A Data-Based Enumeration Technique for Fully Correlated Signals*

H. Krim and J. G. Proakis

Communications and Digital Signal Processing Center

Northeastern University, Boston, MA 02115

Edics 5.2.1, 5.2.2 or 5.2.3

Permission to publish this abstract separately is granted

Abstract

We present a novel method for estimating the number of signals impinging on a uniform linear array using observed sensor data. Unlike other algorithms which apply Rissanen's Minimum Description Length (MDL) principle to the observed data for source enumeration, this method applies it to the *prediction errors* of a linear model which has been fitted to an appropriate data matrix. It is a one-dimensional method which achieves improved performance even for fully correlated signals over contemporary approaches, particularly with short data records and closely spaced signals. Asymptotic consistency is shown and substantiating simulation examples are included.

*This work was supported in part by NSF grant IJR 9102148.

1 Introduction

Parameter estimation techniques, particularly eigenstructure-based techniques, have been the focus of increased research activity because of their wide-ranging applications in a number of fields. Many of them, however, rely on prior knowledge of the number of signals whose parameters are of interest.

Determining the actual number of signals impinging on an array of sensors - the detection problem - is generally a critical first step in solving the source localization problem - estimating their Direction(s)-Of-Arrival (DOA). Two information theoretic criteria for order estimation of an observed process, namely the Akaike Information Criterion (AIC) [1] and Rissanen's Minimum Description Length (MDL) principle [2], have inspired many solutions to the aforementioned problem [11, 14, 9, 15, 16, 7, 5, 6, 17, 8]. Many of the algorithms were derived under a "long data record" assumption and when simplified, include a test of equality of the smallest eigenvalues of the array covariance matrix for an observed process. Most of the forementioned methods fail in (fully) correlated or coherent signal environments, as some of the eigenvalues which would normally correspond to the eigenvectors spanning the signal subspace get confounded with those which correspond to the noise subspace.[1] The MDL-based algorithm as first proposed by Wax and Kailath [5], suffered from this shortcoming. This deficiency was later overcome by Wax and Ziskind in [6] and by Wax in [7]. Their solutions were based on a multidimensional search of a performance surface, which for a large number of degrees of freedom can become computationally intensive. Moreover, the search may not necessarily converge to an absolute extremum even in a moderately noisy environment.

[1] The signal and the noise subspaces are respectively spanned by the eigenvectors of the array covariance matrix corresponding to the large and the small eigenvalues.

When the array is uniform and linear, a computationally more attractive solution consists of first recovering the rank through a smoothing transformation applied to the array covariance matrix, and then applying a Smoothed Rank Profile (SRP)[2] test [9, 8, 14]. This entails tracking the increase and eventual stabilization of the rank as the rank of the smoothing matrix varies. The point at which the rank stabilizes with probability 1 (w.p.1) [14], corresponds to the number of signals present in the observed process . In the presence of closely spaced coherent signals, however, this techniques exhibits a performance threshold which is due in part to the residual correlation among the averaged diagonal submatrices required by the smoothing transformation. As we later explain in section 3, this "vestigial" correlation, if unaccounted for, can further degrade the performance of even those algorithms which are applicable in a coherent signal environement.

The purpose of this paper is to derive a new, data-domain based signals (possibly fully correlated) enumeration method which ameliorates the above shortcoming(s) when the signals are assumed to impinge on a uniform linear array.[3] This method applies the MDL algorithm to the *prediction errors* of a linear model which has been fitted to an appropriate data matrix.[4] This is in contrast to the solutions in [9, 8, 14] which appply the MDL to encode the data and ultimately include an equality test of eigenvalues of the smoothed array covariance matrix. Simulation results which show improved performance over existing similar approaches are included. This improvement is particularly demon-

[2]The ranks of a sequence of smoothed covariance matrices is referred to as a smoothed rank profile.

[3]A similar scalar problem was independently treated in [11].

[4]**Added in Proof:** As pointed out by one of the reviewers, there is an alternative method due to Rissanen [3, 4], called *stochastic complexity* which is also based on coding the prediction errors. It has been shown to be optimal with respect to minimizing the codelength of the encoded observations, but its computational complexity exceeds that of MDL.

strated with short data records and closely spaced signals at a moderate computational cost of two Singular Value Decompositions (SVD)[5]. We also show that the technique is asymptotically consistent.

1.1 Propagation Model

In the following analysis, we shall consider a uniform linear array composed of L identical, equally spaced, omnidirectional sensors with sensor spacing $d = \frac{\lambda}{2}$ where λ is the signal wavelength. We assume that $M < L$ narrowband planewaves (centered about the known frequency ω_0) impinge on the array from (distinct) directions $\theta_1, \theta_2, \ldots, \theta_M$. The (complex envelopes of the) signals received by this array of sensors can then be expressed as

$$x_i(t) = \sum_{k=1}^{M} a_i(\theta_k) s_k(t) + n_i(t), \; i = 1, \ldots, L, \tag{1}$$

where $x_i(t)$ is the output of the i^{th} sensor, $\mathbf{a}(\theta_k) = [1, e^{j\omega_0\tau_k}, \ldots, e^{j\omega_0(L-1)\tau_k}]^{\mathrm{T}}$, $\tau_k = (d/c)\sin\theta_k$, is the $L \times 1$ *Direction Of Arrival (DOA)* or *steering vector* of the k^{th} signal, $s_k(t)$ is the k^{th} deterministic signal as received at the reference point, sensor 1, and $n_i(t)$ is the noise at the i^{th} sensor. The exponent 'T' in the above expression for $\mathbf{a}(\theta_k)$ denotes transposition, and τ_k represents the propagation delay between two successive sensors for a *planar wavefront* impinging on the array from the direction θ_k with propagation speed c. We assume that $M - (M' - 1)$ signals are fully correlated. The effective rank[6] of the covariance matrix $\mathbf{R} = \mathrm{E}\{\mathbf{x}^H\mathbf{x}\}$, where $\mathbf{x} = [x_1, x_2, \cdots, x_L]$ is then M'.

[5] The computations may further be optimized through recursion.

[6] Throughout the remainder of this paper, by the *(effective) rank* of $\mathbf{R}$, we mean the number of "large" eigenvalues (or singular values) which is the rank of the noise-free matrix $\mathbf{R}$. When the signals are not fully correlated, the effective rank corresponds to the number of signals present in the observed process. By abuse of terminology, we will often drop the adjective 'effective'.

We shall assume that the array outputs and the noise are stationary and ergodic, complex-valued normal random processes having zero mean, the noise is uncorrelated with the signals, and the noise terms are mutually uncorrelated with *unknown* but identical variances, σ^2.

The signal enumeration problem is to estimate the number of signals impinging on the above array making a direct use of the data and avoiding to compute the second or higher order statistics. Of greater interest to us is the case where the data record is short, i.e., has length $N < 2L$.

2 A New Data-Domain Detection Algorithm

2.1 Smoothing Transformations

It is well known that signal correlation induces a rank deficiency in the matrix

$$\mathbf{X} = \begin{pmatrix} x_L(1) & x_{L-1}(1) & \cdots & x_1(1) \\ x_L(2) & x_{L-1}(2) & \cdots & x_1(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_L(N) & x_{L-1}(N) & \cdots & x_1(N) \end{pmatrix}, \tag{2}$$

where the i^{th} column vector of $\mathbf{X}$ corresponds to a time record of the signal recorded at the $(L-i+1)^{th}$ sensor, and that it is possible to "unfold" the collapsed column space by applying an appropriate smoothing transformation to (2) (e.g., see [13]). Specifically, if p is a positive integer smaller than L, $n_s = L - p + 1$, and $\mathbf{F} = [\mathbf{F}_{1p}|\mathbf{F}_{2p}|\cdots|\mathbf{F}_{n_s p}]$, is the $p \times n_s L$ windowing matrix defined by $\mathbf{F}_{ip} = [\mathbf{0}_{p\times(i-1)}|\mathbf{I}_p|\mathbf{0}_{p\times(L-i-p+1)}]$, $i = 1, 2, \cdots, n_s$,[7] then

[7] When $i = 1$ (n_s), the first (last) zero matrix in $\mathbf{F}_{ip}$ is not present.

the smoothing transformation induced by (successive) p-element overlapping subarrays[8] is defined to be $[\mathbf{I}_{n_s} \otimes \mathbf{X}]\mathbf{F}^{\mathrm{T}}$ where $\otimes$ denotes the Kronecker product, $\mathbf{I}_{n_s}$ is the $n_s \times n_s$ identity matrix and n_s corresponds to the number of subarrays.

It is easy to see that

$$\mathbf{X}(p) = [\mathbf{I}_{n_s} \otimes \mathbf{X}]\mathbf{F}^{\mathrm{T}}, \tag{3}$$

$$= \begin{pmatrix} \mathbf{X}(p,1) \\ \mathbf{X}(p,2) \\ \vdots \\ \mathbf{X}(p,n_s) \end{pmatrix}, \tag{4}$$

where

$$\mathbf{X}(p,i) = \begin{pmatrix} x_{p+i-1}(1) & x_{p+i-2}(1) & \cdots & x_i(1) \\ x_{p+i-1}(2) & x_{p+i-2}(2) & \cdots & x_i(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_{p+i-1}(N) & x_{p+i-2}(N) & \cdots & x_i(N) \end{pmatrix}, \; i = 1, 2, \cdots, n_s, \tag{5}$$

is the data matrix determined by the i^{th}, p-element subarray comprised of sensors $i, i+1, \cdots, p+i-1$. Note that the above data matrix is related to the spatially smoothed array covariance matrix (determined by p-element subarrays), $\mathbf{R}(p)$, via the formula $\mathbf{R}(p) = \frac{1}{Nn_s}\mathbf{X}(p)^{\mathrm{H}}\mathbf{X}(p)$. Since rank $\mathbf{R}(p)$=rank $\mathbf{X}(p)$, we can adapt the methods described in [14] or [9] to unravel the signal structure from the observed process; equivalently, determine M. To this end, we need an effective way to compute rank $\mathbf{X}(p)$. One way to accomplish this is to fit a linear model to the rank-enhanced matrix $\mathbf{X}(p)$, and then use the fact

[8] These are comprised of the p successive sensors $i, i+1, \cdots, p+i-1$.

that its order equals rank $\mathbf{X}(p)$. The remainder of this paper will show that this can be done very efficiently if we apply the MDL principle to the *prediction errors* instead of the entries of $\mathbf{X}(p)$.

2.2 Computing the Density Function of the Model

Thus, let

$$\mathbf{X}(p)'\mathbf{d} + \mathbf{e} = \mathbf{x}_p, \tag{6}$$

where $\mathbf{X}(p) = [\mathbf{x}_p, \mathbf{X}(p)']$, $\mathbf{x}_p$, the first column of $\mathbf{X}(p)$, $\mathbf{d}$ is the $(p-1)\times 1$ linear prediction vector, and $\mathbf{e} = [\mathbf{e}(p), \mathbf{e}(p+1), \cdots, \mathbf{e}(L)]^{\mathrm{T}}$ is the $n_s N \times 1$ prediction error vector. For $j = 0, \cdots, L-p$, the $N \times 1$ subsvector $\mathbf{e}(p+j)$ of $\mathbf{e}$ is induced by the $N \times p$ submatrix $\mathbf{X}(p, j+1)$. Without loss of generality, we shall assume that $n_s N$ is large, so that the errors (the individual components of $\mathbf{e}$) can be assumed to be normally distributed with covariance matrix $\mathbf{\Gamma} = \mathrm{E}[\mathbf{e}\mathbf{e}^{\mathrm{H}}]$ having minimal "end effects" when using the conditional density [18]. The components of an error N-dimensional subvector (e.g., $\mathbf{e}(p)$) are clearly uncorrelated since the time samples are assumed independent. There exists, however, correlation between the error subvectors since successive subblocks of data - elements of $\mathbf{X}(p)'$ - which are induced by the partitioning of $\mathbf{e}$, have $p-2$ columns in common as can be seen from Eqs. (5) and (6). In order to apply the MDL principle, we need to compute the probability density function of the (fitted) model (e.g., see [6]).

To simplify the computation of this density, we first whiten the model error. This, of course, requires explicit knowledge of $\mathbf{\Gamma}$. We accomplish this by first rewriting (6) as

$$\mathbf{X}(p)\mathbf{d}' = \mathbf{e},$$

or

$$\mathbf{d}'^{\mathrm{T}}\mathbf{X}(p)^{\mathrm{T}} = \mathbf{e}^{\mathrm{T}}, \tag{7}$$

where $\mathbf{d}' = [1, -\mathbf{d}^{\mathrm{T}}]^{\mathrm{T}}$.

Recall that if $\mathbf{Y}$ is an $m \times n$ matrix, $\mathrm{Vec}(\mathbf{Y})$ is defined to be the $mn \times 1$ (row) vector $(y_{11}, \cdots, y_{m1}, \cdots, y_{1n}, \cdots, y_{mn})^T$. Applying the Vec operator to both sides of (7), we obtain

$$\mathrm{Vec}(\mathbf{d}'^{\mathrm{T}}\mathbf{X}(p)^{\mathrm{T}}) \;=\; \mathrm{Vec}(\mathbf{e}^{\mathrm{T}}) = \mathbf{e},$$

or,

$$(\mathbf{I} \otimes \mathbf{d}'^{\mathrm{T}})\,\mathrm{Vec}(\mathbf{X}(p)^{\mathrm{T}}) \;=\; \mathbf{e}. \tag{8}$$

Each data point has a signal component contaminated with noise, namely,

$$x_i(j) \;=\; x_{sci}(j) \;+\; n_i(j),\; j = 1, \ldots, N, \tag{9}$$

where the subscript sc denotes "the signal component." Denoting the noise components of $\mathbf{X}(p)$ by $\mathbf{N}(p)$, we see that $\mathrm{Vec}(\mathbf{X}(p)^{\mathrm{T}}) \;=\; \mathrm{Vec}(\mathbf{X}(p)^{\mathrm{T}}_{sc}) \;+\; \mathrm{Vec}(\mathbf{N}(p)^{\mathrm{T}})$, from which

$$\mathbf{e} \;\approx\; 0 + (\mathbf{I} \otimes \mathbf{d}'^{\mathrm{T}})\,\mathrm{Vec}(\mathbf{N}(p)^{\mathrm{T}}), \tag{10}$$

since the signal components (cisoids) theoretically fit the linear model perfectly. Using this expression for $\mathbf{e}$, we can now express $\mathbf{\Gamma}$ as

$$\begin{aligned} \mathbf{\Gamma} \;&=\; (\mathbf{I} \otimes \mathbf{d}'^{\mathrm{T}})\,\mathrm{E}[\mathrm{Vec}(\mathbf{N}(p)^{\mathrm{T}})\mathrm{Vec}(\mathbf{N}(p)^{\mathrm{T}})^{\mathrm{H}}](\mathbf{I} \otimes \mathbf{d}'^{\mathrm{T}})^{\mathrm{H}} \\ &=\; \sigma^2(\mathbf{I} \otimes \mathbf{d}'^{\mathrm{T}})\,\widetilde{\mathbf{I}}(p)(\mathbf{I} \otimes \mathbf{d}'^{\mathrm{T}})^{\mathrm{H}} \;=\; \sigma^2\widetilde{\mathbf{\Gamma}}, \end{aligned} \tag{11}$$

where $\widetilde{\mathbf{I}}(p)$ is an $n_sNP \times n_sNP$ indicator matrix[9] with 1 and 0 components which result from the covariances of the noise components-e.g. $\widetilde{\mathbf{I}}(1,1) = \widetilde{\mathbf{I}}(1, Np+2) = 1$-.

Premultiplying (7) by $\widetilde{\mathbf{\Gamma}}^{-\frac{1}{2}}$ - *whitening* the term $\mathbf{e}$ - we obtain,

$$\widetilde{\mathbf{\Gamma}}^{-\frac{1}{2}}\mathbf{X}(p)'\mathbf{d} \;+\; \widetilde{\mathbf{\Gamma}}^{-\frac{1}{2}}\mathbf{e} \;=\; \widetilde{\mathbf{\Gamma}}^{-\frac{1}{2}}\mathbf{x}_p \tag{12}$$

[9]The matrix size depends on N and p and is, $\widetilde{\mathbf{I}}(p) \;=\; \mathbf{To}[\mathbf{I}_{Np}, \mathbf{S}\mathbf{I}_{Np}, \mathbf{S}^2\mathbf{I}_{Np} \cdots]$ where $\mathbf{S}$ is a right shifting matrix and $\mathbf{To}$ is the block Toeplitz operator.

or,

$$\mathbf{Y}(p)\mathbf{d} + \varepsilon = \mathbf{y}_p, \tag{13}$$

where $\mathrm{E}[\varepsilon\varepsilon^{\mathrm{H}}] = \sigma^2\mathbf{I}$.

Using the fact that ε is Gaussian, it follows that the conditional probability density function of $\mathbf{y}_p$ is given by

$$f(\mathbf{y}_p \mid \beta) = \frac{1}{(\pi\sigma^2)^{n_s N}} \exp\left(-\frac{(\mathbf{y}_p - \mathbf{Y}(p)\mathbf{d})^{\mathrm{H}}(\mathbf{y}_p - \mathbf{Y}(p)\mathbf{d})}{\sigma^2}\right), \tag{14}$$

where $\beta = [\mathbf{d}^{\mathrm{T}}, \sigma^2]$.

2.3 The MDL-Based Signal Enumeration Criterion

Fitting an optimal linear model to the smoothed observed data is one way of succintly describing (or "best summarizing") the main ("signal") components. The optimality is a result of accounting for the correlation among the data vectors of the various subarrays. Encoding the resulting error is then equivalent to encoding this "best" representation we have of the data. It is thus natural to expect an improved performance when an "optimal" preprocessing of the data is performed, followed by an optimal encoding. The latter optimality is afforded by the MDL principle [2] which intuitively states that the shortest code length which describes all of the relevant data is given by the optimal code. This is equivalent to saying that of all models which can describe an observed data, the most parsimonious one is optimal.

The application of this principle to the *prediction error* can be expressed as

$$L(\epsilon) = \mathcal{L}(\beta) + \frac{1}{2}\nu \log(n_s N), \tag{15}$$

where

$$\mathcal{L}(\beta) \quad = \quad -\log f(\mathbf{y}_p \mid \beta), \tag{16}$$

ν is the number of free parameters and $n_s N$ is the length of the vector. Minimizing the length $L(\epsilon)$ over the number of free parameters results in an MDL-based signal enumeration test. We refer to this test as Optimized Data Domain DETection test (ODDDET). We should also point out that a similar test, referred to as DDDET can also be based on $L(\mathbf{e})$. It can be derived by assuming that the error covariance matrix $\mathbf{\Gamma}$ is diagonal -i.e. the smoothed data vectors are uncorrelated[10]-, avoiding the whitening transformation in Eq. (12) and using $f(\mathbf{x}(p) \mid \beta)$ for the conditional probability density function. The minimization of Eq. (15) entails the derivation of the Maximum Likelihood (ML) estimate of the parameter vector β, $\widehat{\beta}$ in Eq. (16). This can be achieved by maximizing the log-likelihood function with respect to (w.r.t.) $\mathbf{d}^{\mathrm{T}}$ and σ^2. Differentiating $\mathcal{L}([\mathbf{d}^{\mathrm{T}}, \sigma^2])$ w.r.t. $\mathbf{d}^{\mathrm{T}}$ and σ^2, and solving the equations obtained by setting these derivatives equal to zero, produces $\widehat{\beta}$:

$$\begin{aligned} \widehat{\sigma^2} \quad &= \quad \frac{1}{\alpha} \parallel \mathbf{y}_p \, - \, \mathbf{Y}(p)\mathbf{d} \parallel^2, \ \text{and} \\ \widehat{\mathbf{d}} \quad &= \quad \mathbf{Y}(p)^{\#}\mathbf{y}_p, \end{aligned} \tag{17}$$

where $\alpha = n_s \times N$ and '$\#$' is a symbol used for the pseudo-inverse. Substituting this result back into (16) gives

$$\begin{aligned} \mathcal{L}(\widehat{\beta}) \quad = \quad & \alpha \log \pi \, + \, \alpha \log \left(\frac{1}{\alpha} \parallel \mathbf{y}_p \, - \, \mathbf{Y}(p)\mathbf{Y}(p)^{\#}\mathbf{y}_p \parallel^2\right) \\ & + \alpha. \end{aligned} \tag{18}$$

[10] This is implicitly used in [9].

Next, for $1 \leq k \leq L$, we replace in the normed quantity in (18)$\mathbf{Y}(p)$ [21, 19] with its rank-k approximation,[11] thus obtaining a parameterized version of $\mathcal{L}(\widehat{\beta})$, $\mathcal{L}(\widehat{\beta}, k)$. Recall that if $\mathbf{Y}(p) = \mathbf{V}\mathbf{\Sigma}\mathbf{U}^{\mathrm{H}}$, is the SVD for $\mathbf{Y}(p)$, the *rank-k approximation* to $\mathbf{Y}(p)$, $\mathbf{Y}(p)_k$, is given by $\mathbf{Y}(p)_k = \mathbf{V}\begin{bmatrix} \mathbf{\Sigma}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{U}^{\mathrm{H}}$, where the vectors $\mathbf{V}$ and $\mathbf{U}$ are, respectively, the left and the right singular vectors of $\mathbf{Y}(p)$, and $\mathbf{\Sigma}_k$ is the $(k \times k)$ diagonal matrix comprised of the largest singular values. Substituting $\mathbf{Y}(p)_k$ for $\mathbf{Y}(p)$ in the argument of the second term of (18) gives the parameterized version of this expression

$$\varepsilon^k = \left\{\mathbf{y}_p - \mathbf{Y}(p)_k \mathbf{Y}(p)_k^{\#}\mathbf{y}_p\right\} = \mathbf{y}_p - \mathbf{V}\begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{V}^{\mathrm{H}}\mathbf{y}_p.^{12} \tag{19}$$

Since the norm of ε is invariant under multiplication by the unitary matrix $\mathbf{V}^{\mathrm{H}}$ [19], the square of the norm of preceeding expression is equivalent to [11]

$$\| \varepsilon^k \|^2 = \| \mathbf{V}^{\mathrm{H}}\mathbf{y}_p - \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{V}^{\mathrm{H}}\mathbf{y}_p \|^2 = \sum_{i=k+1}^{n_s N} \| \xi_i \|^2, \tag{20}$$

where $\xi = \mathbf{V}^{\mathrm{H}}\mathbf{y}_p$ and the superscript on ε^k represents the (assumed) model order (equivalently, effective rank).

For a model order k - equivalently, that k signals are impinging on the array - the optimal linear prediction vector $\mathbf{d}$ is obtained by minimizing $\| \varepsilon^k \|^2$. Since only k entries of the vector ε are used in the minimization, and, given that complex signals are considered herein, the number of adjustable parameters are the $2k$ entries of ε^k.

[11] We set the smallest singular values by zero and result with the closest matrix in norm to the original matrix.

[12] Note that $\begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ is an $n_s N \times n_s N$ matrix.

By substituting the previous expression into Eq. (18) for $\mathcal{L}(\widehat{\beta}, k)$, and attaching the appropriate penalty function to this expression, it is now possible to write down an MDL-based test which, when minimized over all possible k, provides the model order for the chosen subarray size p:

$$\begin{aligned} \widehat{k} &= \arg\left\{ \min_{k \in \{1,\cdots,p-1\}} \mathrm{MDL}(k) \right\}, \text{ where} \\ \mathrm{MDL}(k) &= \left\{ \alpha + \alpha \log \pi + \alpha \log \{ (\frac{1}{\alpha}) \parallel \varepsilon^k \parallel^2 \} \right. \\ &\quad \left. + (k) \log (n_s N) \right\}, \end{aligned} \tag{21}$$

where α, as previously defined, represents the length of the vector $\mathbf{x}_p$ used in estimating the linear prediction model, and the constant independent of k was left out.

The final step is to repeat this process, successively decreasing p from L to 2 (if necessary), obtaining each time, an order $\widehat{k}$ (one which minimizes the MDL(k)). If the number of degrees of freedom is sufficient, the resulting sequence of model orders will stabilize w.p.1 [14]. The value of $\widehat{k}$ at which stabilization takes place, is shown in the next section to be a consistent estimate of the number of signals.

The resulting algorithm is summarized below:

Summary of Algorithm

1. For a subarray size p (usually chosen to be $(L-1)$, if one assumes there are at least two coherent signals, else choose L), obtain an estimate of the linear prediction vector $\mathbf{d}$ as in equation (6);

2. Obtain an estimate of the error covariance matrix $\mathbf{\Gamma}$ as in (11)

3. Whiten the noise via the inverse hermitian square root of $\widetilde{\mathbf{\Gamma}}$ as described in equation (13);

4. Use the norm of ξ for different model orders as they are tested to minimize the MDL in equation (21) to obtain a model order estimate (or effective rank estimate of $\mathbf{X}(p)$) $\widehat{k}$;

5. Increase (or decrease) p and repeat steps 2-4;

6. If $\widehat{k}$ stabilizes for two or more consecutive subarray sizes (as p varies), it should be picked as the number of signals; if the rank does not stabilize the number of degrees of freedom is not sufficient to obtain a solution (and there is no solution).

2.4 Consistency of the Detection Scheme

An important characteristic of a detection procedure lies in its ability to provide an unbiased estimate of the model order (or number of signals), as the length of the data

vector used for prediction grows without bound (or is very large for practical purposes).

Lemma 1 *The order estimate $\widehat{k}$ obtained via the above algorithm is a consistent estimate of the number of signals incident on a uniform linear array, i.e.,*

$$\arg\{\lim_{\alpha\to\infty} \mathrm{MDL}(\widehat{k})\} \;=\; M \quad w.p.1. \tag{22}$$

Proof: By the criterion developed by Cozzens, *et al.* [14], as p varies, the (effective) ranks of the $\mathbf{X}(p)$ stabilize w.p.1 to the number of signals if the number of degrees of freedom (number of sensors) is sufficient. Therefore, it suffices to prove that the effective rank of $\mathbf{X}(p)$ *determined with the above procedure* is asymptotically consistent; equivalently, that the length $\mathrm{MDL}(k)$ is always greater than the length given by the optimal or true order M (or $\mathrm{MDL}(M)$). This is equivalent to showing that M is the minimum value of the curve described by the MDL function.

In order to proceed, the following result is needed [10]: if

$$\lim_{N\to\infty} C_N/N \to 0, \;\; \text{and} \;\; \lim_{N\to\infty} C_N \to \infty, \tag{23}$$

then,

$$\begin{aligned} \log \frac{\parallel \varepsilon^k \parallel}{\parallel \varepsilon^M \parallel} \;&\geq\; -(k-M)C_N/N, \text{when } k > M \;; \\ \log \frac{\parallel \varepsilon^k \parallel}{\parallel \varepsilon^M \parallel} \;&>\; (M-k)C_N/N, \text{when } k < M. \end{aligned} \tag{24}$$

Clearly, $\log\alpha$ satisfies the properties of C_N given in equation (23). If we evaluate (21) at k and M, the following difference can be deduced:

$$\frac{1}{\alpha}(\mathrm{MDL}(k) \;-\; \mathrm{MDL}(M)) \;=\; \frac{1}{\alpha}\log(\alpha)^{(k-M)} \;+\; \log\frac{\parallel \varepsilon^k \parallel}{\parallel \varepsilon^M \parallel}. \tag{25}$$

There are two distinct cases that one must consider in order to study the behavior of the expression given in equation (25).

Case 1: $k > M$: As $\alpha \to \infty$,

$$\log \frac{\| \varepsilon^k \|}{\| \varepsilon^M \|} \geq -(k-M)\frac{\log \alpha}{\alpha}. \tag{26}$$

by the first relation in (24). Since $k > M$, $\| \varepsilon^k \| < \| \varepsilon^M \|$ (recall that $\alpha - k < \alpha - M$), implying that

$$\log \frac{\| \varepsilon^k \|}{\| \varepsilon^M \|} < 0,$$

and hence,

$$(k-M)\frac{\log \alpha}{\alpha} > 0 > \log \frac{\| \varepsilon^k \|}{\| \varepsilon^M \|} \text{ as } \alpha \to \infty \tag{27}$$

since

$$\lim_{\alpha \to \infty} \frac{\log \alpha}{\alpha} = 0^+, \tag{28}$$

Eqs.(25), (27), and (28) thus show that

$$(\text{MDL}(k) - \text{MDL}(M)) > 0. \tag{29}$$

Case 2: $k < M$: As $\alpha \to \infty$, the second relation in (24) is invoked, yielding

$$\log \frac{\| \varepsilon^k \|}{\| \varepsilon^M \|} > (M-k)\frac{\log \alpha}{\alpha}. \tag{30}$$

Since $k < M$, one has $\| \varepsilon^k \| > \| \varepsilon^M \|$, (the same argument is used as in case 1), and hence,

$$\log \frac{\| \varepsilon^k \|}{\| \varepsilon^M \|} > 0, \tag{31}$$

and it follows that

$$\log \frac{\| \varepsilon^k \|}{\| \varepsilon^M \|} + (k-M)\frac{\log \alpha}{\alpha} > 0 \text{ as } \alpha \to \infty. \tag{32}$$

Finally, (32) together with the result in (29), yield

$$(\text{MDL}(k) - \text{MDL}(M)) > 0, \tag{33}$$

or a positive function, thus proving the consistency of the detection criterion. ■

3 Simulations

3.0.1 Example 1

A 10 element array with equal spacing of $\lambda/2$ is considered. Two coherent narrowband signals with a common normalized frequency of $f_{1(2)} = .25$ Hz are assumed to impinge on the array from distinct directions ($\theta_{1(2)} = \pm 2.5\,\mathrm{deg}$). The data record length is N=15 snapshots. A set of 100 random trials is used for each SNR to evaluate the detection performance for different methods.

The AIC and the MDL were the basis in the derivation of the algorithm described in [5] to detect the number of uncorrelated signals impinging on a uniform linear array. The MDL-based test was later extended to detect coherent signals via a unidimensional search [9, 8]. The latter consisted of applying an MDL-based test on a sequence of smoothed covariance matrices of different subarray sizes p. This approach, as previously discussed, assumed a very similar derivation as that first given by Wax and Kailath [5]. We therefore compare our proposed method to that solution and use the MDL and the AIC criteria in the evaluation.

For each p, an estimate of the effective rank is obtained. The principle of rank stabilization described earlier, is then applied to determine the rank which corresponds to the number of sources. The DDDET (prior to optimization-i.e. without the additional whitening step) algorithm proposed in section 3 is first evaluated in this example. In Table 1, the number of correct order estimates (out of 100) is listed in a column corresponding to each of the algorithms under study, for various subarray sizes. The AIC-based solution turned out to be consistently overestimating the number of sources, and its performance further degrades as the subarray size increases (equivalently the number of subarrays decreases). As the number of subarrays decreases, the averaging procedure which effects the

smoothing is over fewer overlapping submatrices. This results in slower enhancement of the eigenvalues and makes their delineation more difficult. This is particularly true when the coherent signals are closely spaced (spatial correlation), and this correlation is thus not adequately reduced. The MDL-based SRP test performs well at high SNR, and exhibits a sharp and sudden drop in performance in a moderately noisy environment. This sharp degradation in performance is partly due to an underlying asymptotic assumption which clearly is not met, but also, as discussed earlier, to the subarrays correlation which is not taken into account. The latter limiting factor clearly affects the performance of DDDET, which still, shows an improvement in performance through a smoother degradation in performance at similar SNR. The unsatisfactory performance of the DDDET at low SNR may thus be attributed to the invalidity of the assumed model error covariance matrix $\mathbf{\Gamma}$. The effect becomes more pronounced for higher spatial correlation (more closely spaced signals) as illustrated in Fig. 1, where the rank stabilization rule is applied at 17.5 dB to compare the performances of the above methods as a function of spatial correlation.

3.0.2 Example 2

In this example we consider the same signal scenario as in example 1. By accounting for the existing subarrays inter-correlation, we estimate the error covariance matrix and perform the additional processing to achieve the whitening of the modeling error. The resulting test, referred to as ODDDET, leads to an improved detection breakdown threshold as illustrated in Fig. 2. The performance, as noted earlier, will be commensurate with the quality of the estimated noise covariance matrix and with the data record length.

In Fig. 3, the rank stabilization criterion is applied at 15 dB to conclude a rank of 2, as the stabilization has taken place over 3 consecutive subarray sizes. We note

that the spatial correlation combined with the temporal, increase the amount of required smoothing to delineate the smallest signal eigenvalue from the largest noise eigenvalue. The stabilization, however, which holds w.p.1 as discussed in [14], is the significant factor as it is used to determine the number of signals. The overall detection performance is summarized in Table 2.

The efficiency and robustness of signals enumeration (even if fully correlated) via coding of the *prediction errors* of a model fitted to an appropriately constructed data matrix, is reflected by the performance improvement even for relatively short data records.

3.0.3 Example 3

In order to demonstrate the performance of the algorithm in an environment where only a subgroup of the arriving signals are correlated, we consider a scenario with two closely spaced coherent sources and one uncorrelated source. The coherent signals have a common frequency $f_{1,2} = .25$ Hz and arrive at $\theta_{1,2} = \pm 2.5\,\text{deg}$, while the uncorrelated signal has a frequency $f_3 = .35$ Hz and impinges from $\theta_3 = 7.5\,\text{deg}$. The SNR is taken to be 15 dB and the observation time is equivalent to $N = 15$ snapshots. In Table 3, the performance of the algorithm in presence of this scenario of signals (correlated and uncorrelated) is shown. The ability of this algorithm to cope with different signal scenarios and to resolve closely spaced signals is nicely demonstrated.

In Table 4, for the same signal scenario, the performance is monitored for several SNR.

4 Conclusion

It was shown that the "vestigial" correlation of the signals following a smoothing transformation needs to be accounted for if a unidimensional signal enumeration technique is

to be used. Encoding the resulting prediction errors results in a significant performance improvement, with a modest computational requirement.

Acknowledgement

We are greatful to the reviewers for all their comments which led to an improvement of this paper.

References

[1] H. Akaike, A New Look at the Statistical Model Identification, *IEEE Trans. on Automatic Control, Vol. AC-19, No. 6, Dec. 1974, pp.716-723.*

[2] J. Rissanen, Modeling by Shortest Data Description, *Automatica, Vol. 14 , pp. 465-471, 1978.*

[3] J. Rissanen, Stochastic complexity and modeling, *Annals of Statistics*, Vol. 14, pp. 1080-1100, 1986.

[4] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, World Scientific Publisher, 1989.

[5] M. Wax and T. Kailath, Detection of Signals by Information Theoretic Criteria, *IEEE Trans. on ASSP, Vol. 33, No. 2, pp. 387-392, 1985.*

[6] M. Wax and I. Ziskind, Detection of the Number of Coherent Signals by the MDL Principle, *IEEE Trans. on ASSP, Vol. 37, No. 8, Aug. 1989.*

[7] M. Wax, Detection and Localization of Multiple Sources via the Stochastic Signals Model, *IEEE Trans. on ASSP, Vol. 39, No. 11, pp. 2450-2456, 1991.*

[8] S. Prasad and B. Chandna, An Augmented Smoothed Rank Algorithm for Determination of Source Coherence Structure, *IEEE Trans. on ASSP, Vol. 37, No. 7, pp. 1144-1146, 1989.*

[9] T. J. Shan, A. Paulraj and T. Kailath, On Smoothed Rank Profile Tests in Eigenstructure Methods for Direction Of Arrival Estimation, *IEEE Trans. on ASSP, Vol. 35, No. 10, pp. 1377-1385, 1987.*

p=6			
SNR (dB)	AIC	MDL	DDDET
20	63	100	100
17.5	61	98	95
15	63	20	91
12.5	63	0	81
10	58	0	30

p=7			
SNR (dB)	AIC	MDL	DDDET
20	20	100	100
17.5	26	94	100
15	22	10	91
12.5	23	0	78
10	26	0	28

p=8			
SNR (dB)	AIC	MDL	DDDET
20	1	100	100
17.5	2	61	91
15	1	1	84
12.5	1	0	47
10	1	0	11

Table 1: Comparison of enumeration performances of MDL and AIC based SRP tests on the array covariance matrix and of DDDET. Various subarray sizes ($p = 6, 7, 8$) and SNR's are looked at.

Figure 1: Spacing threshold for detection of two coherent signals

ODDDET					
SNR (dB)	p=4	p=5	p=6	p=7	p=8
20	100	100	100	100	100
17.5	100	100	100	100	98
15	95	94	97	93	84
12.5	92	93	95	92	66
10	91	81	79	53	16

Table 2: Detection performance of ODDDET with two coherent signals

ODDDET					
SNR (dB)	p=4	p=5	p=6	p=7	p=8
15	100	90	96	92	94

Table 3: Detection in presence of a coherent signal subgroup and an uncorrelated signal at various subarray sizes and fixed SNR

SNR (dB)	ODDDET (p=6)
20	100
17.5	100
15	96
12.5	90
10	88

Table 4: Detection performance in presence of a coherent signal subgroup and an uncorrelated signal at various SNR

Figure 2: Performance improvement of ODDDET with two coherent signals

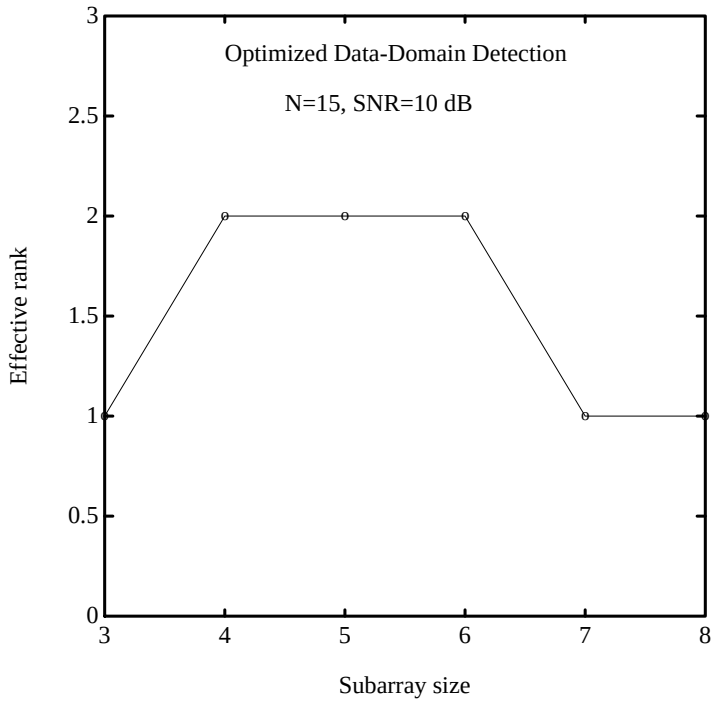

Figure 3: Rank stabilization for a two coherent signal scenario at a fixed SNR

[10] Z. D. Bai, K. Subramanyam and L. C. Zhao, On Determination of the Order of an Autoregressive Model, *Multivariate Statistics and Probability, Vol. 27, 1988, pp. 40-52.*

[11] L. S. Biradar and V. U. Reddy, SVD Based Information Criteria for Detection of the Number of Sinusoids, *Proc. of the ICASSP'90, pp. 2519-2522.*

[12] H. Krim, J. G. Proakis, Enumeration of Fully Correlated Signals, Proc. of Eusipco'92.

[13] H. Krim, Detection and Parameter Estimation of Correlated Signals in Noise, Ph.D. dissertation, Northeastern University, June 1990.

[14] J. H. Cozzens and M. J. Sousa, Enumeration of Fully Correlated Signals, *IEEE Trans. on S.P.*, to appear.

[15] M. Viberg, *Subspace Fitting Concepts in Sensor Array Processing*, Linköping Studies in Science and Technology. Dissertation No. 217, Department of Electrical Engineering, Linköping University, S-581 83, Linköping, Sweden 1989.

[16] M. Viberg, B. Ottersten, and T. Kailath, Direction-of-Arrival Estimation and detection Using Weighted Subspace Fitting, *Proceedings of the Twenty-Third Asilomar Conference on Signals, Systems and Computers*, to appear.

[17] L. C. Zhao, P. R. Krishnaiah and Z. D. Bai, On Detection of the Number of Signals in the Presence of White Noise, *J. Multivariate Anals., Vol. 20, pp. 1-20, 1986.*

[18] M.B. Priestley, *Spectral Analysis and Time Series*, Academic Press, 1989.

[19] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Baltimore, Maryland: The John Hopkins University Press, 1984.

[20] A. Graham, *Kronecker Products and Matrix Calculus*, John Wiley Pub., 1981.

[21] R. A. Horn and C. R. Johnson *Matrix Analysis*, Cambridge University Press, 1990 Edition.