \to \infty \). Since a particular realization of i.i.d. Gaussian signals can be also treated as a set of unknown deterministic parameters, the conditional maximum likelihood (CML) technique can also be applied to this unconditional (i.i.d.) scenario, and will in general provide a different set of DOA estimates than unconditional maximum likelihood (UML).

The detailed comparison of asymptotic \((N \to \infty, M = \text{constant})\) properties of UML versus CML DOA estimation for such a scenario has been conducted by P. Stoica and A. Nehorai in [1]. There, it was analytically predicted that for i.i.d. Gaussian training data, the CML DOA estimate should always be inferior to the UML estimate, although a noticeable difference in those predictions was only demonstrated for highly spatially correlated sources (\( \rho = 0.995 \)), relatively small antenna array dimensions \((M = 5)\), small SNR values, and closely spaced multiple sources. While the analysis described the important region with estimation errors which approach the Cramèr-Rao bound (CRB), it unfortunately provides no insight into the so-called “threshold” behavior of these DOA estimation techniques, where the DOA estimates depart rapidly from the CRB due to the onset of severely erroneous ‘outliers’ as SNR and/or sample support decreases. In this paper, we demonstrate both theoretically and empirically that this conclusion (that UML equals or outperforms CML estimation) remains true for the threshold region as well.

2. G-ASYMPTOTIC UML AND CML DOA ESTIMATION THRESHOLD CONDITIONS

Let us consider a set of i.i.d. complex Gaussian training samples with the true (actual) covariance matrix \( R_0 \)

\[
R_0 = \sigma_n^2 I_M + A(\Theta_0^0) P_m A^H(\Theta_0^0)
\]

(1)

where \( \sigma_n^2 \) is the white noise power,

\[
A(\Theta_m) = [a(\theta_1), \ldots, a(\theta_m)] \in \mathbb{C}^{M \times m}, m < M
\]

(2)

is the \((M \times m)\)-variate “tall” matrix of antenna manifold (steering) vectors in the \( m \) source directions \( \Theta_m = (\theta_1, \ldots, \theta_m) \), and \( P_m \) is the inter-source (spatial) correlation matrix.

This work was partially conducted under DSTO/LMA Collaboration Agreement 290905.
As a result of UML DOA estimation, we reconstruct the estimated covariance matrix

\[ R(\hat{\Omega}_m^{ML}) = (\hat{\sigma}_m^{ML})^2 I_M + A(\hat{\Theta}_m^{ML}) \hat{P}_m^{ML} A^\dagger(\hat{\Theta}_m^{ML}) \]  

(3)

where \( \Omega_m^{ML} \) is the set of all a priori unknown parameter UML estimates; in the extreme case with no a priori knowledge on noise power and interspace covariance matrix, we have \( \Omega_m = \{ \sigma_m^2, P_m, \theta_1, \ldots, \theta_n \} \).

According to the maximum likelihood (ML) principle, we need to seek a likelihood function (or normalized likelihood ratio) which is maximal over the unknown parameter set. \( LR(\hat{\Omega}_m) \) is the normalized likelihood function (likelihood ratio) for multivariate Gaussian (zero mean, covariance \( R_0 \)) i.i.d. data \( X_N \) (i.e. \( x(t) \in CN(0, R_0), t = 1, \ldots, N \)) and its ML solution can be presented as:

\[
LR(\hat{\Omega}_m^{ML}) = \frac{\det R^{-1}(\hat{\Omega}_m^{ML}) R \exp(M)}{\exp \left\{ \text{Tr} R^{-1}(\hat{\Omega}_m^{ML}) R \right\}} \geq \frac{\det R_0^{-1} \hat{R} \exp(M)}{\exp \left\{ \text{Tr} R_0^{-1} \hat{R} \right\}}
\]  

(4)

where \( \hat{R} \) is given by \( \frac{1}{N} X_N X_N^\dagger \). Under the threshold condition, the likelihood ratio associated with a parameter set containing a highly erroneous DOA estimate (denoted \( \hat{\Omega}_m^{ex} \)) can be maximal across all possible parameter sets \( \Omega_m \) and will by definition equal or exceed the likelihood ratio of the unknown true covariance matrix. Therefore, under the threshold condition, we colloquially consider the likelihood principle itself to be “broke”, not because eqn. (4) fails to hold, but because the likelihood ratio no longer always discriminates between accurate and erroneous sets of estimates.

Since the whitened sample covariance matrix is complex Wishart distributed

\[ NR_0^{-\frac{1}{2}} \hat{R} R_0^{-\frac{1}{2}} = \hat{C}_M \sim CW(N, M, I_M)(N \geq M), \]

(5)

the UML threshold condition may be alternatively formulated as a condition where, despite a severely erroneous DOA estimate within the set \( \hat{\Omega}_m^{ex} \), the whitened sample matrix \( \hat{W}_M(\hat{\Omega}_m^{ex}) \)

\[ \hat{W}_M(\hat{\Omega}_m^{ex}) = R^{-\frac{1}{2}}(\hat{\Omega}_m^{ex}) \hat{R} R^{-\frac{1}{2}}(\hat{\Omega}_m^{ex}) \]

(6)

is “statistically indistinguishable” from the white noise-only complex Wishart sample matrix \( \frac{1}{M} C_M^N \). Here, by “statistical similarity”, we mean that the p.d.f. for the LR values associated with the two random matrices \( \hat{W}_M \) and \( \frac{1}{M} C_M^N \) should significantly overlap. Alternatively, it means that the deterministic matrix \( W_M(\hat{\Omega}_m^{ex}) \)

\[ W_M(\hat{\Omega}_m^{ex}) = R^{-\frac{1}{2}}(\hat{\Omega}_m^{ex}) R_0 R^{-\frac{1}{2}}(\hat{\Omega}_m^{ex}) \]

(7)

which is the transform between the two random matrices should be “sufficiently close” to the identity matrix to keep the distribution of (random) eigenvalues in \( \hat{W}_M(\hat{\Omega}_m^{ex}) \) close to the similar distribution of eigenvalues in \( \frac{1}{M} C_M^N \). To more carefully define the similarity of these eigenvalue distributions, we turn to tools derived from General Statistical Analysis (GSA) or Random Matrix Theory (RMT) [2]. In RMT, we consider statistical tests not under standard asymptotic conditions \( (N \to \infty, M = \text{constant}) \), but under Kolmogorov “G-asymptotic” conditions

\[ M, N \to \infty \text{ and } \frac{M}{N} \to c > 0. \]

(8)

In [3], we suggested the G-asymptotic criterion for similarity between two eigenvalue distributions introduced as a “single cluster” criterion. The main motivation for this criterion is that for the white noise-only sample matrix \( \frac{1}{N} C_M^N \), the empirical (sample) distribution of its non-ordered eigenvalues \( \hat{\gamma}_j \) converge almost surely to the non-random limiting distribution known as the Marchenko-Pastur law [4]. Because convergence under the Kolmogorov conditions (8) allows the antenna dimension to grow as the sample support grows, it is considered more appropriate for analysis of “threshold” performance than the conventional asymptotic assumption. While for any finite \( N \) and \( M \), G-asymptotic results may still not be sufficiently representative, a number of applications of RMT methodology demonstrate surprising fast convergence to the asymptotic results.

Not surprisingly, for a random matrix with the mean equal to the identity matrix, this limiting (under (8)) distribution is unimodal, with all eigenvalues \( \hat{\gamma}_j \) confined to the interval (contracting with \( c \to 0 \)):

\[
(1 - \sqrt{c})^2 < \hat{\gamma}_{\min} < \ldots < \hat{\gamma}_{\max} < (1 + \sqrt{c})^2
\]

(9)

When the deterministic matrix \( W_M(\hat{\Omega}_m^{ex}) \neq I_M \) in (7) and has \( M \) different eigenvalues with multiplicity \( K_j \) (the number of times each \( M \) true eigenvalue is repeated) for a sufficiently large training sample support (i.e. for \( c \ll 1 \)), the limiting eigenvalue distribution tends to concentrate around these \( M \) corresponding “clusters” [5,6]. Yet, for a comparatively small deviation of eigenvalues \( W_M(\hat{\Omega}_m^{ex}) \) from a constant number, there is a correspondingly high \( c = M/N(< 1) \) such that these individual clusters broaden, eventually coalescing into only one cluster, as per the white noise-only covariance matrix \( \frac{1}{N} C_M^N \), with its Marchenko-Pastur law. The latter means that eigenvalues in \( \hat{W}_M(\hat{\Omega}_m^{ex}) \) are not separable, exactly as the eigenvalues in \( \frac{1}{N} C_M^N \).

From the RMT perspective, under these conditions, scenario \( \hat{\Omega}_m^{ex} \) may be treated as statistically indistinguishable from the true (actual) scenario \( \Omega_0^{ex} \). For this paper’s model, with i.i.d. (Gaussian) sources, this “single cluster” criterion have been suggested in [3], based on the complementary “eigenvalue splitting” condition derived by X. Mestre in [5,6].

Specifically, for a sample matrix \( \hat{W}_M(\hat{\Omega}_m^{ex}) \), the limiting eigenvalue distribution is a “single cluster”, if and only if

\[
\frac{1}{c} \leq \min_{1 \leq j \leq M} \beta_M(j)
\]

(10)
where $\beta_M(j)$ are calculated as
\[
\beta_M(j) = \frac{1}{M} \sum_{j=1}^{M} K_j \left[ \frac{\gamma_j}{\gamma_j - \alpha_j} \right]^2; \quad \beta_M(0) = \beta_M(M) = 0
\] (11)
and $\alpha_j$ denotes the $(M-1)$ real-valued solutions to
\[
\frac{1}{M} \sum_{j=1}^{M} K_j \left[ \frac{\gamma_j}{\gamma_j - \alpha_j} \right] = 0
\] (12)
and $\gamma_j$ is the $j$-th eigenvalue of the deterministic matrix $W_M(\Omega_m^x)$ (7) with multiplicity $K_j$.

Let us now delineate similar “single cluster” conditions for the CML DOA estimation when applied to the same i.i.d. Gaussian data. Without a loss in generality, we may consider the case with an $a$ priori known white noise power $\sigma_n^2$. According to the ML principle, similarly to the UML case in (4), we have
\[
LR_{DET}(\hat{\Theta}_{m}^{ML}) = \frac{(M-m)\sigma_n^2}{\text{Tr} \left\{ P_\perp(\hat{\Theta}_{m}^{ML})R \right\}} \geq \frac{(M-m)\sigma_n^2}{\text{Tr} \left\{ P_\perp(\Theta_{m})R \right\}}
\] (13)
where
\[
P_\perp(\Theta_{m}) = I_M - A(\Theta_{m})[A^n(\Theta_{m})A(\Theta_{m})]^{-1}A^n(\Theta_{m})
\] (14)
Since the projector $P_\perp(\Theta_{m})$ may be presented as
\[
P_\perp(\Theta_{m}) = U_{M-m}(\Theta_{m})U^H_{M-m}(\Theta_{m})
\] (15)
\[
U^n(\Theta_{m})U(\Theta_{m}) = I_{M-m}
\] (16)
where $U_{M-m}$ are eigenvectors spanning the subspace complementary to the signal subspace $A^n(\Theta_{m})$, we get
\[
\text{Tr} \left\{ P_\perp(\Theta_{m})R \right\} = \text{Tr} \left\{ U^n_{M-m}(\Theta_{m})R U_{M-m}(\Theta_{m}) \right\}
\] (17)
and for $\Theta_{m} = \Theta_{m}^\circ$
\[
\text{Tr} \left\{ U^n(\Theta_{m})R U(\Theta_{m}) \right\} = \text{Tr} \left\{ \frac{1}{N} \hat{C}_{M-m} \right\}
\] (18)
where
\[
\hat{C}_{M-m} \sim \mathcal{CN}(N, M-m, I_{M-m})
\] (19)
While an accurate p.d.f. for $\text{Tr} \left\{ \hat{W}_{M-m}(\hat{\Theta}_{m}^x) \right\}$ where
\[
\hat{W}_{M-m}(\hat{\Theta}_{m}^x) = U^n_{M-m}(\Theta_{m})R U_{M-m}(\Theta_{m})
\] (20)
may be found in some cases, let us compare the “single cluster” conditions for UML matrix $\hat{W}_M(\hat{\Omega}_m^x)$ (6) and CML matrix $\hat{W}_{M-m}(\hat{\Theta}_{m}^x)$. Suppose that for the UML DOA estimation, an alternative scenario $\Omega_m^x$ that turns (10) into equality is found. The covariance for this scenario $R(\Omega_m^x)$ may be presented as
\[
R(\Omega_m^x) = \sigma_n^2[I_M + A(\Theta_m^x)\Gamma_m^s A^n(\Theta_m^x)]; \quad \Gamma_m^s = \sigma_n^{-2}P_m^x
\] (21)
where $P_m^x$ is the inter-source (spatial) covariance matrix as per (1). Since $\Gamma_m^x > 0$, from
\[
\|\Gamma_m^x\| > \|A(\Theta_m^x)A^n(\Theta_m^x)\|
\] (22)
we get the well-known first-order expansion for $R^{-1}(\Omega_m^x)$:
\[
R^{-1}(\Omega_m^x) \approx \sigma_n^2[P_\perp(\Theta_m^x) + A(\Theta_m^x)\times\times[A^n(\Theta_m^x)A(\Theta_m^x)]^{-1}[A^n(\Theta_m^x)A(\Theta_m^x)]^{-1}A^n(\Theta_m^x)]
\] (23)
It is important to note that the projector $P_\perp(\Theta_m^x)$ and the second matrix in (23) are spanned by orthogonal subspaces. Due to this orthogonality, we get the following inequalities
\[
\min_{1<j\leq M} [\text{eig}_j W_m(\Theta_m^x)] \leq \min_{1<j\leq M} [\text{eig}_j W_{M-m}(\Theta_m^x)]
\] (24)
\[
\max_{1<j\leq M} [\text{eig}_j W_m(\Theta_m^x)] \geq \max_{1<j\leq M} [\text{eig}_j W_{M-m}(\Theta_m^x)]
\] (25)
It follows from (23) and (24)-(25) that for any scenario $R(\Omega_m^x) \neq R(\Omega_m^0)$, the eigenspectrum of the “whitened” matrix $W_m(\Omega_m^x)$ should always have a larger span than the eigenspectrum of the $(M - m)$-variate “projected” CML matrix $W_{M-m}(\Theta_m^x)$. This means that under G-asymptotic condition (8)), the UML threshold conditions should always be more favorable (i.e. smaller SNR and/or sample support $N$) than for CML DOA estimation. The following section confirms this conclusion with simulation results.

3. SIMULATION RESULTS

Validity of the theoretical prediction in the last section can be examined using Monte-Carlo simulations. Since, in the analysis in [1] for the UML-CML difference outside the threshold region, differences were most noticeable for small arrays and closely spaced sources, we concentrate on the following scenario. An $M = 8$-element array with $d/\lambda = 0.5$ and $N = 16$ i.i.d. training samples is simulated. Two equal power sources at $\theta_1 = 18^\circ$ and $\theta_2 = 20^\circ$ embedded in additive i.i.d. white noise are varied over the post-beamforming SNR range of 0 to 35dB (i.e. -9 to 26dB per-element SNR), and simulated with no spatial correlation or with a spatial correlation coefficient of either 0.7 and 0.995. Within this SNR range, we observe transition from CRB “compliant” behavior to complete DOA estimation “breakdown” for each spatial correlation case, albeit at different threshold points for different spatial correlations, as expected. In order to assess the genuine ML threshold performance, we conducted exhaustive search for the global LR maximum over all possible angular locations of both sources, both for UML and CML. The demanding computational cost of exhaustive MLE (necessary in the threshold region where more computationally efficient MLE-proxies may not provide accurate representation of MLE performance) limited the simulations to 500 Monte-Carlo trials/SNR. The mean square errors for the two MLE algorithms
are presented in Fig. 1 to Fig. 3, along with the Cramér-Rao bounds. For the final correlation case \( r = 0.995 \), we also present the probability of an outlier occurring as a function of SNR and algorithm type in Fig. 4.

One can see that for low spatial correlation \( (r = 0, r = 0.7) \), no difference between UML and DML threshold behavior is observed. The predicted superiority of UML estimation threshold behavior over DML is only visible for extremely high inter-source correlations \( (r = 0.995) \), again in a similar result to that seen in [1] for non-threshold conditions.

4. SUMMARY AND CONCLUSION

Based on Random Matrix Theory methodologies, we have demonstrated that under the Kolmogorov asymptotic condition, the “single cluster” threshold criterion is more favorable for UML than CML DOA estimation, when applied to i.i.d. (Gaussian) training data. Yet, similarly to conventional asymptotic comparative analysis conducted in [1], the advantage of UML DOA estimation threshold performance is noticeable only for severe spatial correlation conditions.

5. REFERENCES