THE ROLE OF SUBSPACE SWAP IN MAXIMUM LIKELIHOOD ESTIMATION PERFORMANCE BREAKDOWN

*Ben A. Johnson** and *Yuri I. Abramovich†* and *Xavier Mestre‡*

* RLM, Pty. Ltd. Edinburgh, Australia and ITR/Univ. South Australia, e-mail: ben.a.johnson@ieee.org
†DSTO, ISR Division, Edinburgh, Australia, e-mail: yuri.abramovich@dsto.defence.gov.au
‡Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Barcelona, Spain, e-mail: xavier.mestre@cttc.cat

ABSTRACT

Maximum likelihood estimation techniques demonstrate "performance breakdown" at low signal-to-noise ratios where observed estimation errors rapidly depart from the Cramér-Rao bound below a threshold SNR. Rather than rely on the classic asymptotic analysis for prediction of that threshold, Random Matrix Theory (RMT) analysis is employed. Both analytic predictions and direct Monte-Carlo simulations demonstrate that the threshold value can be reliably predicted even for small sample support far removed from classic asymptotic assumptions.

Index Terms— Direction of arrival estimation, maximum likelihood estimation, nonuniformly spaced arrays.

1. INTRODUCTION

It has been known for a long time that under certain "threshold" conditions, MLE may experience "performance breakdown" and generate severely erroneous estimates ("outliers") not consistent with the CRB predictions (see for example [1], pp. 278-286). Historically, analytical studies of such MLE "breakdown" almost always rely on traditional asymptotic ($M =$ const, $T \to \infty$) perturbation analysis. See [2] for an extensive summary of relevant work on prediction of this threshold effect, particularly in single source scenarios.

Recent results in Random Matrix Theory (RMT) (also known as General Statistical Analysis or GSA) provide some powerful tools to predict eigenvalue/eigenvector behavior which do not rely on traditional asymptotic assumptions. The application of these results to MLE breakdown prediction in both single and multiple source scenarios is explored in this paper.

2. PERFORMANCE BREAKDOWN IN MULTIPLE SOURCE SCENARIO

MLE breakdown is observed under conditions when a set of DOA estimates that contains a severely erroneous estimate (an outlier) generates a likelihood function (LF) value that exceeds the local extremum in the vicinity of the true solution. For a solution that contains an outlier to be "more likely" than the actual covariance matrix, the training data should indeed generate a sample signal subspace with some of its elements better represented by the true noise subspace. Therefore, the subspace swap phenomenon is likely to be associated with MLE breakdown. In order to demonstrate this, let us analyze MLE and MUSIC performance in a multiple target scenario, employing the scenario used in [3, 4], with a $M = 20$-element uniform linear array (ULA), $T = 15$ training samples, array element spacing of $d/\lambda = 0.5$ and $m = 4$ independent equal power Gaussian sources (stochastic source model) located at azimuth angles

$$\theta_m = \{-20^o, -10^o, 35^o, 37^o\}, \quad (1)$$

immersed in white noise, with varying per-element source SNRs.

The covariance matrix R_M for this mixture is

$$R_M = \left[\sum_{j=1}^{m} \sigma_j^2 S(\theta_j) S^{\mathrm{H}}(\theta_j)\right] + \sigma_0^2 I_M \quad (2)$$

where the noise power is $\sigma_0^2 = 1$; source SNR is σ_j^2/σ_0^2, $S(\theta)$ is the DOA θ-dependent M-variate "steering" (antenna manifold) vector, and the number of sources ($m = 4$) known *a priori*.

Obtaining an accurate MLE is not as straightforward. In the Gaussian case, MLE is theoretically obtained by selecting the single largest maxima of the multivariate likelihood function (LF) [5]:

$$\mathcal{L}[R_M(\Theta)] = \left[\frac{\exp[-\mathrm{Tr}\, R_M^{-1}(\Theta)\hat{R}_M(\Theta)]}{\pi^M \det R_M(\Theta)}\right] \quad (3)$$

where Θ represents the parameters power (σ_m^2) and angle of arrival (θ_m) for the m sources. However, since the actual global extremum of the LF cannot be guaranteed in practice, MLE performance is assessed using an MLE-proxy algorithm [6]. The essence of this algorithm is to first find a *local extremum* $\hat{R}_L$ of the likelihood function $\mathcal{L}[R_M]$ in the vicinity of the actual parameters Θ for every Monte-Carlo trial. We then make an initial "seed" estimate of the actual parameters using MUSIC derived DOAs and power estimation such as in [7]. This set of DOA estimates $\hat{\theta}_m$ for a given trial is treated as representative of MLE performance if $\mathcal{L}[R_M(\hat{\Theta})] \geqslant \mathcal{L}[\hat{R}_L]$.

Fig. 2 shows the mean-square error (MSE), averaged over 300 trials, for DOA estimates of the two closely spaced sources (at 35^o and 37^o) as source SNR is varied from -15 to $+25$dB. The figure demonstrates the familiar "threshold effect" in MSE for the DOA estimation process, with the sudden degradation in DOA accuracy (due to outliers) as the SNR is decreased. The MLE breakdown is demonstrated with the MLE-proxy algorithm discussed above, using "seeding" solutions produced by MUSIC. Also shown is the stochastic Cramér-Rao bound (CRB) for the two sources at 35^o and 37^o (averaged together). One can observe the large gap in threshold SNRs for MUSIC and MLE for this closely spaced source scenario.

3. SUBSPACE SWAP PREDICTIONS BY GSA/RMT

Tufts, et. al. in [8] associated the threshold effect with the probability of *subspace swap* when some of the noise subspace eigenvectors

This work was partially conducted under DSTO/RLM Collaboration Agreement 290905

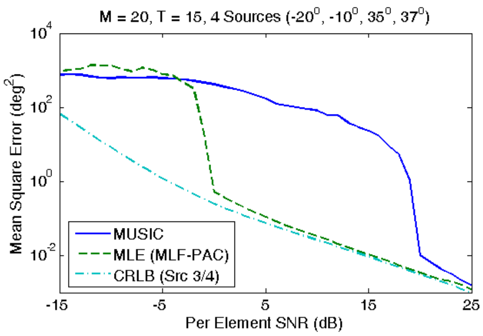


Fig. 1. Multiple-source estimation on a 20-element uniform linear array with $T = 15$ training samples for MUSIC and MLE.

of the sample covariance matrix better represent the signal subspace of the true covariance matrix R_0, than some sample signal subspace eigenvectors [9, 10]. Naturally, this probability depends on how well the signal and noise subspaces are separated in the true covariance matrix, indicated by how large the smallest signal subspace eigenvalue λ_m is compared to the white noise power σ_0^2 in R_0.

For a finite M/T ratio, the fact that the eigensubspaces of the sample covariance matrix $\hat{R}$ may not be the best option to estimate signal and noise subspaces has been recently explored by Mestre in [3]. Based on the General Statistical Analysis (GSA) methodology [11], Mestre derived a G-estimate of the signal subspace and therefore, a G-MUSIC pseudo-spectrum estimate. Importantly, the m-variate signal subspace estimate is found by a weighted sum of all M eigenvectors of the sample matrix $\hat{R}$, and "neighboring" eigenvectors in the sample signal and noise subspaces are weighted accordingly. The weighting coefficients are derived in [12] based on doubly asymptotic assumptions:

$$T, M \to \infty, \quad M/T = \gamma = \text{ const} \tag{4}$$

and allowed for significant improvement in threshold values, as demonstrated in [3].

Note that the GSA methodology provides an important insight into subspace swap mechanisms, at least for these doubly asymptotic (4) conditions. Indeed, Paul in [13], Theorem 4, states that for a real-valued covariance matrix R_0 of the form

$$R_0 = diag(\lambda_1, \lambda_2, \ldots, \lambda_m, 1, \ldots, 1); \tag{5}$$

where

$$\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_m > 1 \tag{6}$$

so that $\mathbf{e}_m(0\ \ldots\ 0\ 1\ 0\ \ldots\ 0)^T$ (with unity in the m-th position) is the m-th eigenvector of R_0, then

a) if $\lambda_m > 1 + \sqrt{\gamma}$ and of multiplicity one (as in (5)),

$$|\langle \mathbf{e}_m^{\mathrm{H}} \hat{\mathbf{e}}_m \rangle| \overset{a.s.}{\to} \sqrt{\left(1 - \frac{\gamma}{(\lambda_m - 1)^2}\right) \Big/ \left(1 + \frac{\gamma}{(\lambda_m - 1)}\right)} \tag{7}$$

b) if $\lambda_m \leqslant 1 + \sqrt{\gamma}$

$$|\langle \mathbf{e}_m^{\mathrm{H}} \hat{\mathbf{e}}_m \rangle| \overset{a.s.}{\to} 0 \text{ as } M, T \to \infty,\ \gamma = \text{ const}. \tag{8}$$

In fact, this theorem reinforces the observation made by Johnstone and Lu [14] who showed that when $M/T \to \gamma \in (0, \infty)$, the sample principle components (eigenvectors) are inconsistent estimates of the true eigensubspace. Furthermore, Paul points out [13] a "phase transition phenomenon", which is clearly analogous to the subspace swap phenomena [15]. A broader definition of the "signal processing" subspace swap phenomena implies

$$\mathbf{e}_4^{\mathrm{H}} \hat{E}_N \hat{E}_N^{\mathrm{H}} \mathbf{e}_4 > \mathbf{e}_4^{\mathrm{H}} \hat{E}_S \hat{E}_S^{\mathrm{H}} \mathbf{e}_4 \tag{9}$$

or equivalently

$$\mathbf{e}_4^{\mathrm{H}} \hat{E}_S \hat{E}_S^{\mathrm{H}} \mathbf{e}_4 < 0.5 \tag{10}$$

i.e. the last signal eigenvector is better represented by the noise subspace than the signal subspace. Therefore, the behavior of the projection $\mathbf{e}_j^{\mathrm{H}} \hat{E}_S \hat{E}_S^{\mathrm{H}} \mathbf{e}_j$ is of relevance to subspace swap (and therefore MLE breakdown) prediction. In order to predict these projection values rather than just the eigenvector correlation predicted by Paul in (7), we have derived in [**?**] (utilizing Theorem 2 of Mestre [16]) the following expression for this projection:

$$\mathbf{e}_j^{\mathrm{H}} \hat{E}_S \hat{E}_S^{\mathrm{H}} \mathbf{e}_j \xrightarrow{a.s.} \frac{\lambda_j}{\lambda_j - (1+\gamma)} - \frac{1}{\lambda_j - 1} = \frac{1 - \gamma \frac{1}{(\lambda_j - 1)^2}}{1 + \gamma \frac{1}{(\lambda_j - 1)}}. \tag{11}$$

For our specific scenario with the minimal signal subspace eigenvalue associated with $\mathbf{e}_4$:

$$\mathbf{e}_4^{\mathrm{H}} \hat{E}_S \hat{E}_S^{\mathrm{H}} \mathbf{e}_4 \xrightarrow{a.s.} \left(1 - \frac{\gamma}{(\lambda_4 - 1)^2}\right) \Big/ \left(1 + \frac{\gamma}{(\lambda_4 - 1)}\right). \tag{12}$$

One can see that we get the same asymptotic expression as in (7), but now for the projection onto the entire sample subspace. This means that when every eigenvalue has a distinct cluster (no "intra-subspace swap") then the power (7) of the true eigenvector $\mathbf{e}_4$ asymptotically resides in the fourth sample subspace eigenvector $\hat{\mathbf{e}}_4$, while the remaining power resides in the sample noise subspace. If instead only the smallest signal subspace eigenvalue is distinct from the noise subspace (no "inter-subspace swap"), then the same power (12) is distributed across multiple sample signal subspace eigenvectors.

4. SIMULATION RESULTS

To examine the relationship of MLE threshold effects to this subspace swap predictions, let us examine scenario (1) at three SNR values: 2dB with $\lambda_4 = 4.97$; 0dB with $\lambda_4 = 3.51$; and -4dB with $\lambda_4 = 2.00$.

For SNR = 2dB and $\lambda_4 = 4.97$, according to (12) for this SNR, we get:

$$\mathbf{e}_4^{\mathrm{H}} \hat{E}_S \hat{E}_S^{\mathrm{H}} \mathbf{e}_4 \xrightarrow{a.s.} 0.69. \tag{13}$$

Monte-Carlo simulations show a mean for $\mathbf{e}_4^{\mathrm{H}} \hat{E}_S \hat{E}_S^{\mathrm{H}} \mathbf{e}_4$ of 0.6815, agreeing well with the prediction and indicating that the 4th eigenvector projects more onto its proper signal subspace than the noise subspace.

For SNR = 0dB and $\lambda_4 = 3.51$, the projection of signal eigenvector onto the signal subspace is forecast via (12) as:

$$\mathbf{e}_4^{\mathrm{H}} \hat{E}_S \hat{E}_S^{\mathrm{H}} \mathbf{e}_4 \xrightarrow{a.s.} 0.51. \tag{14}$$

Monte-Carlo simulations show a mean for $\mathbf{e}_4^{\mathrm{H}} \hat{E}_S \hat{E}_S^{\mathrm{H}} \mathbf{e}_4$ of 0.5098. This indicates that the subspace swap condition given in (10) is essentially satisfied and subspace swap is statistically likely, consistent with the observation in Fig. 2 that MLE breakdown starts to occur at 0dB input SNR.

For SNR = -4dB, $\lambda_4 = 2.0$, and the condition for G-asymptotic convergence ($\lambda_4 > 1 + \sqrt{\gamma} = 2.15$) given in (8) is violated, so the projection should converge to zero asymptotically and (12) is no longer valid. Monte-Carlo simulations show a mean for $\mathbf{e}_4^{\mathrm{H}} \hat{E}_S \hat{E}_S^{\mathrm{H}} \mathbf{e}_4$ of 0.2502. Clearly the scenario is experiencing significant subspace swap, once again consistent with results given in Fig. 2.

Having established that MLE performance breakdown in the examined multiple source scenario (1) is indeed reliably associated with *subspace swap*, we next examine a single source scenario, such as studied in [17, 18], where we expect that subspace swap will indeed be the sole mechanism responsible for both MLE and MUSIC DOA estimation performance breakdown.

To this end, we introduce a second scenario with a single target, based, as in Athley [17, 18], on a sparse minimum redundancy array (MRA) [19], where the generation of outliers is more likely due to poor sidelobe performance. We use the following specific $M = 18$ configuration (d = [0, 2, 10, 22, 53, 56, 82, 83, 89, 98, 130, 148, 153, 167, 188, 192, 205, 216]) [20].

The threshold effect of MLE estimation in this scenario (provided by the Barlett spectrum or conventional beamforming (CBF)) can be observed in Fig. 2 to occur around -5dB for $T = 14$.

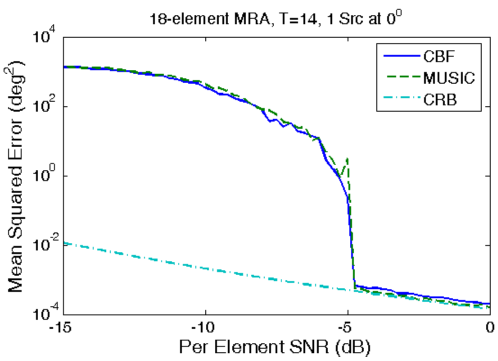


Fig. 2. MSE for MUSIC and MLE (CBF) DOA estimation on a 18-element minimum redundancy array with 1000 trials/SNR step. Note that MUSIC delivers essentially the same performance as MLE in this circumstance, as expected for single sources.

In (9) we defined subspace swap as occurring when the projection of last true eigenvector into the underlying sample noise subspace was higher than into the sample signal subspace. To examine whether this subspace swap is the sole mechanism for MLE breakdown, we can plot for each of 1000 Monte-Carlo trials and a training sample size of $T = 14$, the DOA error of a single source estimated with the MRA versus the correlation between the "maximal" sample and true eigenvector. These plots are shown in Fig. 3 for source SNRs ranging from very low values which result in complete MLE breakdown (input SNR of -18dB, as shown in Fig. 3(a)) to values where there is no MLE breakdown (input SNR of 0dB, as shown in Fig. 3(d)).

Fig. 3 clearly demonstrates that when the projection of the signal true eigenvector onto the sample signal subspace is high, there is no MLE breakdown (i.e. the upper right quadrant of Figs. 3(a) to (d) are all free of any Monte Carlo trials). Interestingly, however, the converse is not true. When the projection of the signal true eigenvector onto the sample signal subspace is low, a DOA outlier estimate may or may not be produced. Thus subspace swap is a necessary but not sufficient condition for MLE breakdown to occur.

Turning our attention back to the uniform line array scenario with closely spaced sources given in (1) with $M = 20$ and $T = 15$, we now conduct a similar examination, using DOA estimates provided by the MLE-proxy algorithm. Because there are multiple sources, but usually only one outlier, we plot the worst observed error versus the projection of the 4th true eigenvector onto the sample signal subspace in Fig. 4. The behavior is remarkably similar to the single source performance shown earlier, and therefore the observation that subspace swap is a necessary but not sufficient condition for MLE breakdown is strongly reinforced.

5. SUMMARY AND CONCLUSION

In this paper we investigated the well-known performance breakdown phenomenon in DOA estimation techniques, which manifests as a dramatic and rapid departure of estimation accuracy from the CRB due to the increasing probability of erroneous "outlier" estimates as the SNR or number of training samples is decreased below certain threshold values.

We analyzed this phenomenon for MLE in multiple and single Gaussian source scenarios with limited i.i.d. sample support. Rather than consider a traditional $T \to \infty$ asymptotic analysis, we specifically considered parameters far removed from the traditional asymptotic regime, focusing on under-sampled scenarios with the number of training samples T less than the antenna dimension M. To provide theoretical analysis of this small-sample regime, we employed the so-called General Statistical Analysis (GSA) methodology that considers the asymptotic regime

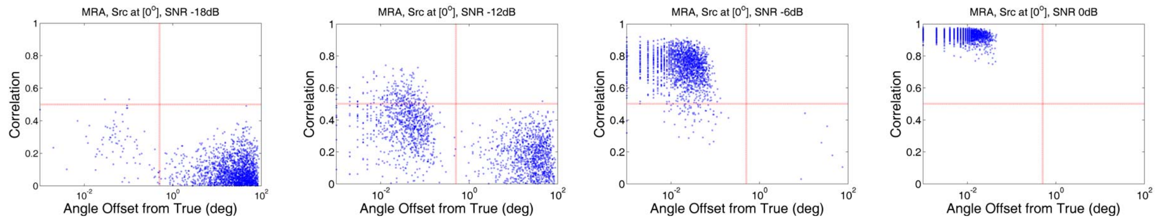
$$M, T \to \infty, \; M/T \to \text{ const.} \tag{15}$$

which differs significantly from the usual M = constant, $T \to \infty$ asymptotic assumptions. This analysis, supported by the results of direct Monte-Carlo simulations, lead to a number of important observations.

First of all, our analysis demonstrated that the GSA methodology very accurately predicts the subspace swap conditions, even for antenna dimensions and sample volume which are far from the G-asymptotic regime.

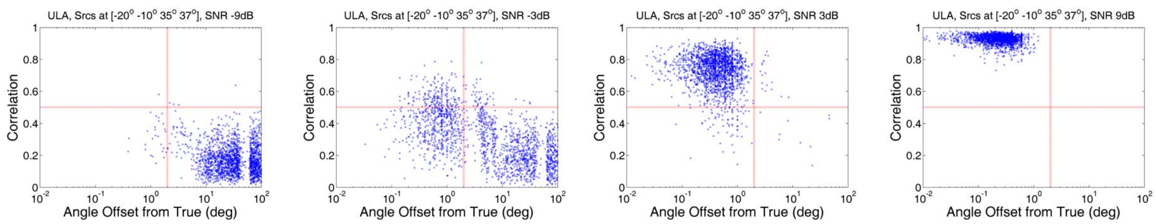
MLE performance breakdown takes place when a set of estimates that contain an outlier is "more likely" than the true parameters or even the local LF maximum in their vicinity. For this to happen, the input data should be insufficient, and therefore with no surprise we established that MLE breakdown is indeed reliably associated with the subspace swap phenomena, well predicted by the GSA methodology.

For scenarios where the MUSIC pseudo-spectrum does not differ significantly from the conventional Barlett spectrum (single or well-separated sources, very low SNR, $T \to \infty$), MUSIC and MLE techniques will demonstrate similar threshold performance, with full subspace swap becoming the common reason for breakdown in both techniques. For single source cases, the similar breakdown point for MLE and MUSIC is well correlated with GSA-derived eigenvalue splitting and subspace swap predictions. It was noted, however, that while high projection values of the minimal signal eigenvector onto the sample signal subspace precludes the formation of outliers leading to performance breakdown, low projection values (indicating in some cases almost complete subspace swap) did not always lead to performance breakdown. Therefore subspace swap is a necessary, but not sufficient, condition for DOA estimation breakdown, with other factors also influencing outlier production. Practically speaking, however, robust prediction by GSA of the sample volume and/or SNR which allows subspace swap to be avoided altogether is the proper strategy to retain high MLE performance.



(a) ∼100% MLE Breakdown (b) ∼50% MLE Breakdown (c) Rare MLE Breakdown (d) No MLE Breakdown

Fig. 3. Distribution of DOA estimation errors versus projection $\mathbf{e}_j^{\mathrm{H}}\hat{E}_S\hat{E}_S^{\mathrm{H}}\mathbf{e}_j$ for a single target scenario on 18-element minimum redundancy array with $T = 14$ training samples. As SNR is increased, the projection approaches unity, and the estimation accuracy improves. Note, however, that trials with low projection values still frequently have low estimation errors.



(a) ∼100% MLE Breakdown (b) ∼50% MLE Breakdown (c) Rare MLE Breakdown (d) No MLE Breakdown

Fig. 4. Distribution of DOA estimation errors versus projection $\mathbf{e}_j^{\mathrm{H}}\hat{E}_S\hat{E}_S^{\mathrm{H}}\mathbf{e}_j$ for worst error in multiple source scenario on 20-element uniform linear array with $T = 15$ training samples. As with the single source case, the projection of the true eigenvector onto the sample signal subspace is a necessary but not sufficient condition for breakdown to occur.

6. REFERENCES

[1] H. L. van Trees, *Detection, Estimation, and Modulation Theory: Part I*. New York: John Wiley and Sons, 1968.

[2] C. D. Richmond, "Mean-squared error and threshold SNR prediction of maximum-likelihood signal parameter estimation with estimated colored noise covariances," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2146–2164, May 2006.

[3] X. Mestre, "An improved weighted MUSIC algorithm for small sample size scenarios," in *Proc. 14th Workshop on Adaptive Sensor Array Processing*. Lexington, MA, USA: MIT-LL, 2006. [Online]. Available: www.ll.mit.edu/asap/asap_06/pdf/Papers/19_mestre_pa.pdf

[4] Y. I. Abramovich, B. A. Johnson, and X. Mestre, "Performance breakdown in MUSIC, G-MUSIC and maximum likelihood estimation," in *Proc. ICASSP*, vol. 2. Honolulu, HI, USA: IEEE, Apr 15–20 2007, pp. 1049–1052.

[5] T. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley and Sons, 1958.

[6] Y. I. Abramovich, N. K. Spencer, and A. Y. Gorokhov, "GLRT-based threshold detection-estimation preformance improvement and application to uniform circular antenna arrays," *IEEE Trans. Signal Processing*, vol. 55, no. 1, pp. 20–31, Jan 2007.

[7] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing," in *Radar Array Processing*. Berlin: Springer-Verlag, 1993, pp. 99–151.

[8] D. Tufts, A. Kot, and R. Vaccaro, "The threshold effect in signal processing algorithms which use an estimated subspace," in *SVD and Signal Processing II: Algorithms, Analysis and Applications*, R. Vaccaro, Ed. New York: Elsevier, 1991, pp. 301–320.

[9] M. Hawkes, A. Nehorai, and P. Stoica, "Performance breakdown of subspace-based methods: Prediction and cure," in *Proc. ICASSP*, vol. 6. Salt Lake, UT, USA: IEEE, 2001, pp. 4005–4008.

[10] J. Thomas, L. Scharf, and D. Tufts, "The probability of a subspace swap in the SVD," *IEEE Trans. Signal Processing*, vol. 43, no. 3, pp. 730–736, Mar 1995.

[11] V. L. Girko, "An introduction to general statistical analysis," *Theory Probab. Applic.*, vol. 32, no. 2, pp. 229–242, 1987.

[12] X. Mestre, "An improved subspace-based algorithm in the small sample size regime," in *Proc. ICASSP*. Toulouse, France: IEEE, 2006.

[13] D. Paul, "Asymptotics of the leading sample eigenvalues for a spiked covariance model," Stanford University, Tech. Rep., Dec 2004. [Online]. Available: http://anson.ucdavis.edu/∼debashis/techrep/eigenlimit.pdf

[14] I. M. Johnstone and A. Y. Lu, "Sparse principal components analysis," 2004. [Online]. Available: www-stat.stanford.edu/people/faculty/johnstone/WEBLIST/AsYetUnpub/sparse.pdf

[15] D. W. Tufts, A. C. Kot, and R. J. Vaccaro, "The threshold analysis of SVD-based algorithms," in *Proc. ICASSP*, New York, NY, USA, Apr 1988, pp. 2416–2419.

[16] X. Mestre, "Improved estimation of eigenvalues of covariance matrices and their associated subspaces using their sample estimates," *Submitted to IEEE Transactions on Information Theory*, 2006. [Online]. Available: http://www.cttc.cat/drafts/cttc-rc-2006-004.pdf

[17] F. Athley, "Performance analysis of DOA estimation in the threshold region," in *Proc. ICASSP*, vol. 3. Orlando, FL, USA: IEEE, 2002, pp. 3017–3020.

[18] ——, "Threshold region performance of maximum likelihood direction of arrival estimators," *IEEE Trans. Signal Processing*, vol. 53, no. 4, pp. 1359–1373, Apr 2005.

[19] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Trans. Antennas Propagat.*, vol. 16 (2), pp. 172–175, 1968.

[20] A. Dollas, W. Rankin, and D. McCracken, "An new algorithm for Golomb ruler derivation and proof of the 19 mark ruler," *IEEE Transactions on Information Theory*, no. 1, pp. 379–382, Jan 1998.