On Information Theoretic Criteria for Determining the Number of Signals in High Resolution Array Processing

KON MAX WONG, SENIOR MEMBER, IEEE, QI-TU ZHANG, MEMBER, IEEE, JAMES P. REILLY, MEMBER, IEEE, AND P. C. YIP

Abstract—An important problem in high resolution array processing is the determination of the number of signals arriving at the array. Information theoretic criteria provide us with a means to achieve this. Two commonly used criteria are the Akaike information criterion (AIC) and minimum descriptive length (MDL) criterion. While the AIC tends to overestimate even at high SNR, the MDL criterion tends to underestimate at low or moderate SNR. By excluding irrelevant parameters, a new log likelihood function has been chosen in this paper. Utilizing this new log likelihood function results in obtaining a set of more accurate estimates of the eigenvalues and in the establishment of modified information theoretic criteria which moderate the performance of the AIC and the MDL criterion. Computer simulations confirm that the modified criteria have superior performance.

1. INTRODUCTION

RECENTLY, high resolution methods of estimating the angles of arrival of signals in sensor array processing have captured the attention of many researchers [1]–[6]. More especially, high resolution estimators utilizing the eigenstructure of the estimated covariance or cross spectral density matrices have gained considerable prominence.

The signal model employed in these methods can be generally formulated such that the received signal vector \( \mathbf{x}(t) \) is an \( M \times 1 \) complex vector given by

\[
\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t)
\]

where \( \mathbf{A} \) is an \( M \times K \) matrix such that

\[
\mathbf{A} = \begin{bmatrix}
    a(\Phi_1) & a(\Phi_2) & \cdots & a(\Phi_K)
\end{bmatrix}
\]

with \( a(\Phi_k) \) being an \( M \times 1 \) complex vector characterized by an unknown parameter vector \( \Phi_k \) associated with the \( k \)-th signal (\( k = 1, \cdots, K \)); \( \mathbf{s}(t) \) is the \( K \times 1 \) vector such that

\[
\mathbf{s}(t) = \begin{bmatrix}
    s_1(t) & s_2(t) & \cdots & s_K(t)
\end{bmatrix}^T
\]

with \( s_k(t) \) being a complex waveform constituting the \( k \)-th signal; and \( \mathbf{v}(t) \) is the \( M \times 1 \) complex noise vector. The assumptions made here are 1) \( K < M \); 2) the \( K \) signals \( s_1(t), \cdots, s_K(t) \) are narrow-band ergodic Gaussian processes with zero mean, and the covariance matrix of \( \mathbf{s}(t) \) is positive definite; 3) the noise vector \( \mathbf{v}(t) \) is an ergodic Gaussian process, independent of the signals, with zero mean and covariance matrix \( \sigma_v^2 \mathbf{I} \) where \( \sigma_v^2 \) is an unknown constant and \( \mathbf{I} \) is the identity matrix; and 4) the matrix \( \mathbf{A} \) contains column vectors \( a(\Phi_k) \), \( k = 1, \cdots, K \), which are linearly independent. It follows from the signal model represented by (1) and the above assumptions that the covariance matrix of \( \mathbf{x}(t) \) is given by

\[
\mathbf{R}_x = E[\mathbf{x}(t)\mathbf{x}^H(t)] = \mathbf{AA}^H + \sigma_v^2 \mathbf{I}
\]

with \( \mathbf{R}_x \) being the covariance matrix of \( \mathbf{s}(t) \) and \( \mathbf{A}^H \) denoting the conjugate transpose. It also follows that the rank of \( \mathbf{AA}^H \) is \( K \), with the smallest \( M - K \) of its eigenvalues being equal to zero. Therefore, if the eigenvalues of \( \mathbf{R}_x \) are ordered in descending order of magnitudes such that \( \lambda_1 \geq \lambda_2 \geq \cdots \), then

\[
\lambda_{K+1} = \lambda_{K+2} = \cdots = \lambda_M = \sigma_v^2.
\]

A problem central to most of the high resolution methods of estimating the angles of arrival of signals in sensor array processing is the determination of the number of signals \( K \) from a finite set of observations \( \mathbf{x}(1), \mathbf{x}(2), \cdots, \mathbf{x}(N) \). This information is necessary for most high resolution methods of estimating the angles of arrival of the signals. One way to solve this problem is based on hypothesis tests using the eigenvalues of the covariance matrix of the observation vector [7]. A threshold is set so that the likelihood ratio statistic can be compared to it in order to decide on the accepted hypothesis. The disadvantage of this method is that the threshold has to be set according to some subjective judgement. To avoid this, Wax and Kailath [8] proposed an approach based on the Akaike information criteria (AIC) [9] and the minimum descriptive length (MDL) criterion introduced by Schwarz [10] and Rissanen [11]. The number of signals is determined by the value for which the AIC or the MDL criterion is minimized:

Let the true number of signals arriving at the sensor array be \( K \). The determination of the number of signals using the approach of information theoretic criteria first formulates a family of conditional probability density functions (pdf) \( f(X | \theta(k)) \) which are dependent on the
assumed number of signals $k$. Here, $X$ is the given set of
$N$ observations $X = \{x(1), \cdots, x(N)\}$, and $\theta(k)$ is
the parameter vector of the model

$$\theta(k) = [\lambda_1, \cdots, \lambda_k, \sigma^2, v_1, \cdots, v_k]$$  \hspace{1cm} (6)

with $v_i$ being the eigenvector corresponding to the eigenvalue $\lambda_i$. Then, the model to be selected is the one which either minimizes the AIC defined by

$$C_A = -\log f(X | \hat{\theta}) + P$$ \hspace{1cm} (7)

or minimizes the MDL defined by

$$C_{DL} = -\log f(X | \hat{\theta}) + \frac{P}{2} \log N$$ \hspace{1cm} (8)

where $\hat{\theta}$ is the maximum likelihood estimate of the parameter vector $\theta$, and $P$ is the number of free parameters in $\theta$. The first term in both (7) and (8), is the well-known log likelihood function. The second term is usually called the penalty function. By assuming that the observation vector $x(1), \cdots, x(N)$ are statistically independent and identically distributed complex Gaussian random vectors of zero mean, and using the results obtained by Anderson [12], Wax and Kailath obtained the maximum likelihood estimates of the parameters in $\theta(k)$ such that

$$\hat{\lambda}_i = l_i, \quad i = 1, \cdots, k$$ \hspace{1cm} (9a)

$$\hat{\sigma}^2_i = \frac{1}{M - k} \sum_{i=k+1}^{M} l_i$$ \hspace{1cm} (9b)

$$\hat{v}_i = u_i, \quad i = 1, \cdots, k$$ \hspace{1cm} (9c)

where $l_1 > l_2 > \cdots > l_M$ and $u_1, \cdots, u_M$ are the eigenvalues and eigenvectors, respectively, of the sample covariance matrix $\Psi$, given by

$$\Psi = \frac{1}{N} \sum_{i=1}^{N} x(i)x^t(i)$$ \hspace{1cm} (10)

and by considering that the complex eigenvectors are orthonormal, the AIC and the MDL criteria could be simplified, respectively, to

$$C_A(k) = L_1(\hat{\theta}(k)) + k(2M - k)$$ \hspace{1cm} (11)

and

$$C_{DL}(k) = L_1(\hat{\theta}(k)) + \frac{1}{2} k(2M - k) \log N$$ \hspace{1cm} (12)

where

$$L_1(\hat{\theta}(k)) = N \log \left[ \frac{1}{M-k} \sum_{i=k+1}^{M} l_i \right] \left[ \prod_{i=k+1}^{M} l_i \right]$$ \hspace{1cm} (13)

The estimated number of signals $\hat{k}$ is determined as the value of $k \in \{0, 1, \cdots, M-1\}$ which minimizes the AIC or the MDL criterion.

In this paper, we suggest that the use of the log likelihood function $\log f(X | \hat{\theta}) = -L_1(\hat{\theta}(k))$ is not the most suitable since the parameter space chosen is too large. We propose that a new likelihood function should be employed. The maximum likelihood estimates of the eigenvalues could be more accurately obtained from this proposed log likelihood function. In Section II, we examine the rationale for such a choice of the log likelihood function and the maximum likelihood estimates of the eigenvalues are then obtained from this log likelihood function. Section III establishes two modified criteria based on this choice of the log likelihood function and the behavior of the modified criteria are examined. Section IV presents some computer simulation results comparing the performance of the modified criteria to those of the AIC and the MDL criterion, and Section V concludes the paper.

II. CHOICE OF THE LOG LIKELIHOOD FUNCTION

The AIC and the MDL criteria given by (7) and (8) both employ the same log likelihood function $-\log f(X | \hat{\theta})$. The criteria use the parameter space which includes the true eigenvalues and the true eigenvectors which span the signal subspace. However, we observe that if a similarity transformation is performed on the covariance matrix of the observed data, the eigenvectors of the resulting matrix would be different from those of the original matrix whereas the eigenvalues would remain the same. But performing a similarity transformation is equivalent to varying the orientation of the array which would not vary the number of signals arriving at the array. Therefore, by the principle of invariance, it is reasonable to assume that the determination of the number of signals arriving at the array is independent of the eigenvectors of the covariance matrix of the observed data. Careful examination of (13) shows that $L_1(\hat{\theta}(k))$ is independent of $v_i, i = 1, \cdots, k$, the eigenvectors spanning the signal subspace and thus confirms our assumption. In other words, the parameters $v_i, i = 1, \cdots, k$, are irrelevant to the decision of the number of signals, and that the true eigenvalues $\lambda_1, \cdots, \lambda_M$ are the only determining parameters. Hence the parameter space can be chosen to exclude the eigenvectors so that

$$\theta(k) = [\lambda_1, \lambda_2, \cdots, \lambda_k, \sigma^2]$$ \hspace{1cm} (14)

As a result, the determination of the number of signals will hinge only on the joint PDF of $\{l_i\}$, the eigenvalues of the sample covariance matrix $\Psi$. Thus, we propose that to obtain the number of signals, we should employ the log likelihood function

$$F = -\log f(l_1, \cdots, l_M | \lambda_1, \cdots, \lambda_M)$$ \hspace{1cm} (15)

To obtain the marginal likelihood function $f(l_1, \cdots, l_M | \lambda_1, \cdots, \lambda_M)$, we review some of the results in multivariate statistical theory.

Suppose the observed complex data $x$ has an $M$-dimensional Gaussian distribution with zero mean and covariance matrix $R_i$. Also let

$$\Phi_i = N \Psi_i$$ \hspace{1cm} (16)

where $N$ is the number of samples of observed data vectors, and $\Psi_i$ is the sample covariance matrix given by (10).
Then the matrix \( \Phi \), has the distribution [23]

\[
f(\Phi) = \frac{1}{\Gamma_M(N) \left( \det R \right)^N} \text{etr} \left( -R^{-1}_{-1} \Phi \right) \left( \det \Phi \right)^{N-M-1} \tag{17}\]

where \( \text{etr} (\cdot) \) denotes \( \exp \left[ \text{tr} (\cdot) \right] \) with \( \text{tr} (\cdot) \) being the trace of a matrix, and

\[
\Gamma_M(N) \approx \pi^{M(M-1)/2} \prod_{i=1}^M \Gamma(N - l + 1). \tag{18}\]

Equation (17) is called the complex Wishart distribution usually denoted by \( W_M(N, R) \), and expresses the joint pdf of the elements \( \Phi_{ij}, i, j = 1, \cdots, M, \) of \( \Phi \). Now, if we obtain the eigenvalues \( l_1, \cdots, l_M \) of the sample covariance matrix \( \Psi \), then by a straightforward extension of the method given in [13], [14] to the case of complex data, the joint pdf of \( l_1, \cdots, l_M \) is given by

\[
f(l_1, \cdots, l_M \mid \lambda_1, \cdots, \lambda_M) = ( \det R )^{-N} F_{0}(R^{-1}_{-1}, L) \cdot \Gamma_M(N) \cdot ( \det \Phi )^{N-M} \prod_{i<j}(l_i - l_j)^2 \tag{19}\]

where \( F_{0}(\cdot, \cdot) \) is the hypergeometric function defined by

\[
F_{0}(R^{-1}_{-1}, L) = \int_{H \in \mathcal{U}(M)} \text{etr} \left[ -NR^{-1}_{-1}HLHL^{*} \right] (H^{*} dH) \tag{20}\]

and \( L \) is the \( M \times M \) diagonal matrix with the diagonal elements being \( l_1, \cdots, l_M \). The matrix \( H \) of (20) is a product of the eigenvector matrix \( V \) of \( R \) and the eigenvector matrix \( U \) of \( \Psi \), and the integration in (20) is performed over the \( M \times M \) unitary matrix group \( \mathcal{U}(M) \). The result of the integration is that the joint pdf of \( l_1, \cdots, l_M \) is averaged over all the eigenvectors in the unitary matrix group, and the parameter space has been reduced to exclude the eigenvectors.

Although (19) is exact, it is difficult to apply since it contains the complicated hypergeometric function \( F_{0} \). To overcome this problem, we can expand this function in terms of \( \lambda_i \) and \( l_i, i = 1, \cdots, M \) for large \( N \) and obtain an asymptotic joint pdf \( f(l_1, \cdots, l_M \mid \lambda_1, \cdots, \lambda_M) \) of (19). Various authors [15], [16] have obtained the asymptotic expansion for the joint pdf of (19) for distinct eigenvalues \( \lambda_1, \cdots, \lambda_M \). However, our signal model here demands that the smallest eigenvalue of \( R \), has multiplicity \( M - K \). The full development of the asymptotic joint pdf \( f(l_1, \cdots, l_M \mid \lambda_1, \cdots, \lambda_M) \) under such a condition of complex signal model is quite lengthy [17], [18]. We outline the procedure in Appendix A. The resulting joint pdf is given by:

\[
f(l_1, \cdots, l_M \mid \lambda_1, \cdots, \lambda_M) = c_1c_2F_1F_2 \tag{21}\]

where

\[
c_1 = \frac{N^{MN-(1/2)(2M-K-1)}}{\prod_{i=1}^M \Gamma(2M - 1)} \tag{22a}\]

\[
c_2 = \frac{\Gamma_M(N-1/2)(2M-K-1)}{\Gamma_M(N) \Gamma_M(M - K)} \tag{22b}\]

\[
F_1 = \exp \left\{ -N \sum_{i=1}^k \frac{l_i}{\lambda_i} + \sum_{i=k+1}^M \frac{l_i}{\lambda_i + \sigma_i^2} \right\} \prod_{i=1}^k \frac{l_i^{N-M}}{\lambda_i^{N-M} \prod_{i=k+1}^M (\lambda_i + \sigma_i^2)} \tag{22c}\]

\[
F_2 = \prod_{i<j} (l_i - l_j)^2 \tag{22d}\]

with

\[
\alpha_i = \lambda_i / \left\{ (\lambda_i - \lambda_j) (l_i - l_j) \right\} \tag{23a}\]

\[
\beta_i = \lambda_i / \left\{ (\lambda_i - \sigma_i^2) (l_i - l_j) \right\} \tag{23b}\]

\( \Gamma_M(M - K) \) is defined in (18).

In deriving (21) we have assumed that the eigenvalues of \( R \) are such that

\[
\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K > \lambda_{K+1} = \cdots = \lambda_M = \sigma^2. \tag{24}\]

To find the estimates of the eigenvalue \( \lambda_i \), we evaluate the values of \( \lambda_i \) and \( \sigma^2 \) which maximize (21). These estimates can be shown (Appendix B) to be given by the simultaneous equations:

\[
\hat{\lambda}_m = l_m - \frac{1}{N} \sum_{i=1}^{M-k} \hat{\lambda}_i - \frac{M-k}{N} \hat{\lambda}_m \hat{\sigma}^2 \tag{25}\]

\[
\hat{\sigma}^2 = \frac{1}{M-k} \sum_{i=1}^{M-k} l_i + \frac{1}{N} \sum_{i=1}^{K} \hat{\lambda}_i \hat{\sigma}^2 \tag{26}\]

Equations (25) and (26) form a set of nonlinear equations which can be solved by Newton's method or its modified version. This issue will be addressed in Appendix C. Comparing the estimates in (9a) and (9b) to the estimates of (25) and (26), it can be observed that (9a) uses only the information of the \( m \)th sample eigenvalue to estimate the \( m \)th signal eigenvalue, and (9b) uses only the \( (M - k) \) noise sample eigenvalues to estimate the noise power. However, (25) and (26) utilize the relationship with other eigenvalues as well, the adjacent ones having the most effect. These extra terms constitute the correction terms in the evaluation of the eigenvalues.

To examine the improvement of accuracy of the eigenvalue estimates, we test the two procedures of estimation as illustrated by the following example:

A linear array with eight sensors \( (M = 8) \) equally spaced at a distance of half the signal wavelength receives three equal strength signals \( (K = 3) \) arriving at \( 10^\circ, 20^\circ, \)
30°, respectively, from the normal to the array. The signal-to-noise ratio, defined as the power of each signal to the spatially white Gaussian noise power \( \sigma_i^2 \), is \(-5\) dB.

1) The information of this signal model is fed into the corresponding covariance matrix \( R \) of (4). The eigenvalues of the covariance matrix are then numerically calculated and these constitute the true eigenvalues, \( \lambda_1, \cdots, \lambda_M \) of the covariance matrix of the signal model.

2) With the above SNR of \(-5\) dB, we now simulate the signal and noise environment of (1), and the sample covariance matrix \( \Psi \) of (10) after 100 snapshots (\( N = 100 \)) is formed. The eigenvalues \( \lambda_1, \cdots, \lambda_M \) of \( \Psi \), are numerically evaluated, and the estimated sample signal eigenvalues \( \tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\lambda}_3 \), and the sample noise power \( \tilde{\sigma}_2 \) are then calculated by (9a) and (9b), respectively.

3) Using the sample covariance matrix \( \Psi \), above, the more refined estimates of the signal eigenvalues \( \hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3 \), and the sample noise power estimate \( \hat{\sigma}_2 \) are evaluated according to (25) and (26), respectively.

Procedures 1) and 2) above are repeated 50 times so that an average of the eigenvalue estimates are obtained. The results of these estimates are shown in Tables I–III.

The values shown in Tables I–III reveal that the refined estimates, while not having much improvement on the variance, have much smaller bias than the sample eigenvalue estimates. Many other examples have been carried out, and similar observations persist. Thus, we can conclude that the new maximum likelihood estimates of the eigenvalues and noise power using the new log likelihood function indeed provides us with more accurate results.

III. CRITERIA FOR DETERMINATION OF THE NUMBER OF SIGNALS USING NEW LOG LIKELIHOOD FUNCTIONS

To exploit the use of the new log likelihood function, we establish modified criteria for the determination of the number of signals by employing the new log likelihood function and utilizing the same penalty functions as suggested by the AIC and the MDL criterion in (7) and (8), respectively. Since the parameter space in our likelihood function has been reduced in size, the number of free parameters is equal to \( k + 1 \) (there being \( k \) distinct eigenvalues \( \lambda_1, \cdots, \lambda_k \), and one multiple eigenvalue \( \sigma_2^2 \)). Hence our criteria to determine the number of signals are to choose \( k \) so that either one of the following functions is minimized:

\[
C_1(k) = -\log f(t_1, \cdots, t_M \mid \tilde{\lambda}_1, \cdots, \tilde{\lambda}_k, \tilde{\sigma}_2^2) + k 
\]

and

\[
C_2(k) = -\log f(t_1, \cdots, t_M \mid \hat{\lambda}_1, \cdots, \hat{\lambda}_k, \hat{\sigma}_2^2) + \frac{1}{2}k \log N
\]

where \( f(t_1, \cdots, t_M \mid \tilde{\lambda}_1, \cdots, \tilde{\lambda}_k, \tilde{\sigma}_2^2) \) is given by (21), \( \hat{\lambda}_k \) and \( \hat{\sigma}_2 \) are the maximum likelihood estimates given by (25) and (26), respectively. Substituting the values of \( \hat{\lambda}_i, i = 1, \cdots, k \), and \( \hat{\sigma}_2^2 \) into (21), and ignoring terms that do not involve \( \hat{\lambda}_1, \cdots, \hat{\lambda}_k, \hat{\sigma}_2^2 \), or \( k \) the two criteria in (27) and (28) can be rewritten as

\[
C_1(k) = \Lambda_a + \Lambda_e + p_a + p_c + k
\]

\[
C_2(k) = \Lambda_a + \Lambda_e + p_a + p_c + \frac{1}{2}k \log N
\]

where

\[
\Lambda_a = N \left( \sum_{i=1}^{k} \frac{t_i}{\hat{\lambda}_i} + \sum_{i=k+1}^{M} \frac{t_i}{\hat{\sigma}_2^2} \right) + (N - M + 1) \]

\[
\cdot \sum_{i=1}^{k} \log \tilde{\lambda}_i + (M - k) (N - k) \log \hat{\sigma}_2^2
\]

\[
\Lambda_e = \sum_{i,j=1}^{k} \log (\hat{\lambda}_i - \hat{\lambda}_j) + \sum_{i=k+1}^{M} (M - k) \log (\hat{\lambda}_i - \hat{\sigma}_2^2)
\]

\[
- \sum_{i<j}^{M} \log (t_i - t_j)
\]

\[
p_a = \frac{1}{2}k(2M - k - 1) \log N
\]

\[
p_c = -\sum_{i=M-k+1}^{M} \log \Gamma(i).
\]

It can be seen that both criteria \( C_1 \) and \( C_2 \) consist of two parts. The part \( (\Lambda_a + \Lambda_e) \) which depends on the sample and estimated eigenvalues is the likelihood function and is the same for both criteria, whereas the part \( p_a + p_c \), together with the original respective penalty function is
dependent only on \(k\), \(M\), and \(N\) and provides the penalty function. It is interesting to observe that in establishing \(C_1\) and \(C_2\) in (27) and (28) we utilize the same penalty functions suggested by the original AIC and MDL criterion, respectively, involving the number of free parameters. However, on substituting the values of \(\hat{\lambda}_i\) and \(\hat{\sigma}_i^2\), the function \(-\log f\) not only provides us with the likelihood functions \(\Lambda_{\hat{\theta}}\) and \(\Lambda_{\hat{\mu}}\), but also provides us with the part \((p_i + p_{i+1})\) which constitute portions of the penalty functions.

In order to gain some insight into the difference between the two modified criteria and the original criteria of AIC and MDL, we examine the asymptotic behavior of the modified criteria by ignoring the \(O(N^{-1})\) terms in (25) and (26) so that

\[
\hat{\lambda}_i \rightarrow \lambda_i = l_i, \quad i = 1, \cdots, k.
\]

and

\[
\hat{\sigma}_i^2 \rightarrow \sigma_i^2 = \frac{1}{M - k} \sum_{i+k+1}^M l_i.
\]

Substituting these asymptotically approximate estimates of the eigenvalues into \(\Lambda_{\hat{\theta}}\) and \(\Lambda_{\hat{\mu}}\), we obtain

\[
\Lambda = \Lambda_{\hat{\theta}} + \Lambda_{\hat{\mu}} = N M + N \sum_{i=1}^k \log l_i + (M - k) N \log \hat{\sigma}_i^2
\]

\[
+ \left( \sum_{i=1}^k \log (l_i - l_j) + \sum_{i=1}^k (M - k) \log (l_i - \hat{\sigma}_i^2) \right)
\]

\[
- \left( \sum_{i+j+1}^M \log (l_i - l_j) \right)
\]

\[
+ \left( \sum_{i=1}^k \log (l_i - l_j) + \sum_{i=1}^k (M - k) \log (l_i - \hat{\sigma}_i^2) \right)
\]

\[
- \left( \sum_{i+j+1}^M \log (l_i - l_j) \right)
\]

\[
= L_1 + L_2 + L_3 + \text{(terms independent of } k) \quad (35)
\]

where

\[
L_1 = N \log \left( \frac{1}{M - k} \sum_{i=1}^M l_i \right) - \frac{M - k}{M} \log M
\]

\[
L_2 = \sum_{i=1}^k \log (l_i - l_j)^2
\]

\[
L_3 = \sum_{i=1}^k \log (l_i - \hat{\sigma}_i^2)^{M-k}. \quad (36c)
\]

Ignoring the terms independent of \(k\) in (35), we can see that the first term on the right-hand side (RHS) of (13) is identical to the log likelihood function \(L_1(\hat{\theta}(k))\) in (13). In addition to this log likelihood function, the modified criteria incorporate two extra terms which are expressed as \(L_2\) and \(L_3\) in (35).

Given a set of sample values such that

\[
W = \{l_m; l_m > l_{m-1}, m = 1, \cdots, M - 1\} \quad (37)
\]

the detection of the number of signals is in fact a partition of this set into two disjoint subsets \(W_s\) and \(W_n\), representing the signals and noise, respectively, where

\[
W_s = \{l_1, \cdots, l_k\} \quad (38a)
\]

and

\[
W_n = \{l_{k+1}, \cdots, l_M\}. \quad (38b)
\]

To evaluate the appropriateness of such a partition, certain metric has to be imposed to measure the elements in the subsets \(W_s\) and \(W_n\). The first term on RHS of (35) is a monotonic function of the ratio of the arithmetic mean of the noise elements to their geometric mean. This ratio measures the mean of the spread of the elements in the noise subset \(W_n\). The second term is a measure of the distance between the elements in the signal subset \(W_s\). The third term evaluates the interrelationship between the elements of the subsets \(W_s\) and \(W_n\).

We now compare the behavior of the log likelihood function \(\Lambda\) of the modified criteria with that of \(L_1\), the log likelihood function for the AIC and MDL criterion. The two likelihood functions are obtained by computer simulations in which 100 snapshots \((N = 100)\) are taken and there are 3 signals \((K = 3)\) of equal power impinging on a linear array having 8 sensors \((m = 8)\) equally spaced at a distance of half wavelength. The angles of arrival of these signals are, respectively, 15°, 27°, and 36° counterclockwise from the normal of the array axis. SNR is defined as the ratio of the power of one signal to the power of the noise. The comparison of the two log likelihood functions under various SNR are shown in Fig. 1(a)-(d). Fig. 1(a) shows the case when SNR = 20 dB. The two log likelihood functions start from the same point and decrease very rapidly to a value when \(k = 3\). However, while \(L_1\) continues to decrease for \(k > 3\) at a much slower rate, \(\Lambda\) actually increases in value for \(k > 3\) rendering the minimum value to occur at \(k = 3\). An enlargement of this portion of the graph is shown on the right-hand corner of Fig. 1(a). This phenomenon shows the effect of the two criterion terms \(L_2\) and \(L_3\), which, at high SNR, produce a combined line of positive gradient with respect to \(k\), thereby resulting in the desirable minimum point of \(\Lambda\) at the correct number of \(k\). Fig. 1(b) shows the two log likelihood functions at SNR = 3 dB. The starting point at \(k = 0\) is much lower than that in Fig. 1(a), however, the
sharp changes in gradient at $k = 3$ are still retained in both log likelihood functions. Again, as in the case when the SNR = 20 dB, while $L_1$ continues to have a negative gradient, $\Lambda$ exhibits a positive gradient for $k \geq 3$. However, this positive gradient is markedly reduced compared to the previous case of Fig. 1(a). As the SNR is further reduced, this effect of the correction terms is further weakened. In Fig. 1(c) when the SNR = 0 dB, the gradient of $\Lambda$ for $k \geq 3$ has actually become negative although it is less negative than the corresponding portion of $L_1$. Further reduction of the SNR (Fig. 1(d), SNR = -3 dB), renders the effect of the correction terms negligible.

Now, let us turn our attention to the penalty functions in the modified criteria. Adopting the modified log likelihood function in the two criteria results in an additional penalty term ($p_b + p_c$) besides the original penalty function. $p_b$ and $p_c$ are given by (33) and (34), respectively. Fig. 2(a) and (b) show how these new terms affect the penalty function of the modified criteria $C_1$ and $C_2$, respectively. In Fig. 2(a) we add the degree of freedom $k$ to the part $p_b$, forming $p_{el}$ which increases quite steeply as the number of signal increases. The total penalty function for criterion $C_1$ is given by $p_1$, which is the sum of $p_{el}$ and the correction function $p_c$. The correction function $p_c$ reduces the slope of $p_{el}$, giving rise to the final penalty function $p_1$. We also plot the penalty function

$$P = k(2M - k)$$  \hspace{4cm} (39)

of the original AIC expressed in (11) for comparison. It can be observed the $p_1$ increases much more steeply than $P$. In Fig. 2(b), we add the originally chosen penalty function $1/2k \log N$ to the part $p_b$. It is interesting to observe that the resulting function $p_{el2}$ is identical to the penalty function of the original MDL criterion given in (12). The total penalty function for criterion $C_2$ is given by $p_2$, and is the sum of $p_{el2}$ and the correction function $p_c$, which
serves to reduce the slope of \( p_{\phi_2} \). It is known [20]–[22] that the AIC tends to underpenalize whereas the MDL criterion tends to overpenalize. As a result, the AIC yields a relatively large penalty of overestimation (false alarm) of the number of signals even when the SNR is high while the MDL underestimates (misses) the number of signals when the SNR is low. The combined penalty functions \( p_1 \) and \( p_3 \), as can be seen in Fig. 2(a) and (b) behave similarly, and are moderators of overestimation and underestimation.

IV. COMPARISON OF THE PERFORMANCES OF THE CRITERIA

We now examine the performances of the modified criteria using computer simulation results. In our computer simulations, we assume a linear array with eight \( (M = 8) \) equally spaced sensors. We also assume that the distance between the sensors is equal to one half of the wavelength of the incoming signals so that the vector \( \mathbf{a}(\Phi_k) \) in (2) becomes an \( M \times 1 \) complex vector characterized by the unknown parameters \( \Phi_k \) such that

\[
\mathbf{a}(\Phi_k) = [1 \ e^{-j \sin \Phi_k} \ \cdots \ e^{-j(M-1) \sin \Phi_k}]^T
\]

with \( \Phi_k \) being the angle of arrival of the \( k \)th signal measured in a counterclockwise direction from the normal to the axis of the array. We define the signal-to-noise ratio as

\[
\text{SNR} = 10 \log \left( \frac{1}{K} \sum_{k=1}^{K} \frac{\sigma_k^2}{\sigma^2} \right)
\]

where \( \sigma_k^2 \) is the power of the \( k \)th signal and \( \sigma^2 \) is the noise power.

The number of snapshots in all the simulations is \( N = 100 \). For various SNR, all the criteria, viz., \( C_4 \), \( C_{DL} \), \( C_1 \), and \( C_2 \), are applied and the integer value of \( k \) which yields the minimum value of each criterion is taken to be the estimated number of signals \( k \) for that criterion. The same experiment is repeated 300 times for the same SNR and the number of errors for each criterion recorded. The probability of error (approximated by the frequency of error) for various SNR is then plotted for each criterion.

We first examine the performance of the modified criteria \( C_1 \) and \( C_2 \). \( C_1 \) and \( C_2 \) are given by (29) and (30) where \( \Lambda_k \) and \( \Lambda_f \) are the likelihood functions in which exact values (25), (26) of the estimated eigenvalues are used. On the other hand, when the estimated eigenvalues are approximated by the sample eigenvalues (9a), (9b), the log likelihood functions are approximated by (35). Fig. 3(a) and (b) show, respectively, the performances of the criteria \( C_1 \) and \( C_2 \) using the exact and approximated log likelihood functions in each case. In each case, there are two signals arriving at 16° and 25°. It can be observed in both cases that the criterion using the exact values of the estimated eigenvalues yields marginally better performance than the one using approximated estimated eigenvalues. Other signal conditions, such as different angles of arrival and unequal signal strengths, have been tested and the same observation persists. This slight difference in performance indicates that the approximated log likelihood function in (35) behaves rather similarly to the exact log likelihood function, and its use is examining the approximate behavior of the modified criteria in the last section is justified.

We now compare the performance of the modified criteria \( C_1 \) and \( C_2 \) with the AIC \( C_4 \) and the MDL criterion \( C_{DL} \). Here the criteria \( C_1 \) and \( C_2 \) are calculated using the exact value of the estimated eigenvalues in the log likelihood function. Figs. 4(a)–(c) show the cases when two, three, and four equal power signals, respectively, are arriving at the array. In each of the three cases, it is observed that the thresholds for \( C_4 \), \( C_1 \), and \( C_2 \) occur at about the same SNR whereas the threshold for \( C_{DL} \) occurs at a point approximately 2 dB higher in SNR. The thresholds of all criteria occur at higher SNR as the number of signals increases since the probability of error is inversely proportional to \( (M - k) \) [22]. The AIC \( (C_4) \) exhibits a characteristic such that there is an irreducible probability
performance in the sense that the thresholds of all criteria occur at higher SNR. However, the performance of the criteria relative to each other remain essentially the same in that the thresholds for $C_a$, $C_1$, and $C_2$ occur at about the same SNR whereas the threshold for $C_{DL}$ occurs at approximately 2 dB higher. While both $C_a$ and $C_1$ maintain irreducible errors even if the SNR is high (the error associated with $C_1$ being lower than that of $C_a$), neither $C_2$ nor $C_{DL}$ exhibit such error under these extreme conditions.

V. CONCLUSIONS

The basic consideration of an information theoretic criterion to determine the number of signals is the balance between the information gained from the data and the uncertainty introduced in the estimation of the parameters. The log likelihood function of the information-theoretic criterion can be regarded as a measure of information gain. In general, the more parameters are included, the more information may be obtained, and yet higher uncertainty is introduced in the estimation of these parameters. Therefore, a penalty function is incorporated into the criterion as a measure of the uncertainty.

The AIC and the MDL criterion use the pdf of the data as the log likelihood function and jointly estimate the eigenvalues and eigenvectors (in the maximum likelihood sense), and arrive in an expression independent of the eigenvectors. However, the inclusion of the eigenvectors renders the parameter space so large that the information gained may not offset the uncertainty introduced.

In this paper, we suggest an alternative approach. We reason from a physical point of view that the number of signals is completely independent of the orientation of the array of sensors. By considering that the change in the orientation of the array changes the eigenvectors while the eigenvalues remain unchanged, we conclude that to determine the number of signals, the eigenvectors would provide little information, and that the information is concentrated in the sample eigenvalues. Therefore, in this paper, we suggest the use of the marginal pdf of the sample eigenvalues as the log likelihood function which is obtained by integrating over all the possible values of the sample eigenvectors. In this way, the estimation of the eigenvectors which gains little information for the determination of the number of signals is bypassed and an appropriate reduction of the dimensionality of the parameter space which decreases the uncertainty of estimation is achieved. The resulting estimates (in the maximum likelihood sense) of the eigenvalues using this new log likelihood function have been demonstrated to be superior in accuracy to the sample eigenvalues which are used in the AIC and MDL criterion. Two modified criteria $C_1$ and $C_2$ are then established employing this likelihood function and respectively the penalty functions of AIC and MDL. The new log likelihood function not only provides us with the likelihood terms in the criteria, but also provides us with terms which constitute parts of the penalty function. Both the resultant likelihood part and the penalty parts...
contain correction terms. At high SNR, the correction terms in the likelihood part render the likelihood function to behave in such a way that a minimum occurs at the correct value of $k$. The correction terms in the penalty functions, on the other hand, moderate the overestimation of the AIC and the underestimation of the MDL criterion.

Computer simulations show that while $C_2$, $C_1$, and $C_3$ have similar performance at low SNR, being superior to that of the MDL criterion, $C_2$ does not exhibit the characteristic of having irreducible errors as both $C_2$ and $C_1$ at higher SNR. It can thus be concluded the criterion $C_2$, under general conditions, is an attractive criterion for the determination of the number of signals.

APPENDIX A
ASYMPTOTIC EXPANSION OF THE JOINT PDF OF (19)

The main problem here is to find an asymptotic expression for the hypergeometric function $F_0(-R'^*_{ss}, L)$ as given in (20). Without loss of generality, we can assume that $R_{ss}$ is a diagonal matrix such that

$$R_{ss} = \text{diag}(\lambda_1, \cdots, \lambda_k, \sigma^2, \cdots, \sigma^2)$$

$$= \begin{bmatrix} R_{ss} & 0 \\ 0 & \sigma^2 I_{M-k} \end{bmatrix}$$  \hspace{1cm} (A.1)

where

$$R_{ss} = \text{diag}(\lambda_1, \cdots, \lambda_k).$$ \hspace{1cm} (A.2)

$I_{M-k}$ is an $(M-k)$ dimensional identity matrix, and the eigenvalues obeys the relationship given in (24). Furthermore, the Hermitian matrix $\Phi_s$ can be transformed such that

$$\Phi_s = N\Psi_s = NH^*LH$$ \hspace{1cm} (A.3)

where

$$L = \text{diag}(l_1, \cdots, l_M)$$ \hspace{1cm} (A.4)
and \( H \) is a unitary matrix having the eigenvectors \( h_m, m = 1, \ldots, M \) as its columns such that
\[
H = [h_1 \cdots h_k | h_{k+1} \cdots h_M] = [H_1 | H_2].
\]  
(A.5)
Using (A.1), (A.2), (A.4), and (A.5), we can write
\[
\text{tr} \left( NR^{-1}_\alpha H' LH \right) = \text{tr} \left( NR^{-1}_\alpha H'_1 L H_1 \right) + \text{tr} \left( \sigma_i^2 N H'_2 L H_2 \right) = \text{tr} \left[ (R^{-1}_\alpha - \sigma_i^2 I_k) N H'_2 L H_2 \right] + \text{tr} \left( \sigma_i^2 N L \right)
\]  
(A.6)
where we have used the facts that
\[
\text{tr} \left( \sigma_i^2 N H'_2 L H_2 \right) = \text{tr} \left( \sigma_i^2 N L H_2 H'_2 \right)
\]  
(A.7)
and that
\[
H_i H'_i = I - H_i H'_i.
\]  
(A.8)
We now turn our attention to the invariant measure \((H' dH)\). Following the same argument used in the case of real variables [13], it can be shown that
\[
(H' dH) = (H'_i dH_i) (G' dG)
\]  
(A.9)
where \( G \) runs over the unitary matrix group \( \mathcal{U}(M - k) \).

Using (A.6) and (A.9), we can, therefore, express the hypergeometric function \( F_0 \) as
\[
F_0(-R^{-1}_\alpha, L) = \int_{H \in \mathcal{U}(M - k)} \text{etr} \left[ -NR^{-1}_\alpha H' LH \right] (H' dH)
\]  
\[
= \exp \left( -\frac{N}{\alpha^2} \sum_{m=1}^{M} l_m \right) \int_{H \in \mathcal{U}(M - k)} \left[ \text{etr} \left[ -R^{-1}_\alpha - \sigma_i^2 I_k \right] N H'_i L H_2 \right] \cdot (G' dG) (H'_i dH_i)
\]
(A.10)
where \( \mathcal{U}(M - k) \), called the Stiefel manifold, is the manifold spanned by the column vectors of \( H_i \). It can be shown [16] that
\[
\int_{G \in \mathcal{U}(M - k)} (G' dG) = \frac{\pi^{(M - k)(M - k - 1)}}{\Gamma_{M - k}(M - k)}
\]  
(A.11)
which represents the volume of \( \mathcal{U}(M - k) \). Hence, substituting (A.11) into (A.10), we have
\[
F_0(-R^{-1}_\alpha, L) = \frac{\pi^{(M - k)(M - k - 1)}}{\Gamma_{M - k}(M - k)} \exp \left( -\frac{N}{\alpha^2} \sum_{m=1}^{M} l_m \right) \int_{H \in \mathcal{U}(M - k)} \text{etr} \left[ (\sigma_i^2 I_k - R^{-1}_\alpha) N H'_i L H_2 \right] \cdot (H'_i dH_i)
\]  
(A.12)
in (A.12) attains its maximum value of
\[
\exp \left( \frac{1}{\sigma_i^2} \sum_{i=1}^{k} l_i - \sum_{i=1}^{k} \frac{l_i}{\lambda_i} \right)
\]  
if and only if \( H_i \), takes the following form:
\[
H_i = H_{i0} = \begin{bmatrix} e^{\theta_i} & 0 \\ 0 & e^{\theta_i} \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}
\]  
(A.13)
where \( 0 \leq \theta_i < 2\pi, i = 1, \ldots, k \).

Since the integrand in (A.12) is evaluated to the \( N \)th power of the quantity \( \text{etr} \left[ (\sigma_i^2 I_k - R^{-1}_\alpha) N H'_i L H_2 \right] \), as \( N \to \infty \), the integral outside the neighborhood of \( H_{i0} \) as given in (A.13) is negligible and the integration may be performed concentrating over the neighborhood of \( H_{i0} \). Using the same technique as described in [13], we can evaluate the integral in (A.12) and express the hypergeometric function for the hypergeometric function is given by
\[
F_0(-R^{-1}_\alpha, L) = \frac{\pi^{(M - k)(M - k - 1)}}{\Gamma_{M - k}(M - k)} \exp \left( -\frac{N}{\alpha^2} \sum_{m=1}^{M} l_m \right) \cdot \exp \left[ -\frac{N}{\alpha^2} \sum_{i=1}^{k} l_i - \sum_{i=1}^{k} \frac{l_i}{\lambda_i} \right] \cdot \prod_{i<j} \left( \frac{\pi \alpha_{ij}}{N} \right) \left( \prod_{i=1}^{k} \prod_{i=1}^{k} \prod_{j=k+1}^{M} \left( \frac{\pi \beta_{ij}}{N} \right) \right)
\]  
(A.14)
where \( \alpha_{ij} \) and \( \beta_{ij} \) are given by (23a) and (23b). Substituting (A.14) into (19) and simplifying, the joint pdf \( f(l_1, \ldots, l_m | \lambda_1, \ldots, \lambda_M) \) in (21) follows.

**APPENDIX B**

**THE MAXIMUM LIKELIHOOD ESTIMATES OF EIGENVALUES**

The joint pdf \( f(l_1, \ldots, l_m | \lambda_1, \ldots, \lambda_M) \) is given by (21). Taking the logarithm and ignoring the terms which are independent of the parameters \( \lambda_i, i = 1, \ldots, k \) and \( \sigma^2 \), we obtain the log likelihood function
\[
F = -\log f(l_1, \ldots, l_m | \lambda_1, \ldots, \lambda_k, \sigma^2)
\]  
\[
= N \left[ \sum_{i=1}^{k} \frac{l_i}{\lambda_i} + \sum_{m=1}^{M} \frac{l_m}{\sigma^2} \right] + N(M - k) \log \sigma^2
\]
\[
= \sum_{i=1}^{k} \log \lambda_i - \sum_{i<j} \log \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)} - (M - k) \sum_{i=1}^{k} \log \frac{\lambda_i \sigma^2}{(\lambda_i - \sigma^2)}.
\]  
(B.1)
To obtain the maximum likelihood conditions, we differentiate \( F \) in (B.1) with respect to \( \lambda_m, m = 1, \ldots, k \), and with respect to \( \sigma^2 \), and equate these differentials to
zero. The differentiation of the RHS of (B.1) is straightforward, giving
\[
\frac{\partial F}{\partial \lambda_m} = \frac{N m}{\lambda_m^2} + \frac{N}{\lambda_m} - \sum_{i=1}^{k} \frac{\lambda_i}{\lambda_m (\lambda_i - \lambda_m)} + \frac{(M - k) \sigma_i^2}{\lambda_m (\lambda_m - \sigma_i^2)} \quad m = 1, \ldots, k \tag{B.2}
\]
\[
\frac{\partial F}{\partial \sigma_i^2} = - \frac{N}{\sigma_i^2} \sum_{i=1}^{k} \frac{\lambda_i}{\lambda_m (\lambda_i - \sigma_i^2)} - \frac{(M - k) \lambda_i}{\lambda_m (\lambda_m - \sigma_i^2)} \tag{B.3}
\]
Equating (B.2) and (B.3) to zero and solving for \(\lambda_m\) and \(\sigma_i^2\) (25) follow.

**Appendix C**

**Numerical Evaluation of the Estimates of the Eigenvalues**

Let \(D_m\) \((m = 1, \ldots, k)\) and \(D_s\) represent the difference of the left-hand and right-hand terms in (25) and (26), respectively, i.e.,
\[
D_m = \frac{\lambda_m}{N} \sum_{i=1}^{k} \frac{\lambda_i}{(\hat{\lambda}_m - \hat{\lambda}_i)} + \frac{M - k}{N(\hat{\lambda}_m - \hat{\sigma}_s^2)} = l_m \quad m = 1, \ldots, k \tag{C.1}
\]
and
\[
D_s = \frac{1}{M - k} \sum_{i=1}^{M} l_i - \frac{1}{N} \sum_{i=1}^{k} \frac{\hat{\lambda}_i \hat{\sigma}_s^2}{(\hat{\lambda}_i - \hat{\sigma}_s^2)} \tag{C.2}
\]
We define a \((k + 1) \times 1\) partial derivative vector operator
\[
\nabla_\lambda \doteq \begin{bmatrix}
\frac{\partial}{\partial \lambda_1} \\
\frac{\partial}{\partial \lambda_2} \\
\vdots \\
\frac{\partial}{\partial \lambda_k} \\
\frac{\partial}{\partial \sigma_s^2}
\end{bmatrix}
\tag{C.3}
\]
and a \(1 \times (k + 1)\) row vector \(D\)
\[
D \doteq [D_1 \ D_2 \ \cdots \ D_k \ D_s] \tag{C.4}
\]
so that the operator \(\nabla_\lambda\) when applied to \(D\) results in the \((k + 1) \times (k + 1)\) matrix
\[
\nabla_\lambda D = \begin{bmatrix}
\frac{\partial D_1}{\partial \lambda_1} & \cdots & \frac{\partial D_1}{\partial \lambda_k} & \frac{\partial D_1}{\partial \sigma_s^2} \\
\frac{\partial D_2}{\partial \lambda_1} & \cdots & \frac{\partial D_2}{\partial \lambda_k} & \frac{\partial D_2}{\partial \sigma_s^2} \\
\vdots & \cdots & \vdots & \vdots \\
\frac{\partial D_k}{\partial \lambda_1} & \cdots & \frac{\partial D_k}{\partial \lambda_k} & \frac{\partial D_k}{\partial \sigma_s^2} \\
\end{bmatrix}
\tag{C.5}
\]
The derivatives of \(D_m\) and \(D_s\) with respect to \(\lambda_m\), \(n = 1, \ldots, k\) and with respect to \(\sigma_s^2\) can be easily obtained such that
\[
\frac{\partial D_m}{\partial \lambda_n} = \begin{cases}
1 - \frac{1}{N} \sum_{i=1}^{k} \frac{\hat{\lambda}_i^2}{(\hat{\lambda}_m - \hat{\lambda}_i)^2} - \frac{M - k}{N(\hat{\lambda}_m - \hat{\sigma}_s^2)^2}, & m = n, n = 1, \ldots, k \\
\frac{\hat{\lambda}_m^2}{N(\hat{\lambda}_m - \hat{\lambda}_n)^2}, & m \neq n, m = 1, \ldots, k
\end{cases}
\]
\[
\frac{\partial D_s}{\partial \lambda_m} = \frac{1}{N(\hat{\lambda}_m - \hat{\sigma}_s^2)^2}, \quad m = 1, \ldots, k
\]
\[
\frac{\partial D_s}{\partial \sigma_s^2} = \frac{1}{N} \sum_{i=1}^{k} \frac{\hat{\lambda}_i^2}{(\hat{\lambda}_i - \hat{\sigma}_s^2)^2}, \quad n = 1, \ldots, k
\]
Denoting the \(j\)th iteration by the superscript \((j)\), we use the following recursive formula to arrive at the estimates of the eigenvalues:
\[
\hat{\lambda}^{(j+1)} = \hat{\lambda}^{(j)} - \gamma \nabla_\lambda D(\hat{\lambda}^{(j)})^{-1} D(\hat{\lambda}^{(j)}) \tag{C.7}
\]
where
\[
\hat{\lambda}^{(j)} \doteq [\lambda_1^{(j)} \ \cdots \ \lambda_k^{(j)} \ \sigma_s^{2(j)}]^T \tag{C.8}
\]
and
\[
\hat{\lambda}^{(0)} = \begin{bmatrix}
l_1 & \cdots & l_k & \frac{1}{M - k} \sum_{i=k+1}^{M} l_i
\end{bmatrix}^T \tag{C.9}
\]
Equation (C.7) is referred to as the Newton’s method.

The coefficient \(\gamma\) is a damping coefficient used to control the convergence of the iteration. At the beginning of each iteration, \(\gamma\) is set to unity. If this value does not reduce \(\|D\|\), \(\gamma\) is reduced by \(1/2\) and the new value of \(\lambda_1^{(j+1)}\) recalculated. If again, \(\|D\|\) is not reduced, set \(\gamma = 1/4\) and continue to calculate \(\lambda_1^{(j+1)}\). The norm \(\|D\|\) in our study has been chosen to be the Euclidean norm.
i.e.,

$$|D^2| = D_1^2 + \cdots + D_k^2 + d^2. \quad (C.10)$$

It is found that usually it takes 3 to 4 iterations of Newton's algorithm to converge.

The Newton's algorithm requires the inversion of the matrix $\nabla D$ for every iteration. The computational burden can be reduced by using the gradient method when the number of snapshots is large. When $N \to \infty$, the first equation in (C.6) becomes

$$\frac{\partial D}{\partial \delta n} = \begin{cases} 1, & m = n \\ 0, & m \neq n \end{cases} \quad m, n = 1, \ldots, k \quad (C.11)$$

and together with the other three equations in (C.6), we have

$$\nabla D = I \quad (C.12)$$

where $I$ is a $(k + 1) \times (k + 1)$ identity matrix. In this case, (C.7) can be reduced to

$$\hat{\lambda}^{(j+1)} = \hat{\lambda}^{(j)} - \gamma D(\hat{\lambda}^{(j)}). \quad (C.13)$$

In our studies, we found that if $N$ is moderately large, say, $N > 100$, (C.7) and (C.13) yield very similar results.

REFERENCES


James P. Reilly (S'76-M'80) received the B.A.Sc. degree in electrical engineering from the University of Waterloo, Canada, in 1973, and the M.Eng. and Ph.D. degrees from McMaster University in 1976 and 1981, respectively. From 1980 to 1983 he was employed as a member of the scientific staff at Bell-Northern Research, Ottawa, Canada, primarily in the area of digital radio systems design. Since September 1985 he has been an Associate Professor in the Department of Electrical and Computer Engineering at McMaster University. His research interests are in the area of array and signal processing and their fast implementations.

P. C. Yip received the B.Sc. degree from Memorial University of Newfoundland in 1965, and the Ph.D. degree from McMaster University, Hamilton, Ont., Canada, in 1970, both in theoretical physics. He is currently a Professor in the Department of Mathematics and Statistics at McMaster University. His main research interests are digital signal processing, transform algorithms, and time series modeling.