Determination of the Dimension of a Signal Subspace from Short Data Records

Abhijit A. Shah and Donald W. Tufts

Abstract— **A nested sequence of constant false alarm rate (CFAR) hypothesis tests is presented for rank determination over *short data records*. The procedure is based on the interpretation of sum of squares of singular values as energy in a particular subspace. The CFAR thresholds are set up based on distributions derived from matrix perturbation ideas. Expressions for probability of overestimation and underestimation are presented.**

I. Introduction

In many problems in signal processing, the observed data can be modeled as a linear combination of finite number of signals corrupted with additive noise. The number of signals, also known as the model order, may be unknown. The knowledge of number of signals or model order is essential information for signal inference algorithms. Here we present a way to estimate the model order when a small number of independent observations of the data vector are available for decision.

For such a case, several automated procedures exist. The procedures can be categorized into two distinct classes: Nonparametric methods and Parametric methods. The early procedures stem from the development of likelihood ratio tests for testing the equality of a subset of squares of singular values of the data matrix [1], [2]. Nonparametric approaches based on hypothesis tests for signal processing applications are [3], [4]. Procedures based on the information theoretic criterion developed by Akaike (AIC) [5] and Rissanen's minimum description length (MDL) [6] also exist, see [7], [8]. A Bayesian approach has been proposed [9]. Parametric approaches that assume knowledge of signal structure are also been proposed [10]–[12, pp. 498].

The nonparametric methods mentioned above suffer from a drawback. The performance of the nonparametric methods degrades significantly under *short data record* conditions, even at moderate to high SNR's. Thus, there is a need for a widely applicable nonparametric procedure that performs well over short data records. We present such a nonparametric method, which is widely applicable, and performs well for short data records over moderate to high SNR's. The proposed method consists of sequential constant false alarm rate (CFAR) tests on the sums of squares of singular values of the data matrix. The thresholds are based on approximate distributions of the sums. The approximations are derived using matrix perturbation ideas, and hold true over short data records and a wide range of SNR, resulting in improved performance for short data records.

In order determination there are two basic types of error probabilities, error due to overestimation and error due to underestimation. The implications and importances of these two errors can be radically different in different applications. The proposed method allows the user to set a bound on error due to overestimation which we call the *false alarm probability*. The user can then determine a value of SNR for which a prescribed value of probability of detection or probability of net error can be obtained. Thus, the user can specify the conditions under which performance goals, specified by error probabilities, can be obtained.

Manuscript received August 8, 1993; revised November 11, 1993. This work was supported by the Naval Research Laboratory under contract N00014-89-K-6003, Donald W. Tufts, Principal Investigator. The associate editor coordinating the review of this paper and approving it for publication was Prof. Kevin M. Buckley.

A. A. Shah is with Wireless Access, Inc., Santa Clara, CA 95054 USA.

D. W. Tufts is with the Department of Electrical Engineering, University of Rhode Island, Kingston, RI 02881 USA.

IEEE Log Number 9403265.

II. Statement of the Problem

A data vector $\tilde{\mathbf{y}}_k = \mathbf{y}_k + \mathbf{n}_k$ is described by the following model

$$\mathbf{y}_k = \mathbf{H}(\beta)\theta_k; \quad \mathbf{H}(\beta) \in \mathbf{C}^{m \times r}; \quad \mathbf{n}_k{:}CN \sim [0, \sigma^2_{\text{noise}}\mathbf{I}] \tag{1}$$

where $\mathbf{y}_k$ is the signal component at instant k, $\mathbf{H}(\beta)$ is the modal matrix [12] with r linearly independent columns, and θ_k is $r \times 1$ vector of coefficients that are samples of a stationary random process, with zero mean and positive definite covariance matrix. The dimension of θ_k (model order) given by r ($r < m$) is unknown. Also, some or all of the model parameters β are unknown. Neither do we assume any functional form for $\mathbf{H}(\beta)$ (nonparametric assumption). We have N such observations of $\tilde{\mathbf{y}}_k$ from which we have to determine the order. We represent the data as follows

$$\begin{aligned}\tilde{\mathbf{Y}}_{m \times N} &= [\tilde{\mathbf{y}}_1|\tilde{\mathbf{y}}_2|\cdots|\tilde{\mathbf{y}}_N] \\ &= \underbrace{\mathbf{H}(\beta)[\theta_1|\theta_2|\cdots|\theta_N]}_{\mathbf{Y}} \\ &\quad + \underbrace{[\mathbf{n}_1|\mathbf{n}_2|\cdots|\mathbf{n}_N]}_{\mathbf{N}}.\end{aligned} \tag{2}$$

Since $\min(m, N) > r$, the rank of the signal-only data matrix $\mathbf{Y}$ will be r, equal to the number of signals. Thus we formulate our problem of order determination as a problem of determining the rank of the signal-only matrix $\mathbf{Y}$. Approaches to solve this problem are based on the singular values of the covariance matrix of $\tilde{\mathbf{y}}_k$, given by $\mathbf{R}$. Its singular value decomposition (SVD) is

$$\begin{aligned}\mathbf{R} &= \mathbf{H}(\beta)\mathcal{E}\left(\theta_k\theta_k^H\right)\mathbf{H}(\beta)^H + \sigma^2\mathbf{I} \\ &= \mathbf{U}_1\Lambda_1\mathbf{U}_1^H + \mathbf{U}_2\Lambda_2\mathbf{U}_2^H\end{aligned} \tag{3}$$

where $\mathbf{U}_1$ is an $m \times r$ matrix consisting of the first r singular vectors of $\mathbf{R}$. They span the column-space of $\mathbf{H}(\beta)$. $\mathbf{U}_2$ contains the remaining $(m - r)$ singular vectors. We denote the space spanned by $\mathbf{U}_1$ as the *signal subspace*, and the space spanned by $\mathbf{U}_2$ as the *orthogonal subspace*. Λ_1 contains the first r *signal* singular values $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_r$, and Λ_2 contains the remaining $m - r$ *noise* singular values that are all equal to σ^2. Thus, an intuitively appealing way to determine the model order is by testing for multiplicity of the smallest singular value of $\mathbf{R}$. However, in practice, the covariance matrix $\mathbf{R}$ is unknown, and the tests are carried out on the singular values of the sample covariance matrix, which is equivalent to carrying out tests on the squares of singular values of the data matrix $\mathbf{Y}$. The SVD of $\tilde{\mathbf{Y}}$ is defined as follows

$$\begin{aligned}&\tilde{\mathbf{Y}}_{(m \times N)} \\ &\quad = \left[\tilde{\mathbf{U}}_s \mid \tilde{\mathbf{U}}_0\right]_{(m \times m)} \begin{bmatrix} \tilde{\Sigma}_s & 0 \\ 0 & \tilde{\Sigma}_0 \end{bmatrix}_{(m \times N)} \left[\tilde{\mathbf{V}}_s|\tilde{\mathbf{V}}_0\right]^H_{(N \times N)}\end{aligned} \tag{4}$$

where $\tilde{\Sigma}_s = \text{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r)$, $\tilde{\Sigma}_0 = \text{diag}(\sigma_{r+1}, \sigma_{r+2}, \cdots, \sigma_{\min(m,N)})$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(m,N)}$, and $[\tilde{\mathbf{U}}_s|\tilde{\mathbf{U}}_0]$, $[\tilde{\mathbf{V}}_s|\tilde{\mathbf{V}}_0]$ are unitary matrices. Note that the $\{\sigma_k\}$ are the singular values of $\tilde{\mathbf{Y}}$. The order determination methods perform various

tests on the squares of these data singular values to determine the underlying multiplicity of population singular values of the covariance matrix, i.e., the model order.

III. Description of the Method

Under high SNR conditions, the *perturbed signal subspace*, spanned by columns of $\hat{\mathbf{U}}_s$ is stable and is more or less determined by the *signal subspace*, the column-space of $\mathbf{U}_1$, i.e., the space spanned by the modes $\mathbf{H}(\beta)$ of the model. This in turn stabilizes the *perturbed orthogonal subspace*, i.e., the space spanned by $\hat{\mathbf{U}}_0$. Hence, in different realizations, the singular vectors of $\hat{\mathbf{U}}_0$ may change erratically but the space spanned by $\hat{\mathbf{U}}_0$ remains relatively unchanged, and is closely related to the the column-space of $\mathbf{U}_2$, the orthogonal subspace. Thus the energy in the perturbed orthogonal subspace given by

$$S_r = \sum_{i=r+1}^{m} \sigma_i^2 = \|\hat{\mathbf{U}}_0 \hat{\mathbf{U}}_0^H \hat{\mathbf{Y}}\|_F^2 \tag{5}$$

is also well-defined. It is closely related to the noise energy in the subspace spanned by $\mathbf{U}_2$, the orthogonal subspace. The noise energy is given by $(m-r)N\sigma^2$, N times the sum of last $m-r$ singular values of $\mathbf{R}$ that correspond to $\mathbf{U}_2$. In fact, in the next section using matrix perturbation approximations we quantify the above idea and evaluate the distribution of $\frac{2}{\sigma^2}S_r$ to be $\chi^2_{2(m-r)(N-r)}$, a central χ^2 with $2(m-r)(N-r)$ degrees of freedom.

Based on this distribution, we can set a threshold T_r such that $S_r < T_r$ with a probability $1-\alpha$ where α is a small positive number.[1] In other words, if rank is r then S_r can be well-explained by the noise energy alone and will be below this threshold with high probability. If rank is $r+1$ or greater then, due to the additional signal energy, S_r will exceed the threshold with high probability. Based on this idea, we develop a recursive procedure on the set of sums of squares of singular values of data matrix $\mathbf{Y}$, that is essentially a signal energy detection procedure in enlarging subspaces.

Outline of the Procedure: We define H_k as a hypothesis with signal rank being k,

$$H_k: \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > \lambda_{k+1} = \lambda_{k+2} = \cdots = \lambda_m = \sigma^2. \tag{6}$$

We define $\{S_k\}_{k=0}^{m-1}$ as the cumulative sums of squared singular values of the data matrix $\mathbf{Y}$. The kth sum is defined as $S_k = \sum_{i=k+1}^{m} \sigma_i^2$. Thus, the first element S_0 in the sequence is the sum of squares of all singular values, and the last element S_{m-1} in the sequence is the square of the smallest singular value. We test on $\{S_k\}_{k=0}^{m-1}$ sequentially, starting from the last element S_{m-1}. The first test in the nested sequence of tests is as follows

$$\begin{aligned} 1^{\text{st}} \text{ test: } H_1^0&: \text{Choose } H_0 \cup H_1 \cup \cdots \cup H_{m-1} \\ &\quad \text{if } S_{m-1} < T_{m-1} \\ H_1^1&: \text{Choose } H_m \\ &\quad \text{if } S_{m-1} \geq T_{m-1}. \end{aligned} \tag{7}$$

If we choose H_m then we stop and decide that the rank is m. Otherwise, we conduct the next test on S_{m-2} and so on. The kth test in the procedure will be

$$\begin{aligned} k\text{th test: } H_k^0&: \text{Choose } H_0 \cup H_1 \cup \cdots \cup H_{m-k} \\ &\quad \text{if } S_{m-k} < T_{m-k} \\ H_k^1&: \text{Choose } H_{m-k+1} \\ &\quad \text{if } S_{m-k} \geq T_{m-k}. \end{aligned} \tag{8}$$

We keep testing until we choose H_k^1 for some k and decide that the rank is $m-k+1$. We select the set of thresholds $\{T_k\}_{k=0}^{m-1}$ according to a level α such that

$$\begin{aligned} P(S_0 \geq T_0 \mid H_0) &= P(S_1 \geq T_1 \mid H_1) \\ &= \cdots = P(S_{m-1} \geq T_{m-1} \mid H_{m-1}). \\ &= \alpha. \end{aligned} \tag{9}$$

At any given test, we interpret the level α as the corresponding individual false alarm level for that test, because $P(S_k \geq T_k \mid H_k)$ is the probability of choosing H_{k+1} when H_k is true. As said earlier, under hypothesis H_k (the number of signals or the signal—only matrix rank is k), S_k corresponds to the energy in the perturbed orthogonal subspace. $\frac{2}{\sigma^2}S_r$ has an approximate distribution of $\chi^2_{2(m-k)(N-k)}$. This is a twell-tabulated probability distribution. For a given level α we can easily obtain a set of thresholds $\{T_k\}_{k=0}^{m-1}$.

In addition, we denote the probability of overestimating the rank as P_{FA} the probability of false alarm for the whole procedure. Note that α and P_{FA} are probabilities of two different events and should not be confused with each other. However, it turns out (see Section V) that probabilities of these two events are approximately equal, and for an acceptable level of Probability of overall false alarm, we can choose an appropriate α.

IV. Perturbation Analysis and Distributions

As said earlier, for setting up thresholds we need to evaluate the distribution of S_r, the sum of squares of noise singular values given that the true rank is r. In the first part of this section, we use matrix perturbation approximations, similar to the approximations proposed by Tufts, Vaccaro and Li [13], to approximate $\hat{\mathbf{U}}_0\hat{\mathbf{U}}_0^H$ and thereby obtain an approximation to S_r given that the true rank is r. In the later part of the section, we show that the approximation $\frac{2}{\sigma^2}S_r$ follows $\chi^2_{2(m-r)(N-r)}$.

First, we obtain an approximation to the perturbed noise subspace, i.e., the space spanned by columns of $\hat{\mathbf{U}}_0$. We say that for SNR above threshold [14] the perturbed noise subspace is spanned by $\mathbf{U}_2 + \mathbf{U}_1\mathbf{X}$ where $\mathbf{X}$ has norm in the order of $\|\mathbf{N}\|_F$. In other words, $\hat{\mathbf{U}}_0$ is well-approximated by $(\mathbf{U}_2 + \mathbf{U}_1\mathbf{X})\mathbf{Q}$, where $\mathbf{Q}$ is an arbitrary $m-r \times m-r$ unitary matrix. We obtain $\mathbf{X}$ by using the following definition of orthogonal subspace

$$\hat{\mathbf{U}}_0 = \arg\min_{\mathbf{W}} \|\mathbf{W}^H \hat{\mathbf{Y}}\|_F^2. \tag{10}$$

The above minimization is carried out with the constraint $\mathbf{W}^H\mathbf{W} = \mathbf{I}$. We substitute $\mathbf{W}$ by $\mathbf{U}_2 + \mathbf{U}_1\mathbf{X}$, and solve the above formula for $\mathbf{X}$. To simplify the notation, we represent the data matrix $\hat{\mathbf{Y}}$ of (2) as follows

$$\begin{aligned} \hat{\mathbf{Y}} &= (\mathbf{U}_1\mathbf{U}_1^H + \mathbf{U}_2\mathbf{U}_2^H)(\mathbf{Y} + \mathbf{N}) \\ &= [\mathbf{U}_1 \mid \mathbf{U}_2]\begin{bmatrix}\mathbf{A} \\ \mathbf{0}\end{bmatrix} \\ &\quad + [\mathbf{U}_1 \mid \mathbf{U}_2]\begin{bmatrix}\mathbf{N}_1 \\ \mathbf{N}_2\end{bmatrix} \end{aligned} \tag{11}$$

where $\mathbf{A} = \mathbf{U}_1^H\mathbf{Y}$ is of size $r \times N$ with statistically independent columns. Each column $\mathbf{a}_i$ is Gaussian distributed with zero mean and $\mathcal{E}(\mathbf{a}_i\mathbf{a}_i^H) = \mathcal{E}(\mathbf{U}_1^H\mathbf{y}_i\mathbf{y}_i^H\mathbf{U}_1) = \Lambda_1$. $\mathbf{N}_{1\,r\times N}$ and $\mathbf{N}_{2\,(m-r)\times N}$ consists of *iid* zero mean and variance σ^2 Gaussian distributed elements. Thus, (10) modifies to

$$\begin{aligned} \hat{\mathbf{U}}_0 &= \arg\min_{\mathbf{X}} \|(\mathbf{U}_2 + \mathbf{U}_1\mathbf{X})^H(\mathbf{U}_1\mathbf{A} + \mathbf{U}_1\mathbf{N}_1 + \mathbf{U}_2\mathbf{N}_2)\|_F^2 \\ &= \arg\min_{\mathbf{X}} \|\mathbf{X}^H(\mathbf{A} + \mathbf{N}_1) + \mathbf{N}_2\|_F^2. \end{aligned} \tag{12}$$

[1] The calculation of T_k requires the knowledge of noise level σ^2. The noise level is assumed to be known or estimated independently. If the noise level is unknown, an estimate obtained thorough Fourier analysis could be used. If from the physics of the problem, it is known that for some $r_{\max}$ we have $r < r_{\max}$, then one can estimate the noise level from $\{\sigma_i^2\}_{i=r_{\max}+1}^{m}$.

The unconstrained solution is $\mathbf{X} = -((\mathbf{A}+\mathbf{N}_1)^H)^\dagger \mathbf{N}_2^H$, where † implies *pseudo inverse*. Hence, we approximate $\tilde{\mathbf{U}}_0$ as follows

$$\tilde{\mathbf{U}}_0 = \left(\mathbf{U}_2 - \mathbf{U}_1\left((\mathbf{A}+\mathbf{N}_1)^H\right)^\dagger \mathbf{N}_2^H\right)\mathbf{Q}. \tag{13}$$

$\tilde{\mathbf{U}}_0$ is supposed to be unitary, the above approximation is unitray to the first order. Using the above approximation we approximate S_r, the sum of squares of noise singular values as follows

$$\begin{aligned} S_r &= \|\tilde{\mathbf{U}}_0\tilde{\mathbf{U}}_0^H\hat{\mathbf{Y}}\|_F^2 \\ &= \|\tilde{\mathbf{U}}_0^H\hat{\mathbf{Y}}\|_F^2 \\ &= \|\mathbf{Q}^H(\mathbf{U}_2+\mathbf{U}_1\mathbf{X})^H(\mathbf{Y}+\mathbf{N})\|_F^2 \\ &= \|\mathbf{N}_2+\mathbf{X}^H(\mathbf{A}+\mathbf{N}_1)\|_F^2 \end{aligned} \tag{14}$$

where $\mathbf{X} = -(\mathbf{N}_2(\mathbf{A}+\mathbf{N}_1)^\dagger)^H$. On substituting this approximation of $\mathbf{X}$ we get

$$\begin{aligned} S_r &= \mathrm{tr}\left(\mathbf{N}_2\mathbf{N}_2^H - \mathbf{N}_2(\mathbf{A}+\mathbf{N}_1)^\dagger(\mathbf{A}+\mathbf{N}_1)\mathbf{N}_2^H\right) \\ &= \mathrm{tr}\Big(\mathbf{N}_2\mathbf{N}_2^H - \mathbf{N}_2(\mathbf{A}+\mathbf{N}_1)^H \\ &\quad \times \left((\mathbf{A}+\mathbf{N}_1)(\mathbf{A}+\mathbf{N}_1)^H\right)^{-1}\cdot(\mathbf{A}+\mathbf{N}_1)\mathbf{N}_2^H\Big). \end{aligned} \tag{15}$$

Since $(\mathbf{A}+\mathbf{N}_1)$ can be considered as N independent observations of a multivariate Gaussian distribution, $(\mathbf{A}+\mathbf{N}_1)$ has full row rank and $(\mathbf{A}+\mathbf{N}_1)^\dagger = (\mathbf{A}+\mathbf{N}_1)^H((\mathbf{A}+\mathbf{N}_1)(\mathbf{A}+\mathbf{N}_1)^H)^{-1}$.

We have obtained an approximation to S_r, the sum of squares of noise singular values. Formulas (14) and (15) bear close resemblance to matrix regression expressions and *Wishart distributions*. By using the following linear transformation we make the relationship more apparent. We define $\mathbf{B}_{m\times N}$ as follows

$$\mathbf{B} = [\mathbf{U}_1 \mid \mathbf{U}_2]^H\hat{\mathbf{Y}} = \begin{bmatrix}\mathbf{A}+\mathbf{N}_1 \\ \mathbf{N}_2\end{bmatrix}. \tag{16}$$

We also define $\mathbf{S} = \mathbf{B}\mathbf{B}^H$ and partition it as follows

$$\begin{aligned} \mathbf{S} &= \begin{bmatrix}\mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22}\end{bmatrix} \\ &= \begin{bmatrix}(\mathbf{A}+\mathbf{N}_1)(\mathbf{A}+\mathbf{N}_1)^H & (\mathbf{A}+\mathbf{N}_1)\mathbf{N}_2^H \\ \mathbf{N}_2(\mathbf{A}+\mathbf{N}_1)^H & \mathbf{N}_2\mathbf{N}_2^H\end{bmatrix}. \end{aligned} \tag{17}$$

Using this representation we can write S_r as

$$S_r = \mathrm{tr}\left(\mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\right). \tag{18}$$

If we interpret $\mathbf{B}$ as N independent realizations of $\mathbf{b}_{m\times 1} \sim CN(0,\Lambda)$ then we realize that $\mathbf{S}$ has Wishart distribution $W_m(N,\Lambda)$. The matrix $\mathbf{S}_{22} - \mathbf{S}_{12}\mathbf{S}_{11}^{-1}\mathbf{S}_{21}$ is the familiar *residual sum of squares error* matrix and obeys $W_{m-r}(N-r,\sigma^2\mathbf{I})$ distribution [15]. Since S_r is the trace of a matrix having $W_{m-r}(N-r,\sigma^2\mathbf{I})$ distribution, using the properties of Wishart distributions we conclude that $\frac{2}{\sigma^2}S_r$ obeys $\chi^2_{2(m-r)(N-r)}$. This is indeed a surprising result. It says that as long as SNR is above the threshold SNR the exact distribution of sum of squares of noise singular values does not depend upon the signal component. It depends only on signal rank, matrix dimensions and the noise level. It is a more exact result than the asymptotic result of $\chi^2_{2(m-r)N}$.[2] For short data records, the asymptotic approximation is poor and a test developed using the asymptotic distributions will lead to poor performance.

[2]For large N (in fact for large $\frac{N}{m}$) we can replace $\tilde{\mathbf{U}}_0\tilde{\mathbf{U}}_0^H$ with $\mathbf{U}_2\mathbf{U}_2^H$, the noise singular vectors of covariance matrix. On substituting we get $S_r = \|\mathbf{U}_2^H\hat{\mathbf{Y}}\|_F$ with $\frac{2}{\sigma^2}S_r$ having $\chi^2_{2(m-r)N}$ distribution, for a more precise argument, see [1]. For finite data records, $\tilde{\mathbf{U}}_0$ depends on realization of $\hat{\mathbf{Y}}$, and varies for each realization. Thus, the distribution of $\frac{2}{\sigma^2}S_r$ is no longer $\chi^2_{2(m-r)N}$.

V. Error Probabilities

We differentiate the probability of error in estimating the rank into two distinct probabilities. If the test gives rank greater than the true rank, we call this event *false alarm* and denote this probability as *probability of false alarm*. If the test gives rank less than the true rank, we call this event a *miss* and denote this probability as *probability of miss*. Since these two events are mutually exclusive, the probability of error $P_e = P_{\mathrm{miss}} + P_{\mathrm{FA}}$.

In the procedure we start testing from the smallest sum S_{m-1} and work our way upwards till for some k, we have $S_k \geq T_k$. When this occurs we stop and decide that the rank is $k+1$. Probability of false alarm is probability of deciding rank to be any one from $r+1\cdots m$ when the true rank is r. In other words, the probability of false alarm is the probability that we exceed threshold in a test before the $(m-r)$th test and decide rank greater than r, the true rank. Let us denote event $S_k \geq T_k$ as E_k. Then the probability of the procedure stopping and deciding rank $k+1$ is $P\{E_k \cap_{j=k+1}^{m-1} E_j^c\}$. Thus, the P_{FA} is given by

$$P_{\mathrm{FA}} = P\left\{\left(\bigcup_{k=r}^{m-2} E_k \bigcap_{j=k+1}^{m-1} E_j^c\right)\bigcup E_{m-1} \mid H_r\right\}. \tag{19}$$

On the lines of earlier definitions, we define P_{miss}, the probability of miss as follows

$$P_{\mathrm{miss}} = P\left\{\bigcap_{j=r-1}^{m-1} E_j^c \mid H_r\right\}. \tag{20}$$

Usually it turns out that if we exceed threshold at the $(m-k)$th test (event E_k has occurred) and instead of stopping and deciding that the rank is $k+1$ we continue with subsequent tests, then in all subsequent tests we exceed the corresponding thresholds, i.e., $E_k \subset E_{k-1}$ for all k, regardless of the underlying true rank.[3] From this assumption, we have $\cap_{j=k+1}^{m-1} E_j^c = E_{k+1}^c$. Thus, P_{miss} simplifies to

$$P_{\mathrm{miss}} = P\{E_{r-1}^c \mid H_r\} = P(S_{r-1} < T_{r-1} \mid H_r). \tag{21}$$

Similarly, using $\cap_{j=k+1}^{m-1} E_j^c = E_{k+1}^c$, we have

$$P_{\mathrm{FA}} = P\{E_r \mid H_r\} = P\{S_r \geq T_r \mid H_r\}. \tag{22}$$

But $P(S_r \geq T_r \mid H_r)$ is nothing but α, the individual false alarm level which we had set earlier. Thus, regardless of the true rank r, this level α is indeed equal to the overall or the actual false alarm probability, or P_{FA} the probability of overestimating the rank. We can control P_{FA} simply by choosing $\alpha = P_{\mathrm{FA}}$, and setting thresholds accordingly, and by doing so we are controlling the probability of overestimation.

VI. Simulation Results

Here, we apply the proposed method to predict the model order for short data records. We consider the data matrix as 15 realizations of signal $\mathbf{y}_k$, a length 10 zero mean, gaussian random vector with a covariance matrix of rank 4 which is corrupted by additive zero mean, white gaussian noise with variance σ^2. The singular values of the covariance matrix $\mathbf{R}$ of (3) are $\lambda_1 = 36+\sigma^2$, $\lambda_2 = 16+\sigma^2$, $\lambda_3 = 9+\sigma^2$, $\lambda_4 = 1+\sigma^2$ and $\lambda_5 = \cdots = \lambda_{10} = \sigma^2$. We define SNR as $10\log_{10}(\frac{1}{\sigma^2})$.

[3]Since $S_{k-1} \geq S_k$, the assumption $E_k \subset E_{k-1}$ will be true if all $\{T_k\}_{k=0}^{m-1}$ are equal. In practice, the thresholds are close to each other and (21) and (22) are good approximations of the true performance over range of SNR's and threshold settings.

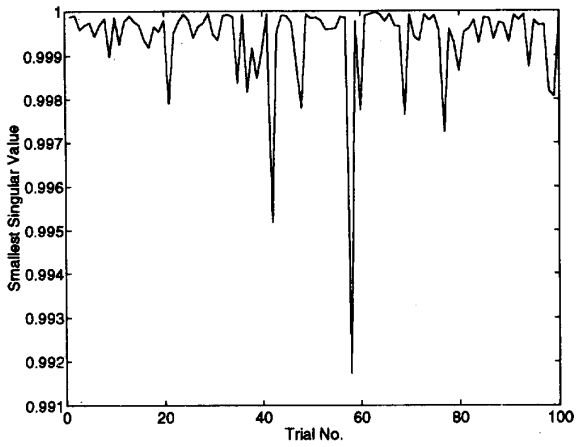

Fig. 1. Accuracy of perturbation approximation. cosine of the largest principal angle between the space spanned by the perturbed signal subspace $\hat{\mathbf{U}}_s\hat{\mathbf{U}}_s^H$ and the space spanned by the perturbation approximation.

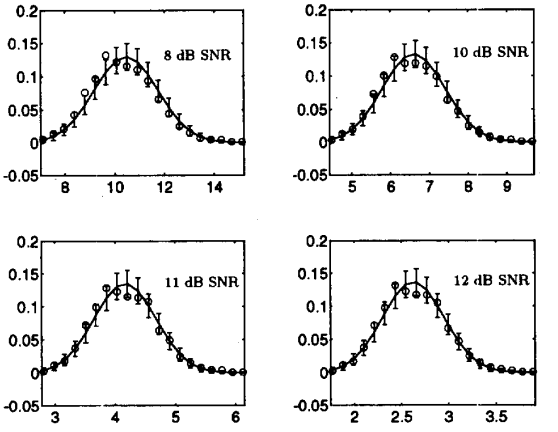

Fig. 2. Histogram of sum of squares of *noise* singular values $\sum_{i=5}^{10}\sigma_i^2$ at 8, 10, 12, 14 dB SNR. The number of bins is 20 and results are based on 2500 independent trials. The confidence intervals are plotted for 99.87% probability (3 standard deviations away from mean).

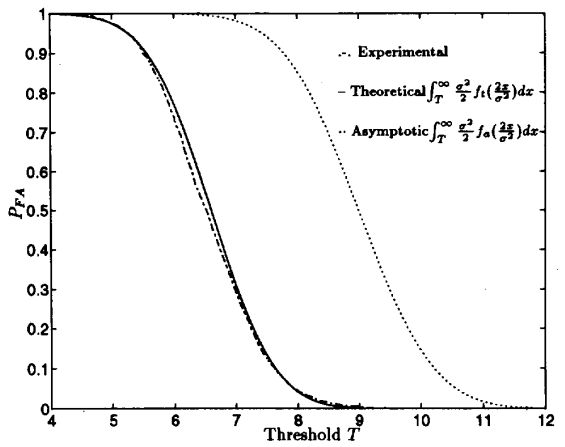

Fig. 3. Comparison of the predicted and experimental values of the overall probability of false alarm, i.e., the probability of overestimation. The results are based on 2500 independent trials and SNR = 10 dB. $f_t(x)$ and $f_a(x)$ are the probability density functions of $\chi^2_{2(m-r)(N-r)}$ and $\chi^2_{2(m-r)N}$ random variables, respectively.

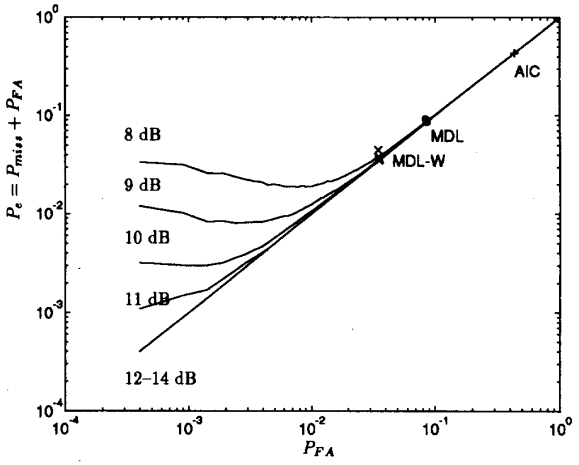

Fig. 4. Comparison of the proposed method with AIC, MDL, and MDL-W criterion. In this case, AIC ($P_e \sim 0.5$), MDL ($P_e \sim 0.1$), and MDL-W ($P_e \sim 0.04$) perform poorly and a more important factis that their performance does not improve much with increase in SNR.The graph of the proposed method is to be interpreted as follows: Given a false alarm rate on the abscissa, the ordinate of the graph gives the resulting probability of error for that particular SNR. As long as the false alarm rate is higher than the minima of the graph, the user can lower the false alarm rate and achieve a better performance.

In order to verify accuracy of the approximation of the perturbed orthogonal subspace given by (13), in Fig. 1 we plot the cosine of the largest principal angle between the approximate and the true subspace, see [16]. In Fig. 2, Histograms of S_4, the sum of squares of the last six singular values are depicted for different SNR's.

In Fig. 3 we compare the actual probability of false alarm, i.e., the probability of overestimation against the theoretical predicted value obtained from the tail probability of $\chi^2_{2(m-r)(N-r)}$ for 10 dB SNR.

For comparing the performance of the proposed method with AIC, MDL, and other methods we plot the error probability of the proposed method against false alarm rate. Note that for very high false alarm levels (very low threshold values), the main contribution to P_e is from error due to overestimation ($P_e \approx P_{\mathrm{FA}}$). The plot of P_e vs P_{FA} is linear and approximately follows the line $P_e = P_{\mathrm{FA}}$. As we lower the false alarm level (increase the threshold values), the contribution to P_e from error due to underestimation increases and the plot deviates upwards from the diagonal line ($P_e = P_{\mathrm{FA}}$). Here, $P_e = P_{\mathrm{FA}} + P_{\mathrm{miss}}$. For very low false alarm levels (very high threshold values) most of the error is due to underestimation and $P_e \approx P_{\mathrm{miss}}$. Thus, the graph of $P_e vs P_{\mathrm{FA}}$ starts at $(P_{\mathrm{FA}}, P_e) = (1,1)$, follows line $P_e = P_{\mathrm{FA}}$ deviating slightly upwards, reaches a certain minimum and then moves upwards with negative slope and ends at $(P_{\mathrm{FA}}, P_e) = (0,1)$, see Fig. 4. The graph of proposed method is to be interpreted as follows: Given a false alarm rate on the abscissa, the ordinate of the graph gives the resulting probability of error for that particular SNR. As long as the false alarm rate is higher than the abscissa of the minima, the user can lower the false alarm rate and achieve a better performance. AIC and MDL are minimization criteria and thus give a single probability of error and false alarm, and user doesn't have any control over the false alarm rate, they will be represented by a single

symbol for each SNR. Since the probability of error of AIC, MDL and MDL-W does not improve with increase in SNR, in the figure the symbols representing the performance overlap considerably.[4]

For 10 000 trials, SNR's above 11 dB S_4 and S_5 are well-separated and there is no occurrence of an underestimation (a miss), and $P_e = P_{FA}$. As the SNR falls below 11 dB, a miss event occurs more frequently, and we can clearly see the deviation from the line $P_e = P_{FA}$. As per the application, the user can choose an appropriate value of false alarm and from the graph can obtain the corresponding value of probability of error for the proposed method. In this example, the performance of the proposed method is superior to the performance of AIC, MDL, and MDL-W [9] over a wide range of false alarm levels. The performance of AIC, MDL and MDL-W, does not improve much as the SNR is increased. But in the proposed method for high SNR cases very low false alarm levels (orders of magnitudes lower than other criterion) can be chosen and still obtain a very low probability of error. Hence, a prudent user armed with probability of error expressions, and the knowledge of minimum SNR level, can design or modify or customize his system to obtain an significant improvement in performance.

REFERENCES

[1] T. W. Anderson, "Asymptotic theory for principal component analysis," *Ann. Math. Statist.*, vol. 34, pp. 122–148, 1963.

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

[3] W. S. Liggett, "Passive sonar: Finding models to multiple time series," in *Signal Processing*, J. W. Griffiths, Ed. New York: Academic, 1980, pp. 135–169.

[4] D. B. Williams and D. H. Johnson, "Using the sphericity test for source detection with narrow-band passive arrays," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 2008–2014, Nov. 1990.

[5] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Auto. Contr.*, vol. AC-19, pp. 716–723, Dec. 1974.

[6] J. Rissanen, "Modeling by the shortest data description," *Automatica*, pp. 465–471, 1978.

[7] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 387–392, Apr. 1985.

[8] Y. Q. Yin and P. R. Krishnaiah, "On some nonparametric methods for detection of the number of signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1533–1538, Nov. 1987.

[9] D. B. Williams, "Comparison of AIC and MDL to the minimum probability of error criterion," in *Proc. Sixth SSAP Workshop in Statistical Signal & Array Processing* (Victoria, Canada), Oct. 1992, pp. 114–117.

[10] M. Wax and I. Ziskind, "Detection of coherant signals by the MDL principle," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-37, pp. 1190–1196, 1985.

[11] Q. Wu and D. R. Fuhrmann, "A parametric method for determining the number of signals in narrow-band direction finding," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 39, pp. 1848–1857, Aug. 1991.

[12] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Identification*, Reading, MA: Addison-Wesley, 1991.

[13] F. Li, R. J. Vaccaro, and D. W. Tufts, "Performance analysis of the state space realization (TAM) and ESPRIT algorithms for DOA estimation," *IEEE Trans. Antenn. Propagat.*, vol. 39, pp. 418–423, Mar. 1991.

[14] 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*, R. J. Vaccaro, Ed. Amsterdam: Elsevier, 1991.

[15] A. M. Kshirsagar, *Multivariate Analysis*. New York: Marcel Dekker, 1972.

[16] G. Golub and C. V. Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: Johns Hopkins University Press, 1989.

[4] The criteria used in simulations do not assume the knowledge of noise level. AIC and MDL criteria can be modified for the case of known noise level.

Performance Analysis of the UCA-ESPRIT Algorithm for Circular Ring Arrays

Cherian P. Mathews and Michael D. Zoltowski

***Abstract*— Statistical performance analysis of the recently developed UCA (uniform circular array)-ESPRIT algorithm is the subject of this correspondence. The UCA-ESPRIT arrival angle estimates are shown to be asymptotically unbiased, and expressions for the asymptotic estimator variances have been obtained. Simulation results that verify the analysis are presented.**

I. INTRODUCTION

The recently developed UCA-ESPRIT algorithm [1] for 2-D angle estimation with uniform circular arrays (UCA's) provides *automatically paired* source azimuth and elevation angle estimates via a *closed-form* procedure. UCA-ESPRIT represents a significant advance in the area of 2-D angle estimation as it does not require the pairing procedures for associating azimuth and elevation estimates, 2-D spectral searches, or iterative optimization techniques characteristic of other 2-D angle estimation algorithms. UCA-ESPRIT is similar to ESPRIT [2] in that the eigenvalues of a matrix (which is the least squares solution to an overdetermined system of equations) directly yield the DOA estimates. The eigenvalues have the form $\mu_i = \sin\theta_i e^{j\phi_i} = u_i + jv_i$, where θ_i and ϕ_i are respectively the elevation and azimuth of the ith source, and u_i and v_i are the direction cosines with respect to the x and y axes, respectively.

The asymptotic performance of UCA-ESPRIT is studied in this paper. The analysis, along the lines of the work of Rao and Hari [3] reveals that the UCA-ESPRIT arrival angle estimators are asymptotically unbiased. Asymptotic expressions for the variances of the DOA estimators are obtained. Results of computer simulations that verify the analysis are presented.

The organization of this correspondence is as follows. The UCA-ESPRIT algorithm employs a phase mode excitation based beamformer, and Section II therefore reviews phase mode excitation based beamforming for UCA's. The UCA-ESPRIT algorithm is described in Section III. The asymptotic performance analysis of UCA-ESPRIT is considered in Section IV. Finally, the results of computer simulations that validate the theoretical performance expressions are presented in Section V.

II. PHASE MODE EXCITATION BASED UCA BEAMFORMERS

The UCA geometry is depicted in Fig. 1. The antenna elements, assumed to be identical and omnidirectional, are uniformly distributed over the circumference of a circle of radius r in the xy plane. A spherical coordinate system is used to represent the arrival directions of the incoming plane waves. The origin of the coordinate system is located at the center of the array. Source elevation angles θ are measured down from the z axis, and azimuth angles ϕ are measured counterclockwise from the x axis. Source DOA's can

Manuscript received August 28, 1993; revised November 30, 1993. This work was supported by AFOSR under contract no. F49620-92-J-0198 in conjunction with Wright Laboratories and the National Science Foundation under contract no. MIP-9320890. The associate editor coordinating the review of this paper and approving it for publication was Dr. R. D. Preuss.

The authors are with the School of Electrical Engineering, Purdue University, West Lafayette, IN 47906 USA.

IEEE Log Number 9403269.