PERFORMANCE BREAKDOWN OF SUBSPACE-BASED METHODS: PREDICTION AND CURE

Malcolm Hawkes and Arye Nehorai[*]

EECS Department
University of Illinois
851 S. Morgan St, Chicago, IL 60607

Petre Stoica[†]

Systems and Control Department
Uppsala University
P.O. Box SE-751 03 Uppsala, Sweden

ABSTRACT

The performance breakdown of subspace-based parameter estimation methods can be naturally related to a switch of vectors between the estimated signal and noise subspaces (a "subspace swap"). In this paper we derive a lower bound for the probability of such an occurrence and use it to obtain a simple data-based indicator of whether or not the probability of a performance breakdown is significant. We also present a conceptually simple technique to determine from the data whether or not a subspace swap has actually occurred, and to extend the range of SNR values or data samples in which a given subspace method produces accurate estimates.

1. INTRODUCTION

All subspace-based estimation methods are known to suffer a rapid degradation in performance as either the signal-to noise ratio (SNR) or the number of available snapshots N drops below a certain value, called the (SNR or N) *threshold*, [1]–[3]. Such a dramatic drop in performance can only be explained by a discontinuity in the parameter estimates, corresponding to the method in question, as functions of the processed data. The sole apparent discontinuity, which is typical of all subspace-based methods, is induced by the splitting of the *measurement space* into a *signal subspace* and a *noise subspace*.

Consider a scenario in which we have N snapshots of an $m \times 1$ complex vector from the measurement model

$$\boldsymbol{y}(t) = A(\boldsymbol{\theta})\boldsymbol{s}(t) + \boldsymbol{e}(t), \tag{1.1}$$

where $\boldsymbol{s}(t)$ is a signal term, $\boldsymbol{e}(t)$ is noise, and $\boldsymbol{\theta}$ is an unknown parameter vector of interest. Suppose that the signal and noise terms are uncorrelated with one another and that the latter has covariance matrix $\sigma^2 I$. Let R and $\hat{R}$ be the theoretical and sample data covariance matrices, respectively. Denote the eigenelements of $\hat{R}$ by $\{\lambda_k, \boldsymbol{v}_k\}_{k=1}^m$, with the eigenvalues $\{\hat{\lambda}_k\}_{k=1}^m$ arranged in non-increasing order. Let $n < m$ be the number of signals, and define

$$\hat{S} = [\hat{\boldsymbol{v}}_1, \ldots, \hat{\boldsymbol{v}}_n], \tag{1.2}$$

$$\hat{G} = [\hat{\boldsymbol{v}}_{n+1}, \ldots, \hat{\boldsymbol{v}}_m]. \tag{1.3}$$

Similarly, let $\{\lambda_k, \boldsymbol{v}_k\}_{k=1}^m$, S and G be the corresponding quantities associated with R. The columns of S and G span the so-called *signal subspace* and *noise subspace*, respectively, while those of $\hat{S}$ and $\hat{G}$ span the corresponding *estimated subspaces*. Under the above conditions *the noise eigenvalues*, $\lambda_{n+1}, \ldots, \lambda_m$ are all equal to σ^2.

Let the symbol $\mathcal{M}$ denote a generic subspace-based parameter estimation method. Suppose the gap between the sets of signal and noise eigenvalues, i.e. between $\{\lambda_1, \ldots, \lambda_n\}$ and σ^2, is large compared with the sampling fluctuations in $\{\hat{\lambda}_k\}_{k=1}^m$. Then the assignment of the estimated eigenvectors into signal and noise subspaces should, with high probability, be done correctly. In that case $\hat{S}$ will be a good approximation to S, the error merely being due to small sampling fluctuation. Similarly, $\{\hat{\lambda}_k\}_{k=1}^m$ will be good estimates of $\{\lambda_k\}_{k=1}^m$.

As either the gap between $\{\lambda_k\}_{k=1}^n$ and σ^2 decreases (which happens when the scenario becomes more difficult, e.g. highly correlated signals that are not well separated), or the values of N and SNR decrease (leading to greater sampling variability in $\{\hat{\lambda}_k\}_{k=1}^m$), the estimates of $\lambda_{n+1}, \lambda_{n+2}$, etc. will with increasing probability be larger than the estimates of λ_n, λ_{n-1}, etc. Whenever this happens a *subspace swap* has occurred. More precisely, in such a case one or more pairs of the set $\{\hat{\lambda}_k, \hat{\boldsymbol{v}}_k\}_{k=1}^n$ actually estimate noise eigenelements instead of signal eigenelements. For the eigenvalues this is usually no big problem: in most cases of interest $\{\hat{\lambda}_k\}_{k=1}^n$ will still be in the vicinity of the signal eigenvalues $\{\lambda_k\}_{k=1}^n$ despite the incorrect association. For the eigenvectors, however, the wrong association leads to the use of one or more sample noise subspace vectors as estimates of *completely different* signal subspace vectors. When this happens, the drop in performance of $\mathcal{M}$ is generally significant.

In Section 2, we present a formula that can be used, along with the observed data, to make inferences about the probability of a subspace swap. We also discuss, in Section 3, how to modify the method $\mathcal{M}$ whenever the subspace swap probability is high and derive methods for determining, based on the observed data

[*] Supported by the Air Force Office of Scientific Research under Grants F49620-99-1-0067 and F49620-00-1-0083, and the Office of Naval Research under Grant N00014-98-1-0542.

[†] Supported by the Swedish Foundation for Strategic Research under a Senior Individual Grant

alone, whether or not a subspace swap has actually occurred. Numerical examples involving the standard ESPRIT [8] algorithm for direction-of-arrival (DOA) estimation are given in Section 4.

2. LOWER BOUND ON BREAKDOWN PROBABILITY

Let $\{\tilde{\lambda}_k\}_{k=1}^m$, $\tilde{S}$, and $\tilde{G}$ denote the estimates of $\{\lambda_k\}_{k=1}^m$, S, and G derived from the eigenelements of $\hat{R}$ by *correct association* of the eigenpairs of $\hat{R}$ and R. So $\{\tilde{\lambda}_k\}_{k=1}^m$ is equal to $\{\hat{\lambda}_k\}_{k=1}^m$ if no swap has occurred otherwise it is a rearrangement of it. Denote the smallest correctly associated signal eigenvalue by $\tilde{\lambda}_{\text{smin}} = \min\{\tilde{\lambda}_k\}_{k=1}^n$, and the largest correctly associated noise eigenvalue by $\tilde{\lambda}_{\text{nmax}} = \max\{\tilde{\lambda}_k\}_{k=n+1}^m$. A subspace swap occurs whenever $\tilde{\lambda}_{\text{smin}}$ is less than $\tilde{\lambda}_{\text{nmax}}$, so it occurs with probability

$$\text{P}_{\text{SS}} = \text{Prob}\ \{\tilde{\lambda}_{\text{smin}} < \tilde{\lambda}_{\text{nmax}}\}. \tag{2.4}$$

If the signal eigenvalues $\{\lambda_k\}_{k=1}^n$ are well separated, then the dominant event in (2.4) corresponds to the case in which $\tilde{\lambda}_n < \tilde{\lambda}_{\text{nmax}}$.

To study the probability in (2.4), we need to know the distribution of $\{\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n, \tilde{\lambda}_{\text{nmax}}\}$. The asymptotic (with respect to N) distribution of $\{\tilde{\lambda}_k\}_{k=1}^n$ is easily derived. The asymptotic distribution of $\tilde{\lambda}_{\text{nmax}}$ can also be derived but it is comparatively complicated [4]. Since we aim at providing a simple formula for predicting a performance breakdown, we replace $\tilde{\lambda}_{\text{nmax}}$ in (2.4) by

$$\tilde{\sigma}^2 \triangleq \frac{1}{m-n} \sum_{k=n+1}^{m} \tilde{\lambda}_k. \tag{2.5}$$

Hence, in lieu of (2.4) we consider

$$\text{P}_{\text{LB}} \triangleq \text{Prob}\ \{\tilde{\lambda}_{\text{smin}} < \tilde{\sigma}^2\} \tag{2.6}$$

$$= 1 - \text{Prob}\ \{\tilde{\lambda}_k \geq \tilde{\sigma}^2, \text{ for all } k \in \{1, \ldots, n\}\}. \tag{2.7}$$

As $\tilde{\sigma}^2 \leq \tilde{\lambda}_{\text{nmax}}$ it follows that P_{LB} is a *lower bound* on P_{SS}

Our approach to studying breakdown probability differs in two ways from that in [2]. In the latter approach P_{SS} is lower bounded by

$$\text{P}_{\text{SS}} \geq \text{Prob}\ \{\boldsymbol{v}_n^H \hat{R} \boldsymbol{v}_n < \frac{1}{m-n} \sum_{k=n+1}^{m} \boldsymbol{v}_k \hat{R} \boldsymbol{v}_k\}, \tag{2.8}$$

which is asymptotically (in N) equal to $\text{Prob}\ \{\tilde{\lambda}_n < \tilde{\sigma}^2\}$, which is a lower bound for P_{LB} and hence P_{SS}. However, generally (2.8) is a less tight bound than P_{LB}. Secondly, although the random variables in (2.8) can be evaluated "exactly" (to within some inherent numerical approximations [2]), the exact expression for (2.8) is quite complex and unenlightening. More importantly, it depends on the eigenvectors $\{\boldsymbol{v}_k\}_{k=n}^m$ and so cannot be well approximated in the breakdown region, implying that it cannot be employed to predict the breakdown effect from the observed data alone.

Theorem 2.1. Under the above assumptions on R and its eigenvalues, and assuming that the snapshots are circularly symmetric Gaussian distributed and temporally white, the asymptotic (in N) distribution of the $n \times 1$ element vector

$$\boldsymbol{\alpha} \triangleq \sqrt{N}[(\tilde{\lambda}_1 - \lambda_1), \ldots, (\tilde{\lambda}_n - \lambda_n), (\tilde{\sigma}^2 - \sigma^2)]^T \tag{2.9}$$

is Gaussian with zero mean and covariance matrix

$$\begin{bmatrix} \Lambda^2 & 0 \\ 0 & \sigma^4/(m-n) \end{bmatrix}, \tag{2.10}$$

where $\Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_n\}$.
Proof: See [5].

The above theorem can be used to compute P_{LB} as follows. Define

$$z_k \triangleq (\alpha_k - \alpha_{n+1})/(\lambda_k^2 + \sigma^4/(m-n))^{1/2} \tag{2.11}$$

$$\mu_k \triangleq \sqrt{N}(\sigma^2 - \lambda_k)/(\lambda_k^2 + \sigma^4/(m-n))^{1/2}, \tag{2.12}$$

for $k = 1, \ldots, n$. It follows from (2.6) that

$$\text{P}_{\text{LB}} = 1 - \text{Prob}\ \{z_k \geq \mu_k \text{ for all } k \in \{1, \ldots, n\}\}. \tag{2.13}$$

The random variables $\{z_k\}_{k=1}^n$ are asymptotically Gaussian distributed with zero mean, unit variance, and so-called "product form" correlations $\text{E}\{z_i z_j\} = \beta_i \beta_j$, for $i \neq j$, where

$$\beta_i = \sigma^2/(\sigma^4 + (m-n)\lambda_i^2)^{1/2}. \tag{2.14}$$

It follows that computation of P_{LB} can be reduced to the calculation of the probability that n Gaussian random variables with zero means, unit variances and the above product correlation structure simultaneously take on values larger than the constants $\{\mu_k\}_{k=1}^n$ defined in (2.12). This probability can be efficiently computed using the MVTIN algorithm in [6], which is specifically designed for Gaussian random variables with product correlation structure.

To use the data to make inferences about the probability of a performance breakdown we first consider replacing $\{\lambda_k\}_{k=1}^n$ and σ^2 in equations (2.12) and (2.14) by $\{\hat{\lambda}_k\}_{k=1}^n$ and

$$\hat{\sigma}^2 = \frac{1}{m-n} \sum_{k=n+1}^{m} \hat{\lambda}_k. \tag{2.15}$$

Hence, we P_{LB} could be estimated by

$$\hat{\text{P}}_{\text{LB}} = 1 - \text{Prob}\ \{z_k \geq \hat{\mu}_k \text{ for all } k \in \{1, \ldots, n\}\} \tag{2.16}$$

where

$$\hat{\mu}_k = \sqrt{N}(\hat{\sigma}^2 - \hat{\lambda}_k)/(\hat{\lambda}_k^2 + \hat{\sigma}^4/(m-n))^{1/2} \tag{2.17}$$

and $\{z_k\}$ are Gaussian random variables with zero mean, unit variance, and product correlation structure given by

$$\text{E}\{z_i z_j\} = \hat{\beta}_i \hat{\beta}_j \qquad \text{for } i \neq j, \tag{2.18}$$

with

$$\hat{\beta}_i = \hat{\sigma}^2/(\hat{\sigma}^4 + (m-n)\hat{\lambda}_i^2)^{1/2}. \tag{2.19}$$

Outside the breakdown region the above formula yields a consistent estimate of P_{LB}. In the breakdown region, however, Monte

Carlo simulations show that $\hat{\mathrm{P}}_{\mathrm{LB}}$ appears to seriously underestimate P_{LB} (see Section 4). There are two reason for this. Intuitively, when there is a non-negligible probability of a subspace swap, $\hat{\sigma}^2$ tends to underestimate σ^2, similarly $\hat{\lambda}_n$ may overestimate λ_n. This combination causes $\hat{\mu}_n$ to significantly underestimate μ_n and, hence, the observed effect. The remaining $\hat{\mu}_k$ will similarly underestimate μ_k, but the overestimation is most pronounced for $\hat{\mu}_n$.

To try to correct for the overestimation of λ_n, we will overestimate σ^2 with the intention that the resulting $\hat{\mu}_n$ will be larger and more accurately estimate μ_n. Therefore, we propose using $\hat{\lambda}_{n+1}$ as the estimate of σ^2 that we use in equations (2.17) and (2.19). We denote the resulting estimate $\tilde{\mathrm{P}}_{\mathrm{LB}}$. Well above the threshold region we expect $\tilde{\mathrm{P}}_{\mathrm{LB}}$ to estimate P_{LB} very accurately. Below the threshold region, it should certainly be a better estimate than $\hat{\mathrm{P}}_{\mathrm{LB}}$, however, its performance in the transition region is not obvious. Therefore Monte Carlo simulations are again be used to asses its efficacy.

3. PREDICTING AND COUNTERACTING BREAKDOWN

For a given m and n, let $\mathcal{G}$ denote the set of all possible partitionings of the eigenelements of $\hat{R}$ into signal and noise subspaces, and let Γ be an element of that set. Denote by $\hat{\boldsymbol{\theta}}_{\Gamma,\mathcal{M}}$, the estimate of $\boldsymbol{\theta}$ given by method $\mathcal{M}$ using the partitioning Γ. Now the deterministic (concentrated) maximum likelihood (DML) method estimates $\boldsymbol{\theta}$ as the argument that minimizes the criterion function $-\operatorname{tr}[A(A^HA)^{-1}A^H\hat{R}]$ e.g. [7]. This minimization requires an n-dimensional search. The DML criterion suggests the followed modified estimator, which we shall term the DML-$\mathcal{M}$ estimator

$$\hat{\boldsymbol{\theta}}_{\mathrm{DML}} = \underset{\Gamma\in\mathcal{G}}{\operatorname{argmin}} -\operatorname{tr}[A(\boldsymbol{\theta})(A^H(\boldsymbol{\theta})A(\boldsymbol{\theta}))^{-1}A^H(\boldsymbol{\theta})\hat{R}], \quad (3.20)$$

In other words we just test the DML criterion at a finite number of points determined by the method $\mathcal{M}$ under the various possible partitionings. Since the DML estimator contains no discontinuous assignment of eigenvectors in signal and noise subspaces, we expect that, generally, the partitioning that leads to the value of $\boldsymbol{\theta}$ that minimizes (3.20), $\bar{\Gamma}$ say, is in fact equal to the correctly assigned partitioning $\tilde{\Gamma}$. If no subspace swap is detected $\bar{\Gamma} = \hat{\Gamma}$ and the DML-$\mathcal{M}$ estimator is the same as that obtained by regular application of $\mathcal{M}$.

Note that $\mathcal{G}$ contains $\binom{m}{n}$ elements, however, unless the probability of a subspace swap is very high we may reduce the computational burden somewhat by safely assuming that the $l < m-n$, say, smallest eigenelements $\{\hat{\lambda}_k, \hat{\boldsymbol{v}}_k\}_{k=m-l+1}^{m}$ are indeed estimates of noise subspace eigenelements. We can then remove from $\mathcal{G}$ all partitions in which one or more of these eigenelements lies in the signal subspace, so reducing its cardinality to $\binom{m-l}{n}$.

4. NUMERICAL EXAMPLES AND DISCUSSION

For the examples in the Section we consider $n = 2$ narrowband Gaussian signals impinging on a $m = 10$ element uniform linear array (ULA) with half wavelength spacing. The unit power sources are located at broadside and 10° from broadside and are highly correlated,

$$\mathrm{E}\, s_1(t)s_2^H(t) = \sqrt{0.9}e^{i\pi/4} \quad (4.21)$$

where $s_1(t)$ and $s_2(t)$ are the source signals. The noise is assumed to be spatially white and $N = 150$ snapshots are used.

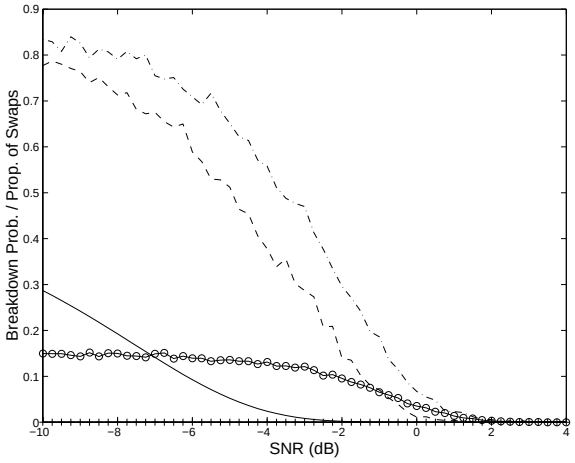

Fig. 1. Lower bound P_{LB} (solid) and empirical estimate of P_{SS} (dash), mean values of $\hat{\mathrm{P}}_{\mathrm{LB}}$ (+) and $\tilde{\mathrm{P}}_{\mathrm{LB}}$ (o), and proportional of subspace swaps indicated by DML criterion (dash-dot) over 1000 Monte Carlo trials.

Figure 1 shows the value of P_{LB}, as given by (2.13), versus SNR. Also shown is a Monte Carlo determination of P_{SS} defined as the number of runs in which an estimated noise eigenvector was closer to the true signal subspace (as measured by the magnitude of its projection onto SS^H) than one of the estimated signal eigenvectors. That P_{LB} is a lower bound for the subspace swap probability is clear.

The estimates $\tilde{\mathrm{P}}_{\mathrm{LB}}$ (which uses $\hat{\sigma}_2 = \hat{\lambda}_{n+1}$) and $\hat{\mathrm{P}}_{\mathrm{LB}}$ (using $\hat{\sigma}_2 = \sum_{k=n+1}^{m}\hat{\lambda}_k/(m-n)$) also appear in Figure 1 (see Figure 2 as well). Clearly, $\hat{\mathrm{P}}_{\mathrm{LB}}$ grossly underestimates P_{LB} in and below the threshold region (note the scale on the upper part of Fig. 2). Furthermore, the standard deviation of $\hat{\mathrm{P}}_{\mathrm{LB}}$ is larger than its mean until 1-2dB above breakdown, at which point it becomes negligible. Therefore, it is possible to set a threshold such that when operating above breakdown the probability that $\hat{\mathrm{P}}_{\mathrm{LB}}$ exceed the threshold (i.e. a false alarm) is small. However, when the system is operating below breakdown, there will always be a significant probability that $\hat{\mathrm{P}}_{\mathrm{LB}}$ will fall below this threshold (a missed event), for any sensible choice of false alarm probability. Note that $\hat{\mathrm{P}}_{\mathrm{LB}}$ tends to a finite value with increasing SNR, in fact it tends to P_{LB} (equation (2.6)) which is also finite for finite N.

On the other hand $\tilde{\mathrm{P}}_{\mathrm{LB}}$ overestimates P_{LB} above breakdown, and it tends to underestimate P_{LB} below breakdown, though is does seem to be reasonably similar to P_{SS} through the transition region. Several dB above breakdown $\tilde{\mathrm{P}}_{\mathrm{LB}}$ and its standard deviation become negligible. Below breakdown, although the standard deviation of

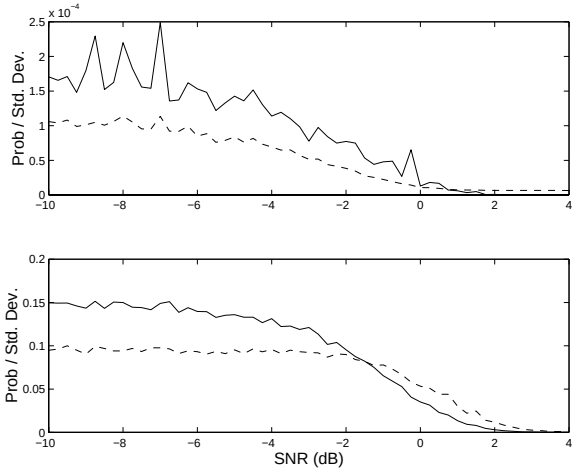

Fig. 2. Mean (solid) and standard deviation (dash) of $\hat{\mathrm{P}}_{\mathrm{LB}}$ (top) and $\tilde{\mathrm{P}}_{\mathrm{LB}}$ (bottom) from Monte Carlo trials.

$\tilde{\mathrm{P}}_{\mathrm{LB}}$ is significant, unlike the case of $\hat{\mathrm{P}}_{\mathrm{LB}}$ it is smaller than its mean. Therefore, if we again set a threshold to achieve a given level of false alarms when operating at a certain SNR above breakdown, a detector based on $\tilde{\mathrm{P}}_{\mathrm{LB}}$ should have fewer missed events when the scenario is really below breakdown that should a detector based on $\hat{\mathrm{P}}_{\mathrm{LB}}$ designed with the same false alarm rate, i.e. the detector based on $\tilde{\mathrm{P}}_{\mathrm{LB}}$ is more powerful.

Figure 3 shows the performance of the standard ESPRIT, and DML-ESPRIT estimators determined from 1000 Monte Carlo trials. The measure of performance is the square root of the mean-squared error (MSE) averaged over both sources. It is clear that DML-ESPRIT outperforms standard ESPRIT. Furthermore, the DML-ESPRIT criterion can be used to determine whether or not a subspace swap has occurred. To illustrate this, the same figure shows the performance of standard ESPRIT averaged over those runs in which the DML criterion indicated no subspace swap had occurred. It is seen that removing the detected subspace swap cases improves the performance quite significantly. However, note that none of the estimators are close to the CRB in the breakdown region. This is partly because the ESPRIT algorithm is not statistically efficient, and partly because the CRB is only a tight bound in small error situations. Tighter lower bounds, such as the large-error Barankin bound [9] and its variants, may be considerably larger than the CRB in the breakdown region. Finally, we notice that the proportion of trials in which DML-ESPRIT indicated a subspace swap (see Fig. 1) is remarkably close to the empirical P_{SS}.

5. CONCLUSION

This paper has presented a simple theoretical lower bound on the probability of a subspace swap. Using this bound we found a data-based statistic that can be used to determine whether the probability of breakdown is low or high. A method for deciding whether a subspace breakdown has actually occurred and a means to correct for it were also presented.

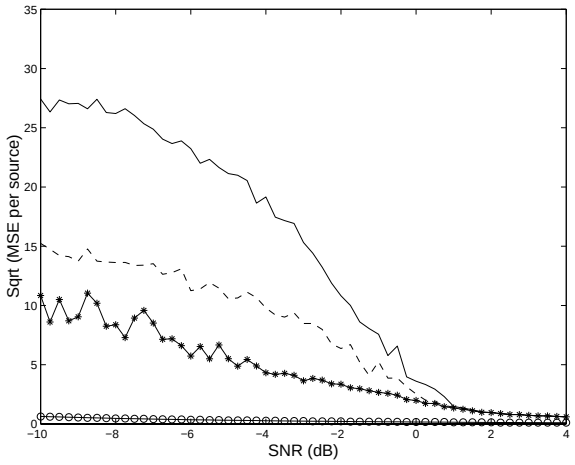

Fig. 3. Performance of standard ESPRIT (solid) and DML-ESPRIT (dash-dot). Also shown is the CRB (o) and the performance of standard ESPRIT averaged over those runs in which the and DML criterion indicated no subspace swap ($*$).

6. REFERENCES

[1] D.W. Tufts, A.C. Kot, and R.J. 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, NY: Elsevier, 1991.

[2] J.K. Thomas, L.L. Scharf, and D. W Tufts, "The probability of a subspace swap in the SVD," *IEEE Trans. Signal Processing*, vol. 43, pp. 730–736, Mar. 1995.

[3] P. Stoica, V. Simonyte, and T. Söderström, "On the resolution performance of spectral analysis," *Signal Processing*, vol. 44, pp. 153–161, June 1995.

[4] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 2nd Ed., New York, NY: Wiley, 1984.

[5] M. Hawkes, P. Stoica, A. Nehorai, "Performance breakdown of subspace-based methods: prediction and cure," in preparation.

[6] C.W. Dunnett, "Multivariate normal probability integrals with product correlation structure,"*Appl. Stats.*, vol. 38, pp. 564–579, 1989. See also correction in *Appl. Stats.*, vol. 42, pp. 709, 1993.

[7] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér-Rao bound," *IEEE Trans. Acoust., Speech, & Sig. Proc.*, vol. 37, pp 720–741, May 1989.

[8] R. Roy and T. Kailath, "ESPRIT - estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, & Sig. Proc.*, vol. 37, pp. 984–995, July 1989.

[9] E.W. Barankin, "Locally best unbiased estimates," *Ann. Math. Stat.*, vol. 20, pp. 477-501, 1949.