SVD-Based Information Theoretic Criteria for Detection of the Number of Damped/Undamped Sinusoids and Their Performance Analysis

V. Umamathi Reddy, Senior Member, IEEE, and L. S. Biradar

Abstract—Recently, Wax and Kailath developed information theoretic criteria for detection of the number of signals received by a sensor array. More recently, Fuchs developed a criterion, based on the perturbation analysis of the data autocorrelation matrix, for detecting the number of sinusoids. In this paper, following the information theoretic approach to model selection, we first develop criteria for detection of the number of sinusoids. These criteria are matched to the singular value decomposition (SVD) based methods, such as modified forward/backward and forward-backward linear prediction, so well that the extra computations needed over and above those required for computing the SVD are marginal. Next, we develop an analytical framework for analyzing the performance of these criteria, following the assumptions made by Wang and Kaveh and the corrections given by Zhang et al. In the development of the analysis, we make some approximations which become better for large signal-to-noise ratio. Simulations are used to verify the usefulness of the analysis, and to compare the performance of our method with that of Fuchs.

I. INTRODUCTION

The problem of estimating the parameters (i.e., frequencies and damping factors) of sinusoids (damped/undamped) in the presence of additive white noise is an important one and several methods have been recently proposed in this regard (see [5]). At high signal-to-noise ratio (SNR), all the existing methods perform equally well asymptotically. The problem, however, becomes very difficult when the number of available data samples is small and/or the SNR is low.

The forward-backward linear prediction (FBLP) method [6] is capable of resolving closely spaced sinusoids with short data records if SNR is sufficiently high. For short data lengths and moderate SNR's, the modified FBLP (MFBLP) method for undamped sinusoids [7], modified backward linear prediction (MBLP) method for damped sinusoids [8], and a recently proposed total least squares (TLS) method [9] perform much better than most of the model-based methods. A quantitative analysis of the method suggested in [8] has been developed in [17]. All these methods, i.e., MFBLP, MBLP, and TLS, require the knowledge of the number of sinusoids a priori. In practice, however, this a priori knowledge is not available. We therefore need a technique to estimate this number from the observed data.

One can use the method proposed in [10] and [11] to estimate the number of sinusoids. But, it involves a subjective decision of the designer making it difficult to implement in practice. Criteria based on the application of information theoretic principles to model selection, introduced by Akaike (AIC) [12] and by Schwartz [13] and Rissanen (MDL) [14], have been recently developed [1] for the detection of the number of signals received by a sensor array. More recently, Fuchs developed a criterion [2] based on the perturbation analysis of the data autocorrelation matrix for detecting the number of sinusoids, which is in some sense a subjective-based method.

In this paper, following the information theoretic approach, we first develop criteria for the detection of the number of damped/undamped sinusoids. Next, we develop an analytical framework for analyzing the performance of the proposed criteria. In both cases, we make some approximations which become better when SNR is large. Simulations are used to verify the usefulness of the analysis.

The paper is organized as follows. In Section II, we develop the criteria. Section III gives the performance analysis of the proposed criteria and Section IV discusses simulation results. Finally, in Section V, we conclude the paper.

II. DEVELOPMENT OF THE CRITERIA

Consider \( N \) uniformly spaced data samples \( y(n) \) of \( M \) real sinusoids corrupted by additive white Gaussian noise

\[
y(n) = \sum_{i=1}^{M} a_i e^{i\omega_i n} \sin (\omega_i n + \phi_i) + v(n),
\]

where \( a_i \) is the amplitude, \( \alpha_i \) is the damping factor, \( \omega_i \) is the normalized angular frequency, and \( \phi_i \) is the initial phase (assumed to be uniformly distributed in the interval
[0, 2\pi) of the \(i\)th sinusoid. \(\{v(n)\}\) are zero-mean, independent and identically distributed real Gaussian random variables with variance \(\sigma^2\).

For the observed data, the set of \(L\)th order \((2M \leq L \leq N - 2M)\) forward and backward linear prediction equations, in matrix form, is given by

\[
\begin{bmatrix}
y(L) & y(L-1) & \cdots & y(1) \\
y(L+1) & y(L) & \cdots & y(2) \\
\vdots & \vdots & \ddots & \vdots \\
y(N-1) & y(N-2) & \cdots & y(N-L) \\
y(2) & \cdots & \cdots & \cdots \\
y(3) & \cdots & \cdots & \cdots \\
\vdots & \vdots & \ddots & \vdots \\
y(N-L+1) & y(N-L+2) & \cdots & y(N)
\end{bmatrix}
\begin{bmatrix}
g_1 \\
g_2 \\
g_3 \\
g_4 \\
g_5 \\
g_6 \\
\end{bmatrix}
= \begin{bmatrix}
y(L+1) \\
y(L+2) \\
\vdots \\
y(N) \\
y(1) \\
y(2) \\
\vdots \\
y(N-L)
\end{bmatrix}
\begin{bmatrix}
e_1(L+1) \\
e_1(L+2) \\
\vdots \\
e_1(N) \\
e_2(1) \\
e_2(2) \\
\vdots \\
e_6(N-L)
\end{bmatrix}
\]

which can be expressed compactly as

\[
Ag = y - \xi.
\]

For FLP/BLP, the data matrix \(A\), the predicted data vector \(y\), and the prediction error vector \(\xi\) will be first/second half (above/below dotted line) in (2.2) with sizes \((N-L)\times L\), \((N-L)\), and \((N-L)\), respectively.

In MFLP, MBLP, and MFBLP methods, the matrix \(A\) is replaced by its \(2M\)-rank approximant. This requires an \textit{a priori} knowledge of the number of sinusoids or a method which derives this information from the given data.

The Akaike's criterion for model selection selects the model which minimizes the AIC, defined by

\[
\text{AIC} = -2 \ln \left( f \left( y \mid \hat{\theta} \right) \right) + 2d
\]

where \(f \left( y \mid \hat{\theta} \right)\) is the maximum of the likelihood function of the observations and \(d\) is the number of free parameters in \(\theta\). The MDL criterion selects the model which minimizes the MDL, defined by

\[
\text{MDL} = -2 \ln \left( f \left( y \mid \hat{\theta} \right) \right) + d \ln N
\]

where \(N\) denotes the number of observations in \(y\) (i.e., the length of the vector \(y\)).

### A. The Detection Criteria

To apply the information theoretic approach, it is necessary to choose an appropriate family of density functions for the observed data \(y(1), y(2), \cdots, y(N)\). From (2.1), the data \(y(n)\) can be expressed as

\[
y(n) = x(n) + v(n)
\]

where \(x(n)\) is deterministic for a given set of initial phases \(\{\phi_i\}\).

Now recall that the samples of a signal consisting of \(M\) real sinusoids (damped/undamped) satisfy a homogeneous difference equation of order \(L \geq 2M\). The set of forward and backward homogeneous difference equations, in matrix form, is given by

\[
A_{x,g} = x
\]

where \(A_x\) and \(x\) are same as \(A\) and \(y\), respectively, with \(y(n)\)’s replaced by \(x(n)\)’s, and \(g\) is the vector of coefficients. We note here that the \(g\) in (2.7) belongs to an affine subspace. From (2.6) and (2.7), we have

\[
y = A_{x,g} + \nu
\]

where (with superscript \(T\) denoting the transpose)

\[
\nu = [v(L+1) \quad v(L+2) \cdots v(N) \quad v(1) \quad v(2) \cdots v(N-L)]^T.
\]

We may point out here that similar relation holds for FLP/BLP excepting that \(A_x\), \(y\) and \(\nu\) are the respective half portions of those corresponding to FBLP. Since \(A_{x,g}\) is deterministic for a given set of initial phases \(\{\phi_i\}\), the family of probability density functions for the vector \(y\) can be found from the density function of the vector \(\nu\).

We note from (2.9) that for \((N-L) \geq (L+1)\), some elements in \(\nu\) repeat making the covariance matrix of \(\nu\) rank deficient. Hence, the density function of \(\nu\) does not exist for this case. However, no problem arises for the case \((N-L) < (L+1)\) in FBLP and for any \(L\) in FLP and BLP. Since the MFBLP has been shown to perform better when \(L\) is in the range \((2N/3, 3N/4)\) [7] and these values of \(L\) correspond to the condition \((N-L) < (L+1)\), we assume in the following development that \(L\) is such that the density function of \(\nu\) exists. The development of the criteria for the case \((N-L) \geq (L+1)\) is discussed in [15].

Thus, the parametrized density function of \(y\) with parameters \(g\) and \(\sigma^2\) is given by

\[
f(y \mid g, \sigma^2) = (2\pi \sigma^2)^{-d/2} \cdot \exp \left( -\frac{(y - A_{x,g})^T(y - A_{x,g})}{2\sigma^2} \right)
\]
where
\[ r = \begin{cases} 2(N - L) & \text{for FBLP} \\ (N - L) & \text{for FLP and BLP}. \end{cases} \]  
(2.11)

In practice, we do not have the matrix \( A \), and what is available to us is the data-matrix \( X \). We now make an approximation by replacing \( A \) with a low-rank approximate of \( A \), which we denote by \( A_{2k} \). Thus, the approximate family of density functions is given by

\[
f(y|\theta, \sigma^2) = \frac{2\pi \sigma^2}{}^{-1/2} \exp(-\frac{(y - \bar{A}_{2k} \theta)^T (y - \bar{A}_{2k} \theta)}{2\sigma^2}).
\]
(2.12)

Now, consider the negative of the logarithm of (2.12)

\[
L^{(k)}(\theta, \sigma^2) = -\ln(f(y|\theta, \sigma^2)) = \frac{r}{2} \ln 2\pi + \frac{r}{2} \ln \sigma^2 + \frac{\|y - \bar{A}_{2k} \theta\|^2}{2\sigma^2}
\]
(2.13)

where \( \| \cdot \| \) denotes the 2-norm. The ML estimates of \( \theta \) and \( \sigma^2 \) are the values of \( \theta \) and \( \sigma^2 \) that minimize \( L^{(k)}(\theta, \sigma^2) \), and are given by

\[
\theta^{*} = \frac{\|y - \bar{A}_{2k} (\bar{A}_{2k}^T \bar{A}_{2k})^+ \bar{A}_{2k}^T \bar{X} \|^2}{r}
\]
(2.14)

\[
\hat{\theta} = (\bar{A}_{2k}^T \bar{A}_{2k})^+ \bar{A}_{2k}^T \bar{X}.
\]
(2.15)

where the superscript denotes pseudoinverse. Observe that the ML estimate \( \hat{\theta} \) is the minimum-norm solution for the predictor coefficient vector in MFPL/MFLP/MFBLP with the assumption that the number of sinusoids is \( k \). We may point out here that the above estimates also follow from the development in [7], [8]. From (2.14), (2.15), and (2.13), we get

\[
L^{(k)}(\hat{\theta}, \hat{\sigma}^2) = \frac{r}{2} \left( 1 + \ln 2\pi + \ln \frac{\|y - \bar{A}_{2k} (\bar{A}_{2k}^T \bar{A}_{2k})^+ \bar{A}_{2k}^T \bar{X} \|^2}{r} \right).
\]
(2.16)

To simplify the last term in (2.16), we use the SVD of \( \bar{A}_{2k} \)

\[
\bar{A}_{2k} = \bar{U} \left( \Sigma_{2k} \begin{array}{c} 0 \\ 0 \\ 0 \end{array} \right) \bar{V}^T
\]
(2.17)

where \( \bar{U} \) and \( \bar{V} \) are the unitary matrices of size \( r \times r \) and \( L \times L \), containing the left and right singular vectors, respectively, of \( \bar{A} \) and \( \Sigma_{2k} \) is a \( 2k \times 2k \) diagonal matrix whose diagonal elements are the \( 2k \) largest singular values of \( \bar{A} \). Substituting (2.17) into the norm-squared term in (2.16) and noting that 2-norm is invariant under unitary transformation, we get

\[
\| \cdot \|^2 = \left\| \bar{U}^T y - \bar{U} \left( \begin{array}{c} I_{2k} \\ 0 \\ 0 \end{array} \right) \bar{U}^T y \right\|^2
\]
\[
= \left\| \left( \begin{array}{c} e \\ h^{2k} \end{array} \right) - \left( \begin{array}{c} I_{2k} \\ 0 \\ 0 \end{array} \right) \left( \begin{array}{c} e \\ h^{2k} \end{array} \right) \right\|^2 = \| h^{2k} \|^2
\]
(2.18)

where \( h^{2k} \) is the vector containing the last \( (r - 2k) \) elements of \( \bar{U}^T y \)

\[
\bar{U}^T y = \left( \begin{array}{c} e \\ h^{2k} \end{array} \right) \{ (r - 2k) \text{elements} \}
\]
(2.19)

Thus, (2.16) simplifies to

\[
L^{(k)}(\hat{\theta}, \hat{\sigma}^2) = \frac{r}{2} \left( 1 + \ln 2\pi + \ln \frac{\| h^{2k} \|^2}{r} \right).
\]
(2.20)

To determine the number of degrees of freedom in the vector \( \hat{\theta} \), consider (2.15). Combining this with (2.17) and (2.19), we get

\[
\hat{\theta} = \bar{V} \left( \Sigma_{2k}^{-1} e \right)
\]
(2.21)

which shows that \( \hat{\theta} \) is the linear combination of the first \( 2k \) vectors of \( \bar{V} \), implying that \( \hat{\theta} \) has \( 2k \) degrees of freedom. We note here that the columns of \( \bar{V} \) remain the same and that indeed \( \hat{\theta} \) is a linear combination of a varying number of them, hence the indicated number of degrees of freedom. Thus, under the assumption of \( k \) sinusoids, the number of free parameters in \( \hat{\theta} = (\hat{\theta}, \hat{\sigma}^2) \) is \((2k + 1)\).

Using this value in place of \( d \) and \( r \) in place of \( N \) in (2.4) and (2.5), and combining with (2.20), we obtain

\[
\text{AIC} (k) = r \left( 1 + \ln 2\pi + \frac{\ln \| h^{2k} \|^2}{r} \right) + 2(2k + 1)
\]
(2.22)

\[
\text{MDL} (k) = r \left( 1 + \ln 2\pi + \frac{\ln \| h^{2k} \|^2}{r} \right) + (2k + 1) \ln r.
\]
(2.23)

B. Computational Load of the Criteria

The method requires one SVD computation of \( A \) and evaluation of \( \| h^{2k} \|^2 \) for different values of \( k \). The computation of \( \| h^{2k} \|^2 \) can be implemented recursively as follows.

After evaluating \( \bar{U} \) through SVD, compute the squares of the elements of \( \bar{U}^T y \) and store them. Let \( \bar{U}^T y \), denote
the $i$th term of $\hat{U}^T y$. We can then use the following recursions to evaluate $\|h^k\|^2$:

$$\|h^0\|^2 = \sum_{i=1}^{r} \{(\hat{U}^T y)_i\}^2$$

$$\|h^{2k}\|^2 = \|h^{2k-2}\|^2 - \{(\hat{U}^T y)_{2k-1}\}^2 - \{(\hat{U}^T y)_{2k}\}^2, \quad k = 1, 2, \ldots, \frac{r}{2} - 1. \tag{2.25}$$

Thus, the additional computations needed over and above those required for the SVD calculation are marginal. In contrast, the method proposed by Fuchs [2] requires computations for 1) estimating the autocorrelation matrix of the observed data, 2) eigen-decomposition of the autocorrelation matrix, 3) estimating the covariance matrix of the perturbed eigenvalues of the autocorrelation matrix, and 4) computing the chi-squared (denoted as $\chi^2$) random variable. Clearly, this method’s computational load is much higher than ours. Furthermore, Fuchs method involves testing of chi-squared random variable against a threshold, thus requiring a subjective decision of the designer. However, it is the first dedicated method to this problem and the resulting test is theoretically justified asymptotically in the number of data.

Each of the criteria, (2.22) and (2.23), is evaluated for different values of $k$ and the $k$ for which the criterion attains minimum is taken as the estimate of the number of sinusoids as given by that criterion. Now, the issue is up to what value of $k$ should we evaluate the criteria.

We note that as $k$ increases, the norm square of the vector $h^k$, $\|h^k\|^2$, decreases and reaches zero value when $k = r/2$, resulting in the minimum value for both the criteria at $r/2$. We therefore have to restrict $k$ to less than $r/2$. Normally, $k$ should be varied up to $L/2$ since an $L$th order predictor can handle up to $L/2$ real sinusoids. In practice, however, the predictor order $L$ is chosen significantly higher than the minimum required value $2M$, and it is usually selected such that the data matrix $A$ is nearly square (when the modified methods are known to perform well). For such values of $L$, we suggest the evaluation of the criteria from $k = 1$ to $L/3$. In our study, we also consider a reasonably small value$^1$ of $L$ which gives rise to an overdetermined set of equations in (2.3). For this value of $L$, we suggest the search up to $L/2$. We should, however, point out that when $L$ is chosen such that $(N - L) \geq (L + 1)$, we need to use certain linear transformations in the development of the criteria if we are working with MFBLP (see [15]). Though we have not considered such a case in the following analysis, it can be handled easily by modifying the intermediate steps.

This case is considered because of the remark made by one of the reviewers.

III. PERFORMANCE ANALYSIS OF THE PROPOSED CRITERIA

In this section, we develop an analytical framework for analyzing the performance of the criteria derived in Section II. To develop a common framework for both criteria, we denote the RHS of (2.22) and (2.23) as

$$\Phi(k) = r \left\{ 1 + \ln 2\pi + \ln \frac{\|h^{2M-1}\|^2}{r} \right\} + p(k) \tag{3.1}$$

where $p(k)$ is the bias correction.

Let $H_M$ denote the hypothesis that the true number of sinusoids is $M$. Then the probabilities of underestimating and overestimating the number of sinusoids, given $H_M$, are defined, respectively, as

$$P_M = P(\hat{M} < M \mid H_M) \tag{3.2}$$

$$P_F = P(\hat{M} > M \mid H_M) \tag{3.3}$$

It has been found via extensive computer simulations that the above assumptions are indeed true for the MDL criterion when the SNR is moderate and higher.

Define the probability of error $P_e$ as

$$P_e = P_M + P_F \tag{3.6}$$

where $P_M$ is the probability of miss and $P_F$ is the probability of false alarm. Following the development as in [4], we can show that

$$P(B_1) = P(B_1) + P(B_2) - P(B_1 \cap B_2) \tag{3.7}$$

where the events $B_1$ and $B_2$ are given by

$$B_1: \Phi (M - 1) < \Phi (M) \tag{3.8}$$

$$B_2: \Phi (M + 1) < \Phi (M). \tag{3.9}$$

In the sequel, we do not show the conditioning on $H_M$ as this has already been reflected in the specification of the events.

From (3.1) and (3.8), we have

$$P(B_1) = P \left\{ 1 + \ln 2\pi + \ln \frac{\|h^{2M-1}\|^2}{r} \right\} + p(M - 1)$$

$$< r \left\{ 1 + \ln 2\pi + \ln \frac{\|h^{2M}\|^2}{r} \right\} + p(M) \tag{3.10}$$
which can be simplified as

\[ P(B_1) = P \left( \sum_{i=2M+1}^{2M} \{ (\tilde{U}^T y)^2 \}^2 \right) \]

Similarly, combining (3.1) and (3.9), we get

\[ P(B_2) = P \left( \sum_{i=2M+1}^{2M} \{ (\tilde{U}^T y)^2 \}^2 \right) > e^{(p(M+1) - p(M))/r - 1} \]  

(3.11)

To evaluate \( P(B_1) \) and \( P(B_2) \), we need to determine the density functions of the quotients in (3.11) and (3.12).

Note that \( U \) is the matrix containing the left singular vectors of \( A \) or the matrix containing the eigenvectors of \( \Lambda A A^T \).

Now, consider

\[ \frac{1}{L} A A^T = \frac{1}{L} A_s A_s^T + \frac{1}{L} A_t A_t^T + \frac{1}{L} A_s A_t^T + \frac{1}{L} A_t A_s^T \]  

(3.13)

where \( A_s \) and \( A_t \) are the signal only and noise only parts of \( A \), respectively. For moderate values of \( L \), second and third terms become relatively small as the SNR increases.

For a finite but moderate SNR, the fourth term is closer to \( a^2 I \) in the FLP/BLP case than in the FBP case because of the following. In FBP case, \( \frac{1}{L} A_s A_s^T \) consists of four \( (N-L) \times (N-L) \) block matrices where the two diagonal blocks will be closer to \( a^2 I \) and the other two off-diagonal blocks contain an element such as \( \frac{1}{L} v^2 \) in addition to cross terms in alternate positions. On the other hand, in FLP/BLP case, \( (1/L) A_s A_s^T \) is simply the first/seond diagonal block of the FBP. In the special case of no noise, the last three terms in (3.13) vanish.

Keeping the above arguments in view, we make the following approximation:

\[ \frac{1}{L} A A^T = \frac{1}{L} A_s A_s^T + \sigma^2 I \]  

(3.14)

which is more justifiable in the FLP/BLP case than in the FBP case. For the given initial phases of the sinusoids and noise variance, the RHS of (3.14) is known and its eigenvector matrix, say \( U \), can be determined. Thus, \( U \) is independent of the noise realization. Thus, for the given initial phases of the sinusoids, \( U \) can be treated as constant. Now, approximating \( \tilde{U} \) to \( U \) in (3.11) and (3.12), and noting that \( y(n) \) is Gaussian with mean \( x(n) \) and variance \( \sigma^2 \), we have

\[ E[(U^T y)_{i}^2] = E[u_i^T v v^T u_i] = u_i^T E[v v^T] u_i = \sigma^2, \quad 1 \leq i \leq r \]  

(3.15)

\[ \text{var} ((U^T y)_i) = E[u_i^T v v^T u_i] = u_i^T E[v v^T] u_i \]

\[ = \sigma^2, \quad 1 \leq i \leq r. \]  

(3.16)

Note that \( x \), like \( A \), depends on the initial phases of the sinusoids.

Under the hypothesis \( H_{3M} \), the eigenvectors \( u_1, u_2, \ldots, u_{3M} \) span the true signal subspace and the remaining \( u_{3M+1}, u_{3M+2}, \ldots, u_r \) span the true noise subspace. It therefore follows that

\[ E[(U^T y)_{i}] = u_i^T x, \quad 1 \leq i \leq 2M \]  

(3.17)

and

\[ E[(U^T y)_{i}] = u_i^T x = 0, \quad 2M + 1 \leq i \leq r. \]  

(3.18)

Further, from the orthogonality of the singular vectors, we get

\[ \text{cov} ((U^T y)_i, (U^T y)_j) = E[u_i^T v v^T u_j] = u_i^T E[v v^T] u_j = 0, \quad 1 \leq i, j \leq r, \quad i \neq j. \]  

(3.19)

Since \( y \) is Gaussian, (3.19) implies that the components of (\( U^T y \)) are independent.

Now, consider

\[ Z_1 = \sum_{i=2M+1}^{2M} \{ (U^T y)^2 \}^2 = \frac{S_1}{T_1} \]  

(3.20)

It is easy to see that \( S_1 \) is a noncentral \( \chi^2 \) random variable with 2 degrees of freedom. Replacing \( q, \lambda, \) and \( n \) in (A.4) (see Appendix A) with \( s_1, \lambda_1, \) and \( 2, \) respectively, the density function of \( S_1 \) is obtained as

\[ f_S(s_1) = \frac{e^{-s_1/2} \lambda_{1/2}^2 e^{2s_1/4}}{2\sigma^2} \left[ 1 + \frac{1}{2} \left( \frac{s_1 \lambda_1}{2\sigma^2} \right) \right] \]

\[ + \frac{1}{2} \cdot 4 \left( \frac{s_1 \lambda_1}{2\sigma^2} \right)^2 + \cdots \]

\[ 0 \leq s_1 < \infty \]  

(3.21)

where

\[ \lambda_1 = \{(U^T x)_{3M+1}^2\}^2 + \{(U^T x)_{2M}\}^2. \]  

(3.22)

Now, consider the random variable \( T_1 \). From (3.16), (3.18), and (3.19), it follows that \( T_1 \) is a \( \chi^2 \) random variable with \( (r - 2M) \) degrees of freedom and, hence, its density function is given by

\[ f_T(t_1) = \frac{t_1^{(r-2M)/2-1} e^{-t_1/2\sigma^2}}{(2\sigma^2)^{(r-2M)/2} \Gamma \left( \frac{r-2M}{2} \right)} \]

\[ 0 \leq t_1 < \infty. \]  

(3.23)
Since $S_i$ and $T_i$ are independent, the density function of $Z_i = S_i/T_i$ is given by

$$f_{Z_i}(z_i) = \int_0^\infty t_i f_{S_i}(t_i z_i) f_{T_i}(t_i) \, dt_i \quad (3.24)$$

which, using (3.21) and (3.23), can be simplified to

$$f_{Z_i}(z_i) = \frac{r - 2M}{2} e^{-\lambda_i/2\sigma^2} \left(1 + z_i\right)^{-(r - 2M + 2)/2} \quad (3.25)$$

Observe that $f_{Z_i}(z_i)$ resembles Rician distribution whose peak location from the origin is proportional to the non-centrality parameter $\lambda_i/\sigma^2$.

From (3.12) and (3.20), we have

$$P(B_i) = P(Z_i < e^{(r(M - 1)/r - 1)} \quad (3.26)$$

Substituting (3.25) in (3.26) and evaluating the integral numerically, we obtain $P(B_i)$ for a given set of initial phases of the sinusoids.

Now, consider the density function of the quotient in (3.12). Let

$$S_i = \sum_{i = 1}^{M+1} \left[U_i^T y_{i+1}\right]^2 \quad (3.27)$$

Here, $S_i$ and $T_i$ are $\chi^2$ random variables with 2 and $(r - 2(M + 1))$ degrees of freedom, respectively. Hence, the density functions of $S_i$ and $T_i$ are given by

$$f_{S_i}(s_i) = \frac{e^{-s_i/2\sigma^2}}{2\sigma^2} \quad 0 \leq s_i < \infty \quad (3.28)$$

and

$$f_{T_i}(t_i) = \frac{t_i^r e^{-t_i/(2\sigma^2)}}{(2\sigma^2)^{r/2} \Gamma(r/2)} \quad 0 \leq t_i < \infty \quad (3.29)$$

Using the fact that $S_i$ and $T_i$ in (3.27) are independent, the density function of $Z_2$ can be shown to be

$$f_{Z_2}(z_2) = \frac{r - 2M}{2} e^{-\lambda_i/2\sigma^2} \left(1 + z_2\right)^{-(r - 2M + 2)/2} \quad 0 \leq z_2 < \infty \quad (3.30)$$

From (3.12), (3.27), and (3.30), the probability of the event $B_2$ is given by

$$P(B_2) = P(Z_2 > e^{(r(M - 1)/r - 1)} = e^{-(r - 2(M + 1))/(r - 2M + 1)} \quad (3.31)$$

Note that $P(B_2)$ does not depend on the signal and noise powers; it depends only on the penalty and $r$ (cf. (2.11)).

Consider now the probability of the event $(B_1 \cap B_2)$. From Appendix B,

$$P(B_1 \cap B_2) \leq P(Z_3 < e^{(r(M - 1)/r - 1)} P(B_2) \quad (3.32)$$

where

$$Z_3 = \frac{S_1}{S_2} = \frac{\left[U_i^T y_{i+1}\right]^2 + \left[U_i^T y_{i+2}\right]^2}{\left[U_i^T y_{i+1}\right]^2 + \left[U_i^T y_{i+2}\right]^2} \quad (3.33)$$

Since $S_i$ and $S_2$ are independent, the density function of $Z_3$ can be shown to be

$$f_{Z_3}(z_3) = \frac{e^{-\lambda_i/(2\sigma^2)}}{2\sigma^2} \left(1 + z_3\right)^{-(r - 2M + 2)/2} \quad 0 \leq z_3 < \infty \quad (3.34)$$

From (3.31) and (3.34), (3.32) can be evaluated numerically for a given set of initial phase angles of the sinusoids.

Combining (3.7), (3.26), (3.31), (3.32), and (3.34), the bounds for the probability of detection $P_D = 1 - P_\phi$, given by

$$1 - P(B_i) = P(B_2) \leq P_D \leq 1 - P(B_1) - P(B_2) + P(Z_3 < e^{(r(M - 1)/r - 1)} P(B_2) \quad (3.35)$$

can be evaluated.

Note that the bounds for $P_D$ are functions of the initial phase angles of the sinusoids, which are assumed to be uniformly distributed in the interval $[0, 2\pi]$. We evaluated these bounds choosing various sets of initial phase angles in the interval $[0, 2\pi]$ and then determined the unconditional probability of detection through numerical integration. To clarify this further, consider the case of two sinusoids. Without loss of generality, let the initial phase of one of the sinusoids be zero and the phase of the other sinusoid be $\phi$. Let $P_D(\phi)$ denote the probability of detection with these initial phases. Then, the unconditional proba-
bility of detection, denoted as $P_{D_{(\text{dec})}}$, is given by

$$P_{D_{(\text{dec})}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_{D_{(\phi)}} d\phi$$

(3.36)

since $\phi$ is assumed to be uniformly distributed in $[0, 2\pi]$. To obtain $P_{D_{(\text{dec})}}$ numerically, we vary the phase $\phi$ in steps of $\Delta \phi$ (say) covering the range 0 to $2\pi$, and determine the probability of detection at each of these steps, and then use the numerical integration technique to evaluate the definite integral part of (3.36).

IV. SIMULATION RESULTS

In this section, we present some simulation results to demonstrate both the effectiveness of our method and the usefulness of the analysis.

To demonstrate the effectiveness of our method compared to that of Fuchs [2], we considered the same example as the one in [2]. The data samples were generated from

$$y(n) = \sqrt{20} \sin 2\pi f_1 n + \sqrt{2} \sin (2\pi f_2 n + \phi) + v(n),$$

$n = 1, 2, \ldots, N$

choosing the normalized frequencies $f_1$ and $f_2$ as 0.2 and $0.2 + \delta$, respectively, with $\delta = 1/N$ or $1/2N$, and the noise variance, $\sigma^2$, as unity. The number of samples $N$ was chosen as 64. We used the FBLP formulation with $L = 32$ (Fuchs used the same formulation for estimating the autocorrelation matrix). The phase of the second sinusoid was varied randomly from trial to trial in the interval $[0, 2\pi]$ and the seed for the noise sequence was selected randomly in each trial. We performed 50 Monte Carlo trials and determined the number of times the criteria attained the minimum value at different values of $k$ (see Tables I and II). Note from the tables that the MDL criterion detected correctly 42 times out of 50 when $\delta = 1/64$ and 20 times out of 50 when $\delta = 1/128$, while the Fuchs method gives 18 and 9 times out of 20, respectively, for $\delta = 1/64$ and $\delta = 1/128$ (these results are taken from [2]). This shows that the two methods perform similarly. The simulation results also show that the detection performance of the MDL is much superior to that of AIC, and that the penalty (or the bias correction) used in AIC is too low to hold the criterion from leaning towards higher model order.

We now present some simulation results to demonstrate the usefulness of the analysis developed in Section III. The undamped sinusoids case is considered first.

A. Case i): Undamped Sinusoids

The data model used is given by

$$y(n) = \sqrt{20} \sin 2\pi f_1 n + \sqrt{2} \sin (2\pi f_2 n + \phi) + v(n),$$

$n = 1, 2, \ldots, N$

where the frequencies $f_1$ and $f_2$ were selected as in the previous example with $\delta = 1/64$, and the variance of the noise, $\sigma^2$, was chosen to give the desired signal-to-noise ratio (SNR), defined as $\text{SNR} = 10 \log_{10} 10/\sigma^2$. Both FLP and FBLP formulations were considered in this example. In each case, we considered two values of $L$ (predictor order).

Tables III through V give the probability of detection (in percentage), obtained from the simulations and the analysis. The simulation results were obtained from 500 Monte Carlo trials. In each trial, we chose randomly the seed for the noise realization and for the initial phase in the interval $[0, 2\pi]$. For the analytical result, $\phi$ was varied uniformly in steps of $\pi/18$ and the conditional $P_{D_{(\phi)}}$ was computed for each setting of $\phi$ as discussed in the previous section. The unconditional probability of detection was then evaluated from these 36 conditional results through the numerical integration as described in the last section (see the discussion in the last paragraph of Section III). We may point out here that the value of $P(Z_1 < e^{r_{(\phi)1}} - r_{(\phi)1} - 1) P(B_0)$ was negligibly small and hence the probability of detection was essentially determined by $P(B_0)$ and $P(B_1)$, which are now (for all practical purposes) the probabilities of miss and false alarm, respectively.

We note the following from the results of the tables. The predicted value of probability of detection is closer
DETECTION PERFORMANCE OF THE CRITERIA FOR $N = 64$ AND $L = 32$
($f_1 = 0.2, f_2 = 0.2 + 1/64$, MONTE CARLO RUNS = 500)

<table>
<thead>
<tr>
<th></th>
<th>$P_{\text{error}}$ (in percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Analysis</td>
</tr>
<tr>
<td></td>
<td>FLP</td>
</tr>
<tr>
<td>AIC</td>
<td>80.31</td>
</tr>
<tr>
<td>MDL</td>
<td>94.00</td>
</tr>
</tbody>
</table>

DETECTION PERFORMANCE OF THE CRITERIA FOR $N = 64$ AND $L = 42$
($f_1 = 0.2, f_2 = 0.2 + 1/64$, MONTE CARLO RUNS = 500)

<table>
<thead>
<tr>
<th></th>
<th>$P_{\text{error}}$ (in percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Analysis</td>
</tr>
<tr>
<td></td>
<td>FLP</td>
</tr>
<tr>
<td>AIC</td>
<td>81.89</td>
</tr>
<tr>
<td>MDL</td>
<td>94.44</td>
</tr>
</tbody>
</table>

due to the enhanced SNR causes a reduction in the value of $P(B_1)$.

Now consider the effect of increase in the predictor order ($L$) on the theoretical value of $P_D$. The results of the Tables III to V show that $P_D$ decreases with increasing $L$. The reasons for this are as follows. A closer examination of the expression for $P(B_2)$ (cf. (3.31)) reveals that its value increases when $r$ decreases. A similar effect takes place with $P(B_1)$ even though it is difficult to see this from the expressions (3.25) and (3.26). Since an increase in $L$ causes a decrease in $r$, it immediately follows from the above that $P_D$ decreases with increasing $L$.

We may point out here that the theoretical and simulated values of $P_D$ (for MDL) are closer in the case of FLP than in FBLP for the reasons given in the paragraph preceding the approximation (3.14), i.e., for a given $L$ and SNR, the approximation is more justifiable in FLP/BLP case than in FBLP case. Further, the SNR of 5 dB is very close to threshold SNR of the MFLP method for the scenarios considered in Tables III and IV. To estimate the threshold SNR, we have applied the MFLP method to the same 500 data realizations as those used in the detection criteria and determined the mean square error performance in the estimates of $f_1$ and $f_2$ for different values of SNR. Though our criteria can be applied at lower values of SNR than the threshold, the analysis will not be useful in that region. Away from the threshold SNR, the predicted values are close to the simulated, particularly in the FLP/BLP case. To support this further, we considered a scenario with closer spacing, $f_1 = 0.2$ and $f_2 = 0.2 + 1/128$. Table VI gives the results for this case which show the closeness between the theoretical and simulated values of $P_D$ for MDL. Once again, the SNR of 15 dB is approximately equal to the threshold SNR of the MFLP method for the scenario of Table VI.

B. Case ii): Damped Sinusoids

The data model used is given by

$$y(n) = \sqrt{20} e^{-0.1n} \sin 2\pi 0.2n$$
$$+ \sqrt{20} e^{-0.05n} \sin (2\pi 0.24n + \phi) + v(n),$$
$$n = 1, 2, \cdots, N$$

where the variance of the noise, $\sigma^2$, was chosen to give the required SNR, defined as $10 \log_{10} \frac{10}{\sigma^2}$. Unlike case...
TABLE VI
DETECTION PERFORMANCE OF THE CRITERIA FOR N = 64 AND L = 32
(f1 = 0.2, f2 = 0.2 + 1/128, MONTE CARLO RUNS = 500)

<table>
<thead>
<tr>
<th></th>
<th>Analysis</th>
<th>Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FLP</td>
<td>FLP</td>
</tr>
<tr>
<td>AIC</td>
<td>80.31</td>
<td>58.6</td>
</tr>
<tr>
<td>MDL</td>
<td>94.01</td>
<td>87.8</td>
</tr>
</tbody>
</table>

TABLE VII
DETECTION PERFORMANCE OF THE CRITERIA WITH BLP FORMULATION (N = 64, L = 32, MONTE CARLO RUNS = 500)

<table>
<thead>
<tr>
<th></th>
<th>Analysis</th>
<th>Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FLP</td>
<td>FLP</td>
</tr>
<tr>
<td>AIC</td>
<td>80.31</td>
<td>53.2</td>
</tr>
<tr>
<td>MDL</td>
<td>94.02</td>
<td>89.0</td>
</tr>
</tbody>
</table>

P_D (in percentage)

V. CONCLUSIONS

In this paper, following the information theoretic approach to model selection, we first developed criteria for detecting the number of damped/undamped sinusoids. These are so well matched to the SVD-based methods of spectral estimation that the additional computations required over and above those needed for the SVD calculation are marginal. The detection performance of these criteria is compared to that of Fuchs [2] and the results show that MDL criterion, developed in the paper, performs nearly same as the Fuchs method. However, Fuchs methods requires much more computations. Next, an analytical framework has been developed for predicting the detection performance of the criteria. In the development of the analysis, we made some approximations which become better for large SNR. Simulation results show that the theoretically predicted values of probability of detection and those of simulation are close with FLP/BLP formulation even at moderate values of SNR.

APPENDIX A

Let w_i ~ N(μ_i, σ^2), i = 1, 2, ⋅⋅⋅, n and assume that they are statistically independent. Then, the density function of w = Σ_{i=1}^n w_i^2/σ^2 is noncentral χ^2, having n degrees of freedom with the noncentrality parameter λ/σ^2 = Σ_{i=1}^n μ_i^2/σ^2 [16], and is given by

\[ f_w(w) = e^{-\lambda/2\sigma^2} \sum_{j=0}^{\infty} \frac{\lambda^j}{2^j j!} G \left( \left| \frac{1}{2} \lambda \frac{w}{\sigma^2} \right| j + \frac{n}{2} \right) \]

0 ≤ w < ∞ (A.1)

where

\[ G \left( \left| \frac{1}{2} \lambda \frac{w}{\sigma^2} \right| j + \frac{n}{2} \right) = \frac{\Gamma(j + \frac{n}{2})}{\Gamma \left( \frac{n}{2} \right)} e^{-\lambda/2\sigma^2 w^2 j + \frac{n}{2} - 1} \]

0 ≤ w < ∞ (A.2)

with Γ denoting the gamma function. Substituting (A.2) into (A.1), we get

\[ f_w(w) = e^{-\lambda/2\sigma^2} \frac{e^{-\lambda/2\sigma^2 w^2 j + \frac{n}{2} - 1}}{\Gamma \left( \frac{n}{2} \right)} \left\{ 1 + \frac{1}{n} \left( \frac{w \lambda}{2 \sigma^2} \right)^2 \right. \\
+ \frac{1}{n(n + 2)} \left( \frac{w \lambda}{2 \sigma^2} \right)^4 \left. \right\}

0 ≤ w < ∞ (A.3)

Now, the density function of σ^2 w = q (say) immediately follows from (A.3) as

\[ f_q(q) = e^{-\mu q/2\sigma^2} \frac{e^{-\lambda/2\sigma^2 q^2 j + \frac{n}{2} - 1}}{\Gamma \left( \frac{n}{2} \right)} \left\{ 1 + \frac{1}{n} \left( \frac{q \lambda}{2 \sigma^2} \right)^2 \right. \\
+ \frac{1}{n(n + 2)} \left( \frac{q \lambda}{2 \sigma^2} \right)^4 \left. \right\}

0 ≤ q < ∞ (A.4)

APPENDIX B

From (3.11), (3.12), (3.20), and (3.27), the probability of the event \( B_1 \cap B_2 \) is given by

\[ P(B_1 \cap B_2) = \sum \left\{ \frac{S_1}{S_1 + T_2} < e^{(p(M_1 - p(M_2))/r - 1)} \right\} \left\{ \frac{S_2}{T_2} > e^{(p(M_1 + 1) - p(M_2))/r - 1} \right\} \] (B.1)
where we used the fact that $T_1 = S_2 + T_2$. It is easy to verify that
\[ e^{(p(M) - p(M-1))/r} - 1 = e^{(p(M+1) - p(M))/r} - 1 = \mu \] (say).
\[(B.2)\]
Substituting (B.2) into (B.1) and simplifying, we get
\[ P(B_1 \cap B_2) = P \left( \frac{S_1}{S_2 + T_2} < \mu \text{ and } \frac{S_2}{T_2} > \frac{S_2}{T_2} \right) \]
\[ = P \left( \frac{S_1}{S_2} - 1 < \frac{T_2}{S_2} \text{ and } \frac{T_2}{S_2} < \frac{1}{\mu} \right) \]
\[ = P \left( \frac{S_1}{S_2} - 1 < \frac{T_2}{S_2} \right) \cdot P \left( \frac{T_2}{S_2} < \frac{1}{\mu} \right) \]
\[ = P \left( S_1 < \mu - 1 \right) P(B_2). \] (B.3)

Denoting $Z_2 = S_1 / S_2$ and substituting for $\mu$ from (B.2) into (B.3), we get
\[ P(B_1 \cap B_2) \leq P(Z_2 < e^{(p(M) - p(M-1))/r}) P(B_2). \] (B.4)

ACKNOWLEDGMENT

The authors wish to thank the reviewers for their critical and helpful remarks which improved the quality of the paper.

REFERENCES


V. Umamathi Reddy (S'68-M'70-SM'82) received the B.E. and M.Tech degrees in electronics and communication engineering from Osmania University and the Indian Institute of Technology (IIT), Kharagpur, in 1962 and 1963, respectively, and the Ph.D. degree in electrical engineering from the University of Missouri in 1971. He was an Assistant Professor at IIT, Madras, during 1972-1976 and Professor at IIT, Kharagpur, during 1976-1979. During 1979-1982 and 1986-1987, he was a Visiting Professor at the Department of Electrical Engineering, Stanford University. In April 1982, he joined Osmania University as a Professor, and was the Founder-Director of the Research and Training Unit for Navigational Electronics, funded by the Department of Electronics, Government of India. Since April 1988, he has been with the Indian Institute of Science, Bangalore, as a Professor of Electrical Communication Engineering and is presently its Chairman. He has served as a consultant in signal processing to Avionics Design Bureau of Hindustan Aeronautics Limited, Hyderabad, and to Central Research Laboratory, Bharat Electronics Limited, Bangalore. His recent research interests are in sensitivity analysis of high-resolution algorithms, adaptive algorithms, adaptive arrays, and wavelet transforms.

Dr. Reddy is a Fellow of the Indian Academy of Sciences, Indian National Academy of Engineering, and Indian National Science Academy, and Fellow of the Institute of Electronics and Telecommunication Engineers (IETE), India. He received the S. K. Mitra Memorial Award (1989) from IETE for the best research paper.

L. S. Biradar was born in Gulbarga, India, on January 27, 1960. He received the B.E. degree in electronics and communication engineering from Gulbarga University in 1983, the M. Tech. degree in information and controls from the Indian Institute of Technology, Kanpur, India, in 1986, and the Ph.D. degree from the Indian Institute of Science, Bangalore, India, in 1992. From 1983 to 1984 and 1986 to 1988 he was a Lecturer in P.D.A. Engineering College, Gulbarga. From 1988 to 1991 he was at the Indian Institute of Science as a Research Scholar. Presently he is working as an Assistant Professor in P.D.A. Engineering College, Gulbarga. His research interests are in statistical signal processing, parameter estimation, and sensor array processing.