Performance Comparison of Subspace Rotation and MUSIC Methods for Direction Estimation

Petre Stoica and Arye Nehorai, Senior Member, IEEE

Abstract—This paper is concerned with studying the statistical performance of subspace rotation (SR) methods (such as the Toeplitz approximation method (TAM) and a variant of ESPRIT) for direction estimation using arrays composed of matched sensor doublers. The distributional properties of these methods are established and a compact explicit formula for the covariance matrix of their estimation errors is provided. Next, using this formula and a similar formula for the MUSIC covariance matrix it is shown that the SR methods are statistically less efficient than MUSIC, at least for a sufficiently large number of snapshots. The difference in statistical performance between the commonly used SR method and MUSIC may be substantial if the number of sensors in the array is large. An optimally weighted SR method which may approach the MUSIC level of statistical performance for one direction parameter (specified by the user) is introduced in this paper.

I. INTRODUCTION

Consider the problem of estimating the directions of $n$ far-field narrow-band sources using a planar array of $m$ sensors that possess a certain displacement invariance. Specifically, it is assumed that the array contains $m$ pairwise matched codirectional sensor doublers (whose locations and sensitivity patterns are otherwise unknown), with $m \in [n, m-1]$. For a uniform linear array of $m$ identical sensors this assumption is satisfied for $m \leq m-1$. For an array composed of $m$ matched sensor doublers whose elements are translationally separated by a (known) constant displacement vector, the assumption above is also satisfied and $m = m/2$.

The aforementioned problem (first mentioned in [21], [22], and [25]) can be reduced to estimating the parameter vector

$$\theta = [\theta_1 \cdots \theta_d]^T$$

in the following model (see, e.g., [7], [9], [10], [21], [22])

$$y(t) = Ax(t) + e(t) \quad t = 1, 2, \cdots, N$$

where $y(t) \in \mathbb{C}^{m \times 1}$ is the array output vector, $x(t) \in \mathbb{C}^{m \times 1}$ is the source signal vector, $e(t) \in \mathbb{C}^{m \times 1}$ is an additive noise, $N$ denotes the number of available snapshots, and the matrix $A \in \mathbb{C}^{m \times m}$ has the following (essential) property:

$$A \doteq \begin{bmatrix} 0 & I_m \end{bmatrix} A \begin{bmatrix} 0 & I_m \end{bmatrix} \Psi \doteq A, \Psi.$$

In (1.3), $I_m$ denotes the $m \times m$ identity matrix, and

$$\Psi = \begin{bmatrix} e^{j\theta} & 0 \\ 0 & e^{j\theta} \end{bmatrix}. \quad (1.4)$$

If $m > \lfloor m/2 \rfloor$, the integer part of $m/2$, the two subarrays corresponding to $A_1$ and $A_2$ are overlapping, and they are non-overlapping if $m \leq \lfloor m/2 \rfloor$. This distinction turns out to be of importance for some later developments.

Remark: Since we will not impose any conditions (other than (1.3)) on the array geometry, there is no restriction to assume that the two subarrays are formed of the first, respectively, the last $m$ sensors. Indeed, this assumption can always be met by appropriately ordering the sensors in the global array (before writing the model (1.2)). A similar observation has been made in [4].

The basic assumptions on the model (1.1)-(1.4) will be introduced in Section II. Note that the number $n$ of sources is assumed to be known in what follows. If $n$ is unknown then it can be estimated from the data by one of the several available techniques (see, e.g., [18], [19]).

The SR direction estimation methods are based on (1.3), as described in Section II. Note that this equation is a rotation transformation of the row subspace of $A_1$ to the row subspace of $A_2$. This observation justifies the name of “subspace rotation” given to the methods based on (1.3). As explained in Section II, this class of methods includes the TAM and ESPRIT-type techniques discussed, for example, in [2]-[10] and [21]-[23]. Section II also contains a brief presentation of the MUSIC method ([13], [14]) and its statistical properties established in [15].

The derivation of the Toeplitz approximation method (TAM) in Section II relies on pure algebraic considerations (similar derivations can be found in [6] and [22]). This derivation is not limited to uniform linear arrays as is the presentation of TAM in [2]-[4], which is based on state-space realizations. Note that an estimation method very related to TAM has been developed independently in [1] for the purpose of state-space modeling of time series.

Concerning the estimation of signal parameters via rotational invariance techniques (ESPRIT), the variant of this method considered in Section II is shown to coincide with TAM (see also [4]). It should be noted that there are several implementations of ESPRIT currently in use (see, e.g., [5], [7]-[12], [21], [22], and [25]). Some of these variants (for example, those which estimate $\theta$ by solving a generalized eigenvalue problem) and the variant considered herein may have different statistical properties; and a separate analysis of them is thus required.

Section III establishes the asymptotic (for $N \gg 0$) distribution of SR estimates of $\theta$ and derives an explicit expression

Manuscript received April 13, 1989; revised March 16, 1990. The work of A. Nehorai was supported by the Air Force Office of Scientific Research under Grant AFOSR-90-0164.

P. Stoica is with the Department of Automatic Control, the Polytechnic Institute of Bucharest, Romania.

A. Nehorai is with the Department of Electrical Engineering, Yale University, New Haven, CT 06520.

IEEE Log Number 904157.
for the covariance matrix of the corresponding estimation errors. A related study has been reported recently in [4]. However, the analysis in [4] is limited to uniform linear arrays and is incomplete (as explained in some detail in Appendix A).

Section IV proves that the asymptotic variance of the SR estimate is greater than the MUSIC asymptotic variance. Furthermore, it is shown that the difference between SR and MUSIC variances may be considerable for large m.

Finally, Section V introduces an optimal SR method which makes use of an initial (consistent) estimate of $\theta_i$ to improve the estimation accuracy of the kth direction parameter. It is shown that the variance of $\hat{\theta}_i$ obtained by this novel SR method may be comparable to the MUSIC variance, whereas its computational burden is still less complex that than associated with MUSIC.

II. SR and MUSIC METHODS

First we introduce the basic assumptions on the data model (1.1)-(1.4) that will be considered to hold throughout the paper.

A. Basic Assumptions

A1: The rank of $A_1$ and $A_2$ is equal to n (this implies $\Re h \geq n$ and $\Im \theta_i \neq \Im \theta_j$ for $i \neq j$).

A2: The source signal vector $x(t)$ is a stationary zero-mean random process with the following covariance properties:

$$E x(t) x^*(s) = P \delta_{t,s}$$

$$E x(t) x^T(s) = 0 \quad \text{for all } t, s.$$

Furthermore, the covariance matrix $P$ is positive definite ($P > 0$), and $x(t)$ and $e(s)$ are uncorrelated for all $t$ and $s$:

$$E x(t) e^*(s) = E x(t) e^T(s) = 0.$$

Note that the superscripts $T$ and $*$ denote transpose and conjugate transpose, respectively; and $\delta_{t,s}$ denotes the Kronecker delta.

A3: The noise $e(t)$ is a stationary zero-mean random process that is both temporally and spatially white:

$$E e(t) e^*(s) = \sigma^2 \delta_{t,s}$$

$$E e(t) e^T(s) = 0 \quad \text{for all } t, s.$$

The above assumptions are commonly made in studies similar to the present one. For some comments on them see [15].

Next we list the notation most frequently used in the following and give some preliminary mathematical results.

B. Basic Notation and Preliminary Results

$$\Re X = \text{the real part of } X,$$

$$\Im X = \text{the imaginary part of } X,$$

$$X_{ij} = \text{the } i, j \text{ element of } X,$$

$$X_{i}^r = \text{the } k \text{ row of } X,$$

$$X_{i}^c = \text{the } k \text{ column of } X,$$

$$X^{1/2} = \text{a square root of a positive definite matrix } X,$$

$$A \succeq B = \text{the difference matrix } A - B \text{ is (Hermitian) positive semidefinite},$$

$$\hat{\theta}_k = \text{estimate of } \theta_k,$$

$$\var \left( \hat{\theta}_k \right) = \text{(asymptotic) variance of } \hat{\theta}_k,$$

$$R = E y(t) y^*(t) = APA^* + \sigma I,$$

$$\lambda_1 > \lambda_2 > \cdots > \lambda_p,$$

$$\Delta = \text{eigenvalues of } R \text{ in decreasing order, }$$

$$\Lambda = \text{diag} \left( \lambda_1, \cdots, \lambda_p \right),$$

$$\bar{\Delta} = \Lambda - \sigma I,$$

$$\{ s_1, \cdots, s_m \} = \text{orthonormal eigenvectors of } R \text{ associated with } \{ \lambda_1, \cdots, \lambda_m \},$$

$$\bar{\Delta} = \text{eigenvalues of } R \text{ in decreasing order},$$

$$\bar{\Delta} = \text{eigenvalues of } R \text{ in decreasing order},$$

$$\hat{\Delta} = \text{eigenvalues of } R \text{ in decreasing order},$$

$$\hat{G} = \text{eigenvalues of } R \text{ in decreasing order},$$

$$\hat{G} = \text{eigenvalues of } R \text{ in decreasing order},$$

$$\hat{\lambda}_i = \text{the } i \text{th largest eigenvalue of } R,$$

$$\hat{\lambda}_i = \text{the } i \text{th largest eigenvalue of } R,$$

$$\left( \hat{\lambda}_1, \cdots, \hat{\lambda}_m \right) = \text{the } m \text{ largest eigenvalues of } R,$$

$$\left( \hat{\lambda}_1, \cdots, \hat{\lambda}_m \right) = \text{the } m \text{ largest eigenvalues of } R,$$

$$\left( \hat{\lambda}_1, \cdots, \hat{\lambda}_m \right) = \text{the } m \text{ largest eigenvalues of } R,$$

$$\left( \hat{\lambda}_1, \cdots, \hat{\lambda}_m \right) = \text{the } m \text{ largest eigenvalues of } R,$$

$$\left( \hat{\lambda}_1, \cdots, \hat{\lambda}_m \right) = \text{the } m \text{ largest eigenvalues of } R,$$

$$\left( \hat{\lambda}_1, \cdots, \hat{\lambda}_m \right) = \text{the } m \text{ largest eigenvalues of } R,$$

$$\left( \hat{\lambda}_1, \cdots, \hat{\lambda}_m \right) = \text{the } m \text{ largest eigenvalues of } R,$$

$$\left( \hat{\lambda}_1, \cdots, \hat{\lambda}_m \right) = \text{the } m \text{ largest eigenvalues of } R,$$

$$\left( \hat{\lambda}_1, \cdots, \hat{\lambda}_m \right) = \text{the } m \text{ largest eigenvalues of } R,$$

$$\left( \hat{\lambda}_1, \cdots, \hat{\lambda}_m \right) = \text{the } m \text{ largest eigenvalues of } R,$$

$$\left( \hat{\lambda}_1, \cdots, \hat{\lambda}_m \right) = \text{the } m \text{ largest eigenvalues of } R,$$

$$\left( \hat{\lambda}_1, \cdots, \hat{\lambda}_m \right) = \text{the } m \text{ largest eigenvalues of } R,$$

$$\left( \hat{\lambda}_1, \cdots, \hat{\lambda}_m \right) = \text{the } m \text{ largest eigenvalues of } R,$$
Step 3: Estimate \( \{ \theta_i \} \) as the angular positions of the eigenvalues of \( \Phi \).

It is sometimes claimed that the matrix \( H \) in the SR procedure should be chosen with care. For example, TAM uses \( H = \tilde{A}^{1/2} \), whereas a possible variant of ESPRIT corresponds to the above procedure with \( H = I \) (see, e.g., [2]-[6]). For the moment, let \( \Phi_H \) denote the matrix (2.4) corresponding to a certain \( H \), and let \( \Phi \) be \( \Phi_H \) with \( H = I \). Then, it is easy to see that the estimated matrices \( \hat{\Phi}_H \) and \( \hat{\Phi} \) are related by a similarity transformation
\[
\hat{\Phi}_H = H^{-1} \hat{\Phi} H
\]
and therefore that they have the same eigenvalues. This means that the theoretical performance of the SR method does not depend on \( H \). (This simple observation has also been made recently in [44].)

According to the above discussion we will consider \( H = I \) in the following. Then, from (2.4) we get
\[
\hat{\Phi} = C^{-1} \Psi C.
\]
(2.7)

For later use, denote
\[
a_k^* = [C_k^*]^{(r)}
\]
(2.8)
\[
\beta_k = [C_k^{-1}]^{(c)}
\]
(2.9)

and observe that \( a_k^* \) and \( \beta_k \) are, respectively, left and right eigenvectors of \( \Phi \) associated with the eigenvalue \( e^{\lambda_k} \), and also that \( a_k^* \beta_k = 1 \).

The SR procedure outlined above can be extended in several ways. Let \( \tilde{W} \) be an \( m \times \tilde{m} \) positive definite matrix which could be data dependent. The limit of \( \tilde{W} \) when \( N \) tends to infinity is assumed to exist and is denoted by \( W \). We can replace step 2 of the SR procedure by the following more general calculation:
\[
\hat{\Phi} = (S_H \tilde{W} S_H)^{-1} (S_H \tilde{W} S_H).
\]
(2.10)

As will be shown in Section V, the weighting matrix \( \tilde{W} \) in (2.10) can be chosen so as to maximize the estimation accuracy of an element \( \theta_i \) of \( \tilde{\theta} \), at the user's choice. However, it is not possible to optimize the accuracy of the whole estimated parameter vector by selecting \( \tilde{W} \). The latter operation would become possible by considering a more general estimate of \( \Phi \) than (2.10). To be specific, let
\[
\varphi = \text{vec} (\Phi) \quad (n^2 \times 1)
\]
\[
\hat{\varphi} = \text{vec} (\hat{S}_H) \quad [((m \cdot n) \times 1]
\]
\[
(\text{where vec}(X) = [X_1^{(r)} \cdots X_n^{(r)}])
\]
and
\[
\hat{S}_H = \left[
\begin{array}{ccc}
S_1 & 0 & \cdots \\
0 & S_2 & \cdots \\
\vdots & \ddots & \ddots \\
0 & \cdots & S_H
\end{array}
\right] \quad [((m \cdot n) \times n^2)].
\]

Using this notation the system of equations (2.3) with \( S \) replaced by \( \hat{S}_H \) can be written as
\[
\hat{\varphi} = \hat{S}_H \varphi + \text{residuals}.
\]

Invoking the Gauss–Markov theory, it then follows that the minimum variance linear estimate of \( \varphi \) is of the form (see, e.g., [17])
\[
\hat{\varphi} = (\hat{S}_H^* B \hat{S}_H)^{-1} (\hat{S}_H^* \hat{\varphi})
\]
(2.11)
where \( B \) is a positive definite \((m \cdot n) \times (m \cdot n)\) matrix. However, in multiple-source applications the computational burden associated with (2.11) may be significantly greater than that corresponding to (2.10). For this reason we limit ourselves to (2.10) in what follows.

**Remark.** The SR method considered in this paper determines the estimate of \( \theta \) by solving the overdetermined system of equations \( S_1 \hat{\theta} = \hat{S}_2 \) in the least squares sense (see (2.6)). Alternatively, this system can be solved by the total least squares method [21], [22]. However, it can be shown that the least squares and the total least squares SR methods have the same asymptotic accuracy [4]. Thus, the explicit expression derived in this paper for the estimation error covariance matrix of the former applies also to the latter.

### D. MUSIC Method

If the sensors are omnidirectional and their locations are known, the transfer vector between the \( i \)th source signal \( x_i(t) \) and the array output \( y(t) \) depends on \( \theta_i \) only. Let this vector be denoted by \( a(\theta_i) \in C^{n \times 1} \). Then
\[
A = [a(\theta_1) \cdots a(\theta_n)]
\]
(2.12)
and it follows from R3 that
\[
a^*(\theta_i) GG^* a(\theta_i) = 0 \quad \text{for } i = 1, \cdots , n.
\]
(2.13)

MUSIC relies on (2.13). The MUSIC estimates of \( \{ \theta_i \} \) are determined as the locations of the \( n \) deepest nulls of the function
\[
f(\omega) = a^*(\omega) GG^* a(\omega).
\]
(2.14)

The minimization of \( f(\omega) \) is usually done by evaluating (2.14) at the points of a fine grid covering the domain of interest, and it appears to be a more costly operation than the calculations in steps 2 and 3 of the SR method. Note that the first step of both MUSIC and SR consists of computing the eigendecomposition of \( \tilde{K} \) (more details about the computational advantages of SRM over MUSIC can be found in [22]).

The statistical accuracy of the MUSIC method has been studied in [15]. It was shown there that for large \( N \), the MUSIC estimation errors \( (\hat{\theta}_i - \theta_i) \) are jointly Gaussian distributed with zero means and variances – covariances given by
\[
E(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) = \frac{\sigma}{2N} \text{Re} \left[ (d_i^* GG^* d_j) (a_k^* U_{ij}) \right]
\]
(2.15)

where
\[
d_k = da(\theta_k)/d\theta_k
\]
(2.16)
\[
U = S \tilde{A}^{-2} S^* \quad \tilde{A} = GG^*
\]
(2.17)

and \( a_k \) is a short notation for \( a(\theta_k) \).

In the next section we establish a similar result for the SR method. Then it becomes possible to compare the statistical performance of SR and MUSIC methods, which is done in Section IV.

### III. STATISTICAL ANALYSIS OF SR METHOD

**Theorem 3.1.** The SR estimation errors \( (\hat{\theta}_i - \theta_i) \) are asymptotically (for large \( N \)) jointly Gaussian distributed with zero means and variances – covariances given by
\[
E(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) = \frac{\sigma}{2N} \text{Re} \left[ e^{i(\theta_i - \theta_j)} (a_k^* a_l) (a_k^* U_{ij}) \right]
\]
(3.1)
where
\[ \rho^* = \left[ (A^T W A)_i^{-1} A^T W F_1 \right]_{(r)} \]
and
\[ F_1 = \begin{bmatrix} 0 & I_n \end{bmatrix} - e^{i\theta} [I_n, 0] \quad (m \times m). \]

Proof: Appendix A establishes the above result and also contrasts it to a related result derived recently in [4].

It is interesting to observe from (3.1) that the SR variance may take large values in any of the following cases: a) the sources are closely spaced (then 4 is almost rank deficient; consequently \( \{ \lambda_i - \sigma \}_1 \) is a small, \( (A^T W A)_i^{-1} \) has large elements, and these imply that the elements of \( \{ \rho^*_k \} \) are large); b) the signal-to-noise ratio is low (then again \( \{ \lambda_i - \sigma \}_1 \) are small); and c) the source signals are highly correlated (then \( P \) is nearly singular and as a consequence \( \{ \rho^*_k \} \) are small once again).

The above observations emerge perhaps more clearly, if the following equivalent expression of \( a^*_k U_{n_k} \) is used in (3.1) (see [15]):
\[ a^*_k U_{n_k} = \left[ P \right]_{n_k} + \sigma \left[ P \left( A^A A^{-1} P \right) \right]_{n_k}. \]

Note that the formula (3.1), (3.4) of the SR covariance matrix depends on the original parameters \( \theta \), \( P \), and \( \sigma \) of the problem only, unlike (3.1), (2.17) which also depends on the eigenelements of \( R \).

IV. PERFORMANCE COMPARISON OF SR AND MUSIC METHODS

The main result of this section states that the SR asymptotic variance cannot be less than the MUSIC asymptotic variance (this result lends theoretical support to the empirical observations in [22] and [24]; see also the references therein).

Theorem 4.1: Define
\[ \gamma_k = \frac{\text{var}_{\text{SR}}(\hat{\theta}_k)}{\text{var}_{\text{MUSIC}}(\hat{\theta}_k)}. \]

Then
\[ \gamma_k = (\rho^*_k \rho_k) (d^* G G^* d) \geq 1 \quad k = 1, \cdots, n. \]

Proof: See Appendix B.

It is worth noting that the variance ratio \( \gamma_k \) depends on \( m \) and \( \{ \hat{\theta}_k \} \) but not depend on \( \sigma \) or \( P \). This ratio may take large values for arrays with many sensors. Thus, in such a case the commonly used SR method is significantly less efficient from the statistical standpoint than MUSIC, in large samples. We illustrate this claim by means of a simple example.

Example 4.1: Consider the case of a single source signal (i.e., \( n = 1 \)) impinging on a uniform linear array of \( m \) omnidirectional sensors, and choose \( m = m - 1 \) and \( W = I \). Then
\[ a = \begin{bmatrix} 1 & e^\theta & \cdots & e^{i(m-1)\theta} \end{bmatrix}^T \]
\[ d = \begin{bmatrix} 0 & \frac{\alpha a^*}{a^* a} & \cdots & \frac{\alpha^* d d^*}{d d^*} \end{bmatrix}^T \]
\[ d^* G G^* d = d^* \left( I - \frac{\alpha a^*}{a^* a} \right) d = d^* d - \frac{|d^* d|^2}{a^* a} \]
\[ = \frac{m(m-1)(2m-1) - m^2 (m-1)^2}{6} \]
\[ = \frac{m^2 - 1}{12} \]

Thus, in such a case the optimal weighting matrix is
\[ W = I \]

(since multiplication of \( W \) by a scalar does not change the weighted SR estimate (2.10)).

Overlapping Subarrays: In the case of overlapping subarrays (i.e., \( m > \lfloor m/2 \rfloor \)) the optimal weighting matrix (5.3) is different from \( W = I \), in general. It is worth noting that the choice (5.3) of \( W \) minimizes the variance of the \( k \) direction estimate only. The other variances might be larger for \( W = V_k^{-1} \).
than for \( W = I \). Thus, the additional computational burden associated with using \( W = V \) given by (5.3) instead of \( W = I \) is motivated only if the user is faced with the requirement of more accurately estimating a single specified direction. Of course, the optimally weighted SR (OWSR) method introduced above could be used with weights \( W = V^{-1} \), \( W = V^2 \), etc., to provide accurate estimates of all directions. However, this repeated use of OWSR leads to a significant increase of the computational burden and in addition it is an open question whether it provides the minimum variance estimate of \( \theta \) (see the discussion in Section II).

Concerning implementation of OWSR, the matrix \( V \) which needs to be inverted in (5.3) has a special structure (it is Hermitian, banded, and Toeplitz). Thus, its inversion can be done efficiently. In Appendix D an explicit formula of \( V^{-1} \) is derived for the case of \( m = m - 1 \). Note that in practice \( \theta_0 \) in \( V \) will be replaced by a consistent estimate (as shown in Section III, this replacement does not affect the asymptotic accuracy properties of the SR estimate). A consistent estimate of \( \theta_0 \) can be obtained in a previous stage (e.g., by using the SR procedure with \( W = I \)), or it may be available from a priori information.

The OWSR method can provide significantly more accurate estimates than the SR method commonly used (with \( W = I \)). The next example, which is a continuation of Example 4.1, lends some support to this claim.

**Example 5.1:** Consider the situation described in Example 4.1. Using the calculations in Appendix D (\( Z \) and \( \tilde{a} \) appearing below are as defined there), we get

\[
Z_A = e^{-\theta} \left[ \tilde{m} \ (m - 1) e^{\theta} \ ... \ e^{i(m-1)\theta} \right]^T
\]

\[
\tilde{a}^T Z_A = e^{i(m-1)\theta} \left[ 1 + 2 \ + \ ... \ + (m - 1) \right]
\]

\[
A^*_s Z A_s = (m - 1)^2 + (m - 2)^2 + \ ... \ + 1
\]

\[
m(m - 1)(2m - 1)/6
\]

\[
A^*_s Z \tilde{a}^T Z A_s / (m + 1) = (m - 1)^2 m/4
\]

and

\[
A^*_s V^{-1} A_s = m(m - 1)(2m - 1)/6
\]

\[-m(m - 1)/4 = m(m - 1)/12.\]

Comparing with the calculations in Example 4.1, we conclude that in this case

\[
\text{var}_{\text{OWSR}} (\hat{\theta})/\text{var}_{\text{MUSIC}} (\hat{\theta}) = 1
\]

and

\[
\text{var}_{\text{SR}} (\hat{\theta})/\text{var}_{\text{OWSR}} (\hat{\theta}) > 1 \quad (\text{and} \gg 1 \text{for large} \ m).
\]

In the above example, OWSR and MUSIC had identical performance. However, this is not generally true as shown in the next example.

**Example 5.2:** Consider a situation in which a single source signal impinges on a uniform linear array containing \( m = 5 \) sensors. The array is (fictitiously) split into two overlapping subarrays of \( m = 3 \) sensors each. This situation corresponds to

\[
A = \begin{bmatrix} 1 & e^{i\theta} & e^{2i\theta} \\ 1 & e^{i\theta} & e^{2i\theta} \end{bmatrix}
\]

\[
A_s = \begin{bmatrix} 1 & e^{i\theta} \\ 1 & e^{i\theta} \end{bmatrix}^T.
\]

Thus,

\[
d = 0 \ i3e^{i\theta} \ i\theta \ i4e^{i\theta} \ i2e^{i\theta}\]

\[
d^*GG^*d = d^*d - \frac{1}{A^*A} \left[ \begin{array}{c} -e^{i\theta} \\ 0 \\ -e^{i\theta} \\ 0 \\ -e^{i\theta} \end{array} \right] \left[ \begin{array}{c} -e^{i\theta} \\ 0 \\ -e^{i\theta} \\ 0 \\ -e^{i\theta} \end{array} \right] = 10
\]

\[
F = \begin{bmatrix} -e^{i\theta} & 0 & 1 & 0 & 0 \\ 0 & -e^{i\theta} & 0 & 1 & 0 \\ 0 & 0 & -e^{i\theta} & 0 & 1 \end{bmatrix}
\]

\[
V^{-1} = (FP)^{-1} = \begin{bmatrix} 2 & 0 & e^{-i\theta}^{-1} \\ 0 & 2 & 0 \\ e^{-i\theta} & 0 & 2 \end{bmatrix}
\]

\[
A^*_s V^{-1} A_s = \frac{1}{6}
\]

\[
\begin{bmatrix} 4 & 0 & 2e^{-i\theta} \\ 0 & 3 & 0 \\ 2e^{-i\theta} & 0 & 4 \end{bmatrix}
\]

and

\[
A^*_s V^{-1} A_s = 5/2
\]

which gives

\[
\text{var}_{\text{OWSR}} (\hat{\theta})/\text{var}_{\text{MUSIC}} (\hat{\theta}) = (A^*_s V^{-1} A_s)^{-1} (d^*GG^*d) = 4.
\]

**VI. Concluding Remarks**

The explicit formula derived in this paper for the covariance matrix of the large-sample errors in the SR-based direction estimates can also be used for numerical studies of performance. For example, numerical comparisons of SR, OWSR, and MUSIC variances over a wide range of scenarios (corresponding to different SNR values, source separations, array configurations, etc.) would provide more insight into the differences in performance between these methods. Another aspect that is more difficult to study analytically but can also be studied numerically is the choice of \( m \) in those cases where the user is faced with this problem. This choice can be made in specific situations by numerically evaluating the error variance formula over the range of possible values of \( m \) and for a given scenario that is thought of being feasible from a priori information.

The behavior in finite samples of the OWSR method introduced in this paper would also be interesting to study. This method focuses on a direction parameter (at the user’s choice) for which it determines an improved estimate, and appears to be of interest for a number of applications such as surveillance, radio communication and remote sensing.

**APPENDIX A**

**Proof of Theorem 3.1**

Let \( \mu_k = e^{i\theta} \) denote the \( k \)th eigenvalue of \( \Phi \) (by assumption \( \Phi \) has simple eigenvalues) and let \( \mu_k = \mu_k e^{i\theta} \) denote the \( k \)th eigenvalue of \( \Phi \). Since \( \mu_k \) is a consistent estimate of \( \mu_k \) we can write, for sufficiently large \( N \)

\[
\theta_k = \text{Im} [\ln \mu_k] = \text{Im} \left[ \ln \left( \frac{\mu_k - \mu_k + \mu_k}{\mu_k} \right) \right]
\]

\[
= \text{Im} [\ln \mu_k] + \text{Im} \left[ \ln \left( 1 + \frac{\mu_k - \mu_k}{\mu_k} \right) \right]
\]

\[
= \theta_k + \text{Im} \left[ \frac{\mu_k - \mu_k}{\mu_k} \right]. \quad (A.1)
\]
Note that the approximate equalities in this Appendix are to be interpreted as first-order approximations.

Next we make use of a result in [16] according to which, for large $N$

$$
\hat{\mu}_k - \mu_k = \alpha_k^* (\hat{\phi} - \phi) \beta_k
$$

where $\alpha_k$ and $\beta_k$ are as defined in Section II (see (2.8) and (2.9) there). Using the definition (2.10) of $\hat{\phi}$ and (2.3), gives

$$
\hat{\phi} - \phi = (S^* S W S)^{-1} S^* W (S - S) \beta_k
$$

$$
= (S^* S W S)^{-1} S^* W [(S - S) - (S - S)] \beta_k
$$

(A.3)

Since $\beta_k$ is a right eigenvector of $\Phi$ associated with the eigenvalue $e^{\theta_k}$ (see Section II), i.e.,

$$
\Phi \beta_k = e^{\theta_k} \beta_k
$$

(A.4)

it follows that

$$
\hat{\mu}_k - \mu_k = \alpha_k^* (S^* S W S)^{-1} S^* W [10 I_N]
$$

$$
- e^{\theta_k} [I_N 0]) (S - S) \beta_k
$$

(A.5)

From the definition (2.8) of $\alpha_k^*$ and (2.9), we get

$$
\alpha_k^* (S^* S W S)^{-1} S^* = C_k^{-1} (C^* A_k^* W A_k C)^{-1} C^* A_k^*
$$

$$
= C_k^{-1} C^{-1} (A_k^* W A_k)^{-1} A_k^*
$$

$$
= [(A_k^* W A_k)^{-1} A_k^*]^{(r)}
$$

(A.6)

Inserting (A.6) in (A.5) gives

$$
\hat{\mu}_k - \mu_k = \rho_k^* (S - S) \beta_k
$$

$$
= \left[ \begin{array}{c}
\rho_k^* (S - s_1) \\
\vdots \\
\rho_k^* (S - s_n)
\end{array} \right]
$$

$$
= \left[ \beta_{k,1} \\
\vdots \\
\beta_{k,n}
\right]
$$

(A.7)

The eigenvector estimation errors $(\hat{s}_i - s_i)$ for $i = 1, \cdots, n$ are asymptotically (for large $N$) jointly Gaussian distributed with zero means and covariances given by

$$
E(\hat{s}_i - s_i)(\hat{s}_j - s_j)^* = \frac{\lambda_i}{N} \left[ \sum_{p=1}^{n} \frac{\lambda_p}{(\lambda_p - \lambda_i)^2} \hat{s}_p^2 \rho_p^2 + \frac{\sigma}{(\lambda_i - \lambda_j)} \right] \delta_{i,j}
$$

(A.8a)

$$
E(\hat{s}_i - s_i)(\hat{s}_j - s_j)^* = -\frac{\lambda_i \lambda_j}{N(1 - \lambda_i - \lambda_j)^2} \delta_{i,j}
$$

(A.8b)

(see, e.g., [20]). From the above result and (A.1) and (A.7), the asymptotic Gaussian distribution of $(\hat{\theta}_i - \theta_i)$ follows immediately. It remains to verify the expression (3.1) for the covariance matrix of this distribution.

It is readily verified that for $\varepsilon, \omega \in C$

$$
\text{Im}(\varepsilon) \cdot \text{Im}(\omega) = \frac{1}{2} \text{Re}(\varepsilon \omega^* - \omega \varepsilon^*)
$$

(A.9)

Thus,

$$
\text{cov} (\hat{\theta}, \hat{\theta}) = \frac{1}{2} \text{Re} \left[ e^{i(\theta_k - \theta)} E(\hat{\mu}_k - \mu_k)(\hat{\mu}_p - \mu_p)^* - e^{-i(\theta_k - \theta)} E(\hat{\mu}_k - \mu_k)(\hat{\mu}_p - \mu_p) \right].
$$

(A.10)

Next, we note the following important fact:

$$
\rho_k^* A = \left[ (A_k^* W A_k)^{-1} A_k^* W F_k A_k \right]^{(r)}
$$

$$
= \left[ (A_k^* W A_k)^{-1} A_k^* W (A_k \Psi - A_k e^{\theta_k}) \right]^{(r)}
$$

$$
= [\Psi - e^{\theta_k} I_N]^{(r)} = 0
$$

(A.11)

which by R2 implies

$$
\rho_k^* S = 0 \quad k = 1, \cdots, n.
$$

(A.12)

From (A.7), (A.8), and (A.12) we get

$$
E(\hat{\mu}_k - \mu_k)(\hat{\mu}_p - \mu_p) = 0
$$

(A.13)

and

$$
E(\hat{\mu}_k - \mu_k)(\hat{\mu}_p - \mu_p)^* = \frac{\sigma}{N} \left[ \beta_{k,1} \rho_k^* \cdots \beta_{k,n} \rho_k^* \right]
$$

$$
= \frac{\sigma}{N} \left[ \begin{array}{c}
\beta_{k,1} \\
\vdots \\
\beta_{k,n}
\end{array} \right]
$$

$$
= \left[ \begin{array}{c}
\lambda_i \\
(\lambda_i - \sigma)
\end{array} \right]
$$

$$
= \left[ \begin{array}{c}
\rho_k^* G G^* \\
0
\end{array} \right]
$$

$$
= \left[ \begin{array}{c}
\beta_{k,1} \rho_k^* \\
\vdots \\
\beta_{k,n} \rho_k^*
\end{array} \right]
$$

$$
= \frac{\sigma}{N} \left[ \rho_k^* G G^* \right] \left[ \begin{array}{c}
\beta_{k,1} \\
\vdots \\
\beta_{k,n}
\end{array} \right]
$$

(A.14)

Using R2 and (2.9) gives

$$
\rho_k^* \hat{\lambda}_k^{-2} \Delta \beta_k = (A_k^* \hat{\lambda}_k^{-2} A_k^* \Delta \beta_k) = \rho_k^* U_{\lambda_k}.
$$

(A.15)

Further, from R3

$$
\rho_k^* G G^* = \rho_k^* [I - A(A_k^* A_k)^{-1} A_k^*] = \rho_k^*.
$$

(A.16)

Inserting (A.13)–(A.16) in (A.10) we obtain

$$
\text{cov} (\hat{\theta}, \hat{\theta}) = \text{cov} (\hat{\theta}, \hat{\theta})
$$

$$
= \frac{\sigma}{2N} \left[ e^{i(\theta_k - \theta)} (\rho_k^* \rho_k^*) (\rho_k^* \rho_k^*) \right]
$$

(A.17)

and the proof is complete.

Remark: A related study of the SR large-sample error variances has appeared recently in [4]. This study is limited to uniform linear arrays and is fairly incomplete. Specifically, the main result obtained in [4] is essentially (A.5) (for $W = I$).
This preliminary result (derived here in a much simpler manner than in [4]) can be used, along with (A.1) and (A.8), to compute the SR direction estimate variances, but it does not provide a compact explicit formula for these variances. Indeed, the variance formula obtained in [4] from (A.5) only, is unenlightening and rather involved (inter alia, it contains complicated expressions for terms such as (A.13) which, in fact, are equal to zero!).

Note that to obtain our compact variance formula (A.17) from (A.5), the key results R2, (2.8), (2.9), (A.12), and (A.16) have first to be observed.

**APPENDIX B**

**Proof of Theorem 4.1**

From (2.15) and (3.1) we get

\[
\gamma_k = (\rho^*_k \rho_k) (d^*_k G G^* d_k).
\]  

(B.1)

By the Cauchy–Schwarz inequality (see, e.g., [16])

\[
\gamma_k \lesssim (\rho^*_k \rho_k) \left[ (G G^*)^2 d_k^* d_k \right] \]

\[
\lesssim | \rho^*_k G G^* d_k |^2 = | \rho^*_k d_k |^2
\]

(B.2)

where the last equality follows from (A.16). Now, from (A.11)

\[
\rho^*_k d_k = 0.
\]

(B.3)

Taking the derivative of (B.3) with respect to \( \theta_k \) gives

\[
\rho^*_k d_k + \rho^*_k a_k = 0.
\]

(B.4)

Since

\[
F_k a_k = [F_k A_k]^{(c)} = [A_k \Psi - e^{i \phi_k} A_k]^{(c)} = 0
\]

(B.5)

we note in passing that (B.3) results also from (B.5), it follows that

\[
\rho^*_k a_k = [A^*_k W A_k]^{(c)} d^*_k a_k
\]

\[
= -i e^{i \phi_k} [A^*_k W A_k]^{(c)} [I_0 0] a_k
\]

\[
= -i e^{i \phi_k} [A^*_k W A_k]^{(c)} [A_k]^{(c)} = -i e^{i \phi_k}.
\]

(B.6)

Finally, from (B.2), (B.4), and (B.6)

\[
\gamma_k \gtrsim | i e^{i \phi_k} |^2 = 1
\]

which is the stated result.

**APPENDIX C**

**Proof of Theorem 5.1**

Since

\[
\rho^*_k \rho_k = [(A^*_k W A_k)^{-1} A^*_k W V_k W A_k (A^*_k W A_k)^{-1}]^{(c)}
\]

is the only factor of \( \var(X_k) \) that depends on \( W \), and since

\[
(A^*_k W A_k)^{-1} A^*_k W V_k W A_k (A^*_k W A_k)^{-1} - (A^*_k V_k^{-1} A_k)^{-1}
\]

\[
\cdot V_k W A_k (A^*_k W A_k)^{-1} - V_k^{-1} A_k (A^*_k W A_k)^{-1} A^*_k
\]

is a positive semidefinite matrix, the inequality (5.1) follows at once.

The fact that the lower bound in (5.1) is attained for the choice (5.3) of \( W \) is obvious.

**APPENDIX D**

**Analytical Inversion of \( V \) for \( \hat{m} = m - 1 \)**

In this Appendix we assume that \( \hat{m} = m - 1 \). We also omit the index \( k \) of \( V \), \( \Psi \), etc., to simplify the notation. Let

\[
Z = \begin{bmatrix} e^{i \phi} & -1 & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{i \phi} \end{bmatrix} \quad (\hat{m} \times \hat{m})
\]

and let

\[
u = [0 \cdots 0 1]^T \quad (m \times 1)
\]

Then

\[
V = (Z^* Z)^{-1} + u u^*
\]

and by the matrix inversion lemma (see, e.g., [17])

\[
V^{-1} = Z^* Z + Z^* u u^* Z/(1 + u^* Z Z u)
\]

(D.1)

Some straightforward calculations give

\[
Z = \begin{bmatrix} e^{-i \phi} & e^{-i \phi} & \cdots & e^{-i \phi} \\ e^{i \phi} & e^{i \phi} & \cdots & e^{i \phi} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & e^{i \phi} \end{bmatrix}
\]

(D.2)

and

\[
u^* Z Z u = \hat{m}.
\]

(D.3)

Denoting the last column of \( Z \) by \( \bar{u} \)

\[
\bar{u} = [Z]^{(c)}
\]

(D.4)

and inserting (D.3) and (D.4) in (D.1), we get

\[
V^{-1} = Z^* [I - \bar{u} \bar{u}^*/(\hat{m} + 1)] Z
\]

(D.5)

Finally, a simple calculation gives

\[
Z^* Z = \begin{bmatrix} 1 & e^{i \phi} & \cdots & e^{i(\hat{m} - 1) \phi} \\ e^{i \phi} & 2 & \cdots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ e^{i(\hat{m} - 1) \phi} & \cdots & (\hat{m} - 1) e^{i \phi} & \hat{m} \end{bmatrix}
\]

and

\[
Z^* \bar{u} = e^{i(\hat{m} - 1) \phi} [1 2 e^{i \phi} \cdots \hat{m} e^{i(\hat{m} - 1) \phi}]^T
\]

and the evaluation of \( V^{-1} \) is complete.

**Acknowledgment**

The authors thank one of the reviewers for bringing [4] to their attention and for highlighting the main differences between the results in [4] and in Section III of this paper.
REFERENCES


Petre Stoica received the M.Sc. and Ph.D. degrees, both in automatic control, in 1972 and 1979, respectively.

Since 1972 he has been with the Department of Automatic Control, the Polytechnic Institute of Bucharest, Romania. His research interests include various aspects of system identification, time series analysis, and signal processing. For papers on these topics he received three national prizes. He is coauthor of five books, the most recent being System Identification (Englewood Cliffs, NJ: Prentice-Hall, 1989).

Dr. Stoica is a member of the Board of Directors of the Time Series Analysis and Forecasting (TS&AF) Society, and an Associate Editor for the Journal of Forecasting. He has been given the Member of TS&AF (MTS&F) honors award. He was co-recipient with A. Nehorai of the IEEE Signal Processing Society's Senior Award for the paper, "MUSIC, Maximum Likelihood, and the Cramér–Rao Bound," published in 1989.

Arye Nehorai (S’80–M’83–SM’90) received the B.Sc. and M.Sc. degrees in electrical engineering from the Technion—Israel Institute of Technology, in 1976 and 1979, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1983.

From 1983 to 1984, he was a Research Associate at Stanford University. From 1984 to 1985, he was a Research Engineer at Systems Control Technology Inc., in Palo Alto, CA. Since 1985 he has been with the Department of Electrical Engineering and the Applied Mathematics Program at Yale University, New Haven, CT, where he is an Associate Professor. During parts of 1989 and 1990 he held Visiting Professorships at Uppsala University, Sweden, and the Technion, Israel. His areas of interest are adaptive filtering, sensor array processing, system identification, and biomedical engineering.

Dr. Nehorai is an Associate Editor of the journal Circuits, Systems and Signal Processing and was an Associate Editor of the IEEE Transactions on Acoustics, Speech, and Signal Processing. He is a member of the Technical Committee on Spectrum Estimation and Modeling and the Education Committee in the IEEE Signal Processing Society, and is the Chairman of the Connecticut IEEE Signal Processing Chapter. He was co-recipient with P. Stoica of the IEEE Signal Processing Society's Senior Award for the paper, "MUSIC, Maximum Likelihood, and Cramér–Rao Bound," published in 1989. He is a member of Sigma Xi.