STATISTICAL NONIDENTIFIABILITY OF CLOSE EMITTERS: MAXIMUM-LIKELIHOOD ESTIMATION BREAKDOWN AND ITS GSA ANALYSIS

Yuri Abramovich, Ben Johnson, Nick Spencer

Intelligence, Surveillance and Reconnaissance Division, Defence Science & Technology Organisation (DSTO), PO Box 1500, Edinburgh SA 5111, Australia
Yuri.Abramovich@dsto.defence.gov.au

Lockheed Martin Pty Ltd, PO Box 22, Edinburgh SA 5111, Australia
Ben.A.Johnson@ieee.org

Adelaide Research & Innovation Pty Ltd (ARI),
200 Labs, ISRD, DSTO, PO Box 1500, Edinburgh SA 5111, Australia
Nick.Spencer@adelaide.edu.au

ABSTRACT

We investigate the "ambiguity region" associated with erroneous maximum-likelihood (ML) direction-of-arrival (DOA) estimates of closely spaced signals, near and below the "resolution limit". We demonstrate that the general statistical analysis (GSA) technique can accurately predict the ambiguity region for a given scenario. We consider that this prediction may be used together with the Barankin bound (BB) to form a more comprehensive description of maximum-likelihood estimation (MLE) performance in the problematic "threshold region" where ML techniques suffer from a dramatic failure rate ("performance breakdown").

Index Terms— Parameter estimation, signal detection, maximum likelihood estimation, array signal processing.

1. INTRODUCTION AND BACKGROUND

In [1], Cramér–Rao bound (CRB) analysis on two independent Gaussian sources and an M-sensor antenna array was used to predict that there is always a small enough separation such that information-theoretic criteria (ITC) reliably detect the correct number of sources ($\hat{m} = m$) down to the SNR "detection threshold" ε_{ITC}. At the same time, reliable ML DOA estimation with accuracy better than the intersource separation is possible only for much higher SNRs, above some ε_{ML} threshold. By a direct exhaustive two-dimensional ML search for the two-source scenario of [1], we found that within the SNR range

$$\varepsilon_{ITC} \lesssim \text{SNR} \lesssim \varepsilon_{ML} \tag{1}$$

the ML DOA estimate, and in fact, detection-estimation based on the generalised likelihood-ratio test (GLRT), suffer from *statistical nonidentifiability*. This means that, within this range of SNRs, both ITC- and GLRT-based detection-estimation always give an ML model with the correct number of sources ($\hat{m} = m = 2$), but one of the MLE DOA estimates is severely erroneous (is an "outlier"). We showed that the typical MLE result during performance breakdown is one DOA estimate close to the midpoint of the true DOAs ($\hat{\theta}_1 \simeq \bar{\theta} \equiv (\theta_1 + \theta_2)/2$), with the other DOA estimate (outlier) chosen by the ML search from the "ambiguity region", which is an extensive area of DOA values that depends on SNR and training data sample number T.

We have chosen to experiment with one of the scenarios studied by Lee and Li [1], namely

$$M = 3, \quad T = 100, \quad m = 2,$$
$$p_1 = p_2, \quad p_0 = 1, \quad \{\theta_1, \theta_2\} = \{0^\circ, 1.08^\circ\} \tag{2}$$

where the three-sensor antenna array is uniformly spaced at half-wavelength units, p_1 and p_2 are the powers of the sources, and p_0 is the white-noise power.

Fig. 1(a) shows the histogram over 1000 Monte-Carlo simulations at 25 dB SNR of the quantity

$$\min\left[\left|\hat{\theta}_1 - \theta_1\right|, \left|\hat{\theta}_2 - \theta_2\right|\right] \tag{3}$$

which is the error associated with the midpoint DOA estimate. The CRB for this scenario (for arbitrary source powers and known white-noise power) is 2.9° for each DOA. Despite this, the midpoint DOA error is seen to be much smaller, and is generally less than half of the intersource separation $(\theta_2 - \theta_1)/2 = 0.54^\circ$.

On the contrary, the sample histogram presented at Fig. 1(b) for the "outlier DOA estimate error"

$$\max\left[\left|\hat{\theta}_1 - \theta_1\right|, \left|\hat{\theta}_2 - \theta_2\right|\right] \tag{4}$$

shows a long tail that extends beyond 20°; this represents the "ambiguity region".

Analysing each DOA estimate separately, we find that in about half of the trials both estimates are confined to a very small range of θ around the midpoint DOA $\bar{\theta}$, while in the other half of trials they are distributed over some ambiguity region

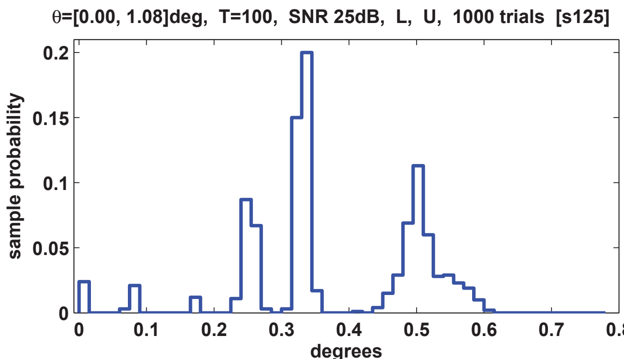
$$\theta_{\min} < \hat{\theta}_1 < \bar{\theta}, \qquad \bar{\theta} < \hat{\theta}_2 < \theta_{\max} \tag{5}$$

This statistical nonidentifiability means that for a given scenario $\{\theta_1, \theta_2, p_1, p_2\}$ there exists a continuum of alternate models, that includes the extreme ones at

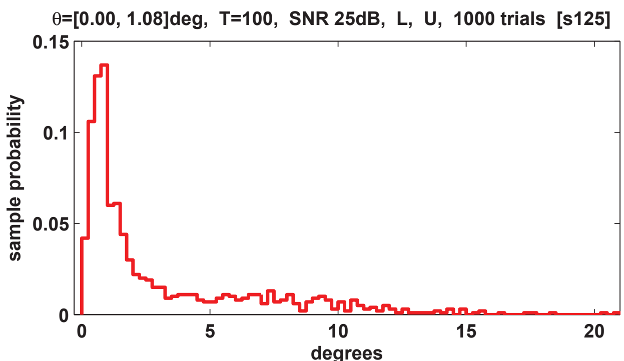
$$\{\theta_{\min}, \bar{\theta}, p_1 \ll p_2\} \quad \text{and} \quad \{\bar{\theta}, \theta_{\max}, p_1 \gg p_2\} \tag{6}$$

with an associated likelihood ratio (LR) that is statistically indistinguishable from the LR of the true (exact) covariance matrix R. The definition of these quantities is

$$R = SPS^H + p_0 I_M \tag{7}$$



(a)



(b)

Fig. 1. Sample distribution of the error associated with (a) the "midpoint" and (b) the "outlier" DOA estimate.

where

$$S = [\boldsymbol{s}(\theta_1), \ldots, \boldsymbol{s}(\theta_m)], \qquad P = \text{diag}\,[p_1, \ldots, p_m] \tag{8}$$

and $\boldsymbol{s}(\theta_j) \in \mathcal{C}^{M\times 1}$ is the antenna array manifold (steering) vector for the DOA (azimuthal angle) θ_j. For $T > M$, we use the LR as our (normalised) likelihood function [2], which is

$$LR(R_\mu) = \frac{\det(R_\mu^{-1}\hat{R})\exp M}{\exp\,\text{tr}\,(R_\mu^{-1}\hat{R})} \le 1 \tag{9}$$

where the estimated covariance matrix is $\hat{R} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{y}(t)\boldsymbol{y}^H(t)$, as usual, where $\boldsymbol{y}$ are the T independent identically distributed training data samples with complex Gaussian distribution, and the model covariance matrix is

$$R_\mu = S_\mu P_\mu S_\mu^H + p_0 I_M. \tag{10}$$

Recall that $LR(R_\mu)$ tests the hypothesis $H_0: \ \mathcal{E}\{\hat{R}\} = R_\mu$ versus the alternative hypothesis $\mathcal{E}\{\hat{R}\} \neq R_\mu$ [2].

Clearly, statistical nonidentifiability has a reciprocal quality: for any true scenario that belongs to the ambiguity region, the scenario with close DOAs and equal powers is ambiguous, as well as any other scenario from the ambiguity region. Therefore, for any given scenario with two closely separated sources and equal powers, and a given sample volume T, we have to specify the extent of the ambiguity region

$$\{\theta_{\min}, \bar{\theta}\} \quad \text{or} \quad \{\bar{\theta}, \theta_{\max}\} \tag{11}$$

for the MLE DOAs with arbitrary source powers. In fact, this definition of statistical nonidentifiability can be viewed as a version of an earlier likelihood-based interpretation of ambiguity function [3].

2. GSA PREDICTION OF AMBIGUITY REGION

For a given model covariance matrix $R_\mu \neq R$, we wish to directly find the (non-null) probability density function (pdf) of its LR (9). The standard approach is to find the j^{th} moment of $LR(R_\mu)$, and then to apply an inverse Mellin transform. Unfortunately, the analytic solution is in the form of a poorly converging series of Meijer G-functions [4] and so are inappropriate for a search of extreme scenarios with certain statistical properties of its LR.

In our case, the value $LR(R_\mu)$ (9) is just an (ML) metric that measures the "distance" between the random matrices $\hat{V} = R_\mu^{-\frac{1}{2}}\hat{R}R_\mu^{-\frac{1}{2}}$ and $\hat{C}_T = R^{-\frac{1}{2}}\hat{R}R^{-\frac{1}{2}}$, where $T\hat{R} \sim \mathcal{CW}(T \ge M, M, R)$ i.e. has a complex Wishart distribution. The properties of the matrix

$$T\hat{C}_T = \hat{C}, \qquad \hat{C} \sim \mathcal{CW}(T \ge M, M, I_M) \tag{12}$$

defines the pdf of $LR(R)$ (which is independent of R) [5], and so the properties of the matrix $\hat{V}$, where

$$T\hat{V} \sim \mathcal{CW}(T \ge M, M, R_\mu^{-\frac{1}{2}}RR_\mu^{-\frac{1}{2}}), \tag{13}$$

are "sufficiently close" to the properties of $T\hat{C}_T$ if the matrix

$$V = R_\mu^{-\frac{1}{2}}RR_\mu^{-\frac{1}{2}} \tag{14}$$

is "sufficiently close" to the identity matrix I_M, so that the random matrices $\hat{V}$ and $\hat{C}_T$ are statistically indistinguishable.

While the pdf's for $LR(R)$ and $LR(R_\mu)$ yield the full solution to this problem, we may consider tackling it via generalised statistical analysis (GSA), also known as G-analysis and random matrix theory (RMT). The GSA technique analyses the asymptotic behaviour of the eigen-decomposition of a random sample matrix $\hat{V}$ under the particular (Kolmogorov) assumption/condition

$$M, T \to \infty \quad \text{with} \quad \frac{M}{T} \to c \quad \text{for} \quad c \in (0, \infty). \tag{15}$$

GSA theory has established that the empirical distribution of the eigenvalues of $\hat{V}$ tend almost surely to a deterministic probability density G-asymptotically. It turns out that this asymptotic density of sample eigenvalues is organised into clusters located around the positions of the true eigenvalues. Most importantly, if M/T is too small then the asymptotic sample eigenvalue distribution is a single cluster, as per the matrix $\hat{C}_T$. Indeed, the empirical distribution for $\hat{C}_T$ $(T > M)$ converges almost surely to the non-random limiting distribution whose density is specified by the Marĉenko–Pastur law [6]:

$$f(x) = \frac{\sqrt{(x-a)(b-x)}}{2\pi c x} \tag{16}$$

where

$$a = (1-\sqrt{c})^2 \gtrsim \lambda_{\min}, \quad b = (1+\sqrt{c})^2 \lesssim \lambda_{\max} \tag{17}$$

Moreover, in [7, 8] it was demonstrated that the cluster for the k^{th} eigenvalue λ_k ($k = 1, \ldots, M'$, M' distinct true eigenvalues of V with multiplicities m_k) is separate from its neighbouring clusters $(\lambda_{k-1}, \lambda_{k+1})$ if the "eigenvalue splitting condition" is met:

$$\frac{T}{M} > \max_{j=k-1,k} \beta(j), \qquad j = 1, \ldots, M'-1 \tag{18}$$

where

$$\beta(j) = \frac{1}{M}\sum_{i=1}^{M'} m_i \left[\frac{\lambda_i}{\lambda_i - f_j}\right]^2 \tag{19}$$

$$\beta(0) = \beta(M') = 0 \tag{20}$$

and where f_j denotes the $(M'-1)$ real-valued solutions of the equation

$$\frac{1}{M}\sum_{i=1}^{M'} m_i \frac{\lambda_i^2}{[\lambda_i - f_j]^3} = 0. \tag{21}$$

In our case, statistical nonidentifiability means the contrary condition, since in order for $\hat{V}$ to have all sample eigenvalues in a single cluster, none of the eigenvalues of V can meet the splitting condition. The following theorem "inverts" the condition to suit our application.

Theorem
Let the distinct eigenvalues of the matrix V be $\lambda_1 < \lambda_2 < , \ldots, < \lambda_{M'}$ with multiplicity m_j, and consider

$$T\hat{V} = V^{\frac{1}{2}}\hat{C}V^{\frac{1}{2}}, \qquad \hat{C} \sim \mathcal{CW}(T \geq M, M, I_M) \tag{22}$$

$$\frac{M}{T} = c < 1 \tag{23}$$

then as $T, M \to \infty$ all sample eigenvalues $\hat{\lambda}_j$ of $\hat{V}$ belong to a single cluster if and only if

$$\frac{1}{c} < \min_{1 \leq j \leq M'-1} \beta_M(j) \tag{24}$$

$$\beta_M(j) = \frac{1}{M}\sum_{i=1}^{M'} m_i \left[\frac{\lambda_i}{\lambda_i - f_j}\right]^2 \tag{25}$$

with f_j calculated by (21).

For the given set $\{\theta_1, \theta_2, p_1 = p_2 = p\}$, let us consider the pair of DOAs $\{\theta_{\min}, \bar{\theta}\}$ or $\{\bar{\theta}, \theta_{\max}\}$ with the power $(2p - x)$ attributed to the source at $\bar{\theta}$ and the remaining power x to the second source. For the model

$$R_\mu(\bar{\theta}, \theta_{\max}) = (2p - x)\boldsymbol{s}(\bar{\theta})\boldsymbol{s}^H(\bar{\theta}) + x\boldsymbol{s}(\theta_{\max})\boldsymbol{s}^H(\theta_{\max}) + p_0 I_M \tag{26}$$

or the model

$$R_\mu(\theta_{\min}, \bar{\theta}) = x\boldsymbol{s}(\theta_{\min})\boldsymbol{s}^H(\theta_{\min}) + (2p - x)\boldsymbol{s}(\bar{\theta})\boldsymbol{s}^H(\bar{\theta}) + p_0 I_M \tag{27}$$

we can now calculate x_{opt} as

$$x_{opt} = \arg\max_x \min_j \beta_M(x, j) \tag{28}$$

with $\beta_M(x, j)$ as per (25). We see that the set of parameters $\{\bar{\theta}, \theta_{\max}, 2p - x_{opt}, x_{opt}\}$ is ambiguous if

$$\frac{T}{M} \leq \max_x \min_j \beta_M(x, j) \tag{29}$$

and by finding $\theta_{\max}$ or $\theta_{\min}$ such that

$$\frac{T}{M} = \max_\theta \max_x \min_j \beta_M(\theta, x, j) \tag{30}$$

we then have mapped the maximum extent of the ambiguity region.

As a special case, we may also find the SNR threshold below which the single-source model

$$R(\bar{\theta}) = (\lambda_{\max} - 1)\boldsymbol{s}(\bar{\theta})\boldsymbol{s}^H(\bar{\theta}) + p_0 I_M \tag{31}$$

is statistically indistinguishable from the true model R, whose eigenvalues are

$$1, \ldots, 1, \lambda_{M-1}, \lambda_M \equiv \lambda_{\max}. \tag{32}$$

In this case, the matrix V (14) has the single eigenvalue $\lambda_{M-1} \neq 1$, hence the single-cluster condition can be calculated analytically as

$$T \leq \frac{M-1}{\lambda_{M-1}^2}\left[1 + \sqrt[3]{\frac{(1+\lambda_{M-1})^2}{M-1}}\right]^3. \tag{33}$$

Finally, in the two-source scenario, where the eigenvalue $(1 + \lambda_{M-1})$ of the true covariance matrix R is the smallest signal subspace eigenvalue, we can apply the "non-splitting" condition to obtain the SNR threshold below which any ITC will fail, since this sample eigenvalue will belong to the "noise eigenvalues cluster". Putting

$$\lambda_{\max} \simeq 2p - x, \qquad \lambda_{M-1} \simeq x, \qquad \lambda_{\max} \gg \lambda_{M-1} \tag{34}$$

we find

$$T - 1 \lesssim \frac{M-2}{\lambda_{M-1}^2}\left[1 + \sqrt[3]{\frac{(1+\lambda_{M-1})^2}{M-2}}\right]^3. \tag{35}$$

For $M \gg 1$, the difference between the GLRT threshold (33) and the ITC threshold (35) becomes negligible, as would be expected. Yet, for the very small $M = 3$ studied by Lee and Li [1], there is a noticeable difference.

3. COMPARISON OF GSA AMBIGUITY REGION PREDICTIONS WITH MONTE-CARLO RESULTS

For the scenario (2), we conducted 1000 Monte-Carlo trials each for the SNR values 37 dB, 30 dB, 25 dB and 22 dB, calculating the global ML estimate of the DOAs and powers. The following table compares the results of the ambiguity region predictions of $\theta_{\min}$, $\theta_{\max}$ and x_{opt} with the Monte-Carlo results. Note that for these GSA predictions we have

$$\left|\theta_{\min} - \theta_1\right| = \left|\theta_{\max} - \theta_2\right|. \tag{36}$$

	GSA prediction			extreme Monte-Carlo results			
SNR	$\theta_{\min}$	$\theta_{\max}$	x_{opt} (dB)	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{p}_1$	$\hat{p}_2$
37 dB	$-3.5°$	$4.6°$	22.7	$-3.2°$	$0.5°$	23.3	40.0
30 dB	$-10.4°$	$11.5°$	7.4	$-9.4°$	$0.6°$	8.0	33.2
25 dB	$-22.8°$	$23.8°$	-2.7	$0.51°$	$21.3°$	28.2	-2.4
22 dB	$-38.2°$	$39.3°$	-7.2	$0.51°$	$38.3°$	24.9	-7.9

Finally, Fig. 2 plots $\min M\beta_M(j)$ as a function of SNR for (1) the extreme "orthogonal" outlier $\theta_{\max} = \theta_\perp$ with

$$\boldsymbol{s}^H(\theta_\perp)\boldsymbol{s}(\bar{\theta}) = 0, \qquad \theta_\perp = 42.55°, \quad \bar{\theta} = 0.54°; \tag{37}$$

(2) the single-source model with $R(\bar{\theta})$, i.e. the GLRT SNR threshold; and (3) the second eigenvalue in R, i.e. the ITC SNR threshold.

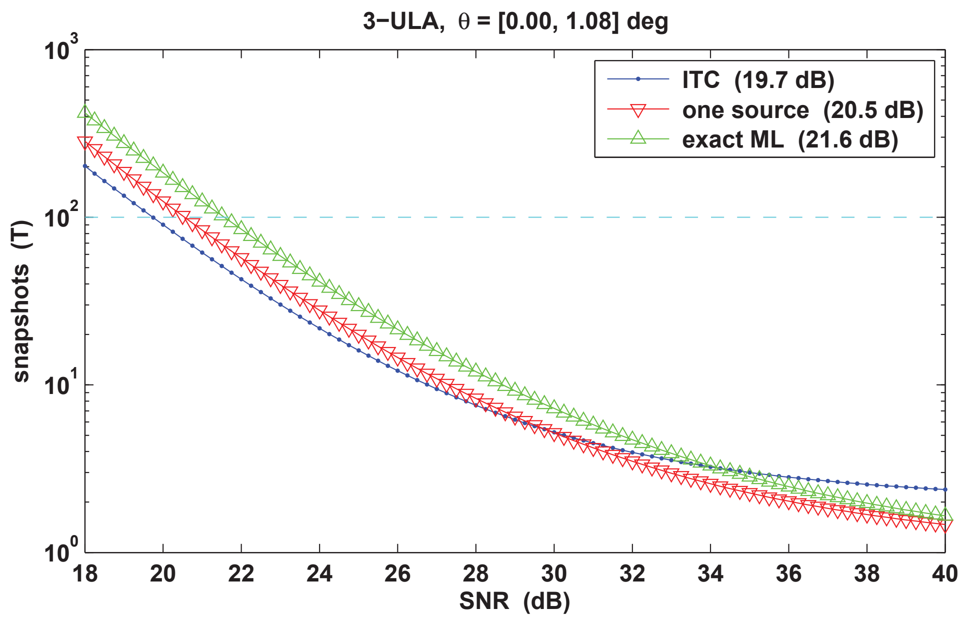


Fig. 2. GSA prediction of "single-cluster" nonidentifiability for the "orthogonal outlier" ("exact ML"), the single-source model, and ITC.

The figure highlights the intersection of these curves with our $T = 100$ snapshots, giving rise to the SNR thresholds annotated.

We see that the single-cluster nonidentifiability condition predicts the Monte-Carlo simulation results with remarkable accuracy. In fact, the extreme ML DOA estimation outliers are predicted down to probability 10^{-3} despite the considered scenario ($M = 3$, $T = 100$) being well below G-asymptotic requirements ($M, T \to \infty$, $M/T \to c$).

Similarly, G-asymptotic prediction of the ITC SNR threshold at 19.7 dB and the single-source threshold at 20.5 dB accurately coincide with the Monte-Carlo results. Indeed, MDL starts to breakdown at 22 dB, while both MDL and MAP fail completely at 19.7 dB, with AIC having a 50% failure rate.

4. SUMMARY AND CONCLUSIONS

We have demonstrated that ML DOA estimation of two closely spaced emitters within the SNR range

$$\varepsilon_{ITC} \lesssim \text{SNR} \lesssim \varepsilon_{ML} \tag{38}$$

suffers from "statistical nonidentifiability", which means that the global ML search gives two estimates: one is always located close to the midpoint DOA of the sources, and the other is an outlier, which is selected from the large "ambiguity region" whose extent depends on SNR and sample volume T. We have "inverted" the GSA eigenvalue-splitting condition, derived in [7, 8], to give the"single-cluster" statistical nonidentifiability condition. We showed that even for the tiny antenna array dimension $M = 3$, this G-asymptotic nonidentifiability condition very accurately describes the bounds of the ambiguity region compared with what was observed by Monte-Carlo simulation results. A similar correspondence has been found between the SNR where ITC fails to detect the correct number of sources and its G-asymptotic prediction. We suggest the single-cluster criterion is used instead of (or in conjunction with) the much more complicated "large error" Barankin bound analysis [9] to gain a more comprehensive understanding of ML detection-estimation performance for scenarios with close sources.

5. REFERENCES

[1] H. Lee and F. Li, "Quantification of the difference between detection and resolution thresholds for multiple closely spaced emitters," *IEEE Trans. Sig. Proc.*, vol. 41, no. 6, pp. 2274–2277, Jun. 1993.

[2] R. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.

[3] M. Rendas and J. Moura, "Ambiguity in radar and sonar," *IEEE Trans. Sig. Proc.*, vol. 46, no. 2, pp. 294–305, Feb. 1998.

[4] C. Khatri and M. Srivastava, "On exact non-null distribution of likelihood-ratio criteria for sphericity test and equality of two covariance matrices," *Indian J. Statistics*, vol. A33, pp. 201–206, 1971.

[5] Y. Abramovich, N. Spencer, and A. Gorokhov, "Modified GLRT and AMF framework for adaptive detectors," *IEEE Trans. Aero. Elect. Sys.*, vol. 43, no. 3, pp. 1017–1051, Jul. 2007.

[6] V. Marĉenko and L. Pastur, "Distribution of the eigenvalues in certain sets of random matrices," *Mat. Sb.*, vol. 1, pp. 457–483, 1967, in Russian.

[7] V. Girko, *An introduction to statistical analysis of random arrays*. Utrecht, Netherlands: VSP, 1998.

[8] X. Mestre, "On the asymptotic behaviour of the sample estimates of eigenvalues and eigenvectors of covariance matrices," *IEEE Trans. Sig. Proc.*, vol. 56, no. 11, pp. 5353–5368, 2008.

[9] I. Reuven and H. Messer, "The use of the Barankin bound for determining the threshold snr in estimating the bearing of a source in the presence of another," in *Proc. ICASSP-95*, vol. 3, Detroit, 1995, pp. 1645–1648.