

Rational Invariant Subspace Approximations with Applications

Mohammed A. Hasan, *Member, IEEE*, Mahmood R. Azimi-Sadjadi, *Senior Member, IEEE*, and Ali A. Hasan

Abstract—Subspace methods such as MUSIC, Minimum Norm, and ESPRIT have gained considerable attention due to their superior performance in sinusoidal and direction-of-arrival (DOA) estimation, but they are also known to be of high computational cost. In this paper, new fast algorithms for approximating signal and noise subspaces and that do not require exact eigendecomposition are presented. These algorithms approximate the required subspace using rational and power-like methods applied to the direct data or the sample covariance matrix. Several ESPRIT- as well as MUSIC-type methods are developed based on these approximations. A substantial computational saving can be gained comparing with those associated with the eigendecomposition-based methods. These methods are demonstrated to have performance comparable to that of MUSIC yet will require fewer computation to obtain the signal subspace matrix.

Index Terms—DOA, ESPRIT, frequency estimation, minimum norm, MUSIC, power methods, rational subspace approximation.

I. INTRODUCTION

THE SIGNAL subspace approach has found applications in several fields such as harmonic retrieval [1], [2], spectral estimation and autoregressive moving average (ARMA) modeling [3], [4], sensor array processing [5], [6], system identification [7], and even in filter design applications [8]. Signal subspace algorithms can usually provide much better performance than traditional least squares methods; however, associated computational load make them less attractive for real-time implementation. Among the most attractive ones are MUSIC [5], MIN-NORM linear prediction [9], [10] and ESPRIT [11]–[14]. In subspace methods, the data matrix or a matrix of some statistics of the data is normally decomposed into two orthogonal subspaces. Then, the direction of arrival (DOA) is estimated using the orthogonality of the noise subspace and the array manifold (MUSIC and MIN-NORM) or the rotation invariance over the signal subspace (ESPRIT). This decomposition is usually carried out using the singular value or eigenvalue decomposition. Several exact methods have been presented in [15] and [16]. However, the computation of these exact decompositions is often very intensive, which may make the subspace algorithms prohibitive.

Manuscript received October 9, 1998; revised November 11, 1999. This work was supported in part by the Office of Naval Research [ONR32ITS] under Grant N61331-94-K-0018. The associate editor coordinating the review of this paper and approving it for publication was Dr. Eric Moulines.

M. Hasan is with the Department of Electrical and Computer Engineering, University of Minnesota Duluth, Duluth, MN 55812 USA.

M. R. Azimi-Sadjadi is with the Department of Electrical Engineering, Colorado State University, Fort Collins, CO 80521 USA.

A. A. Hasan is with the College of Electronic Engineering, Nasser University, Bani Waleed, Libya.

Publisher Item Identifier S 1053-587X(00)07645-5.

Several techniques that seek to determine signal subspace-based estimates without eigendecomposition have been proposed. These include the principal component AR method [17], which uses a polynomial approximation to the pseudoinverse of the covariance matrix. The method of [18] approximates a basis of the signal subspace from a set of transformed data vectors where the signal subspace was approximated using the discrete Fourier and Cosine transforms. In [19], a basis for signal subspace were obtained by solving a linear least squares prediction problem. In [20], Tuft and Melissinos used Lanczos and power-type method to approximate the signal subspace. Ermolaev and Gershman [9], [10] used powers of the sample covariance matrix based on Krylov subspaces to approximate the noise subspace when the number of impinging signals and a threshold that separates the signal and noise eigenvalues are known *a priori*. These methods are shown to have significant computational saving over those that explicitly compute eigen or singular vectors. However, in most applications, the above two parameters (a threshold and the number of signals) are unknown and, thus, place a serious limitation on the usefulness of these techniques. It should be mentioned that none of these methods are developed in the context of the less-costly ESPRIT method.

The objective of this paper is to extend the results of [9], [10], and [20] in approximating the signal and noise subspaces and apply them for the derivation of ESPRIT- as well as MUSIC-type methods. We will present arbitrarily accurate approximations of subspace decomposition using rational approximations and a power-like method. Minimum norm, MUSIC-, and ESPRIT-type estimators will be derived and shown to provide numerically efficient and accurate solutions. This include a method of estimating signal subspace when the number of sources are known, and no *priori* knowledge of a threshold is required. This is an improvement of the method in [9], where both the number of sources and the threshold must be known. The approach presented here is also useful in other DOA estimators such as beamspace MUSIC [21], FINE [22], and generalized MIN-NORM [10].

This paper is organized as follows. Section II describes the data model of the frequency estimation problem and background of high-resolution methods. Approximated subspaces based on rational approximations are given in Section III. Minimum-norm, MUSIC-, and ESPRIT-type estimators are developed in Section IV. Subspace approximations using power-like methods are given in Section V. Operation count is discussed in Section VI. Finally, Section VII contains several simulation results designed to show that although our algorithms use significantly less computation than methods based

on exact decomposition, they give comparable performance. All proofs are given in the Appendices.

II. PROBLEM FORMULATION

A. Model Description

The directions of arrival (DOAs) and the harmonic retrieval problems can be described as follows. In the DOA problem, we consider M narrowband planewaves simultaneously incident on a uniform linear array of L sensors. The signals arriving at all sensors during N snapshots are

$$\mathbf{y}(k) = \mathbf{x}(k) + \mathbf{v}(k) \quad (1a)$$

where

$$\mathbf{x}(k) = \begin{bmatrix} 1 & \cdots & 1 \\ e^{j(2d/\lambda)\sin(\theta_1)} & \cdots & e^{j(2d/\lambda)\sin(\theta_M)} \\ \vdots & \vdots & \vdots \\ e^{j(2d/\lambda)(L-1)\sin(\theta_1)} & \cdots & e^{j(2d/\lambda)(L-1)\sin(\theta_M)} \end{bmatrix} \cdot \begin{bmatrix} s_1(k) \\ \vdots \\ s_M(k) \end{bmatrix} = \mathbf{A}\mathbf{s}(\theta)(k) \quad (1b)$$

where $s_k(t)$ is the k th narrowband signal (with center wavelength λ_c) arriving at angle θ_k , d is the spacing between adjacent sensors, and the array manifold matrix

$$\mathbf{A}(\theta) = [a(\theta_1) \quad a(\theta_2) \quad \cdots \quad a(\theta_M)]$$

with

$$a(\theta) = [1 \quad e^{j(2d/\lambda)\sin(\theta)} \quad \cdots \quad e^{j(2d/\lambda)(L-1)\sin(\theta)}]^T$$

is the steering vector of the array toward the direction θ . Here, $\mathbf{s}(k) \in \mathcal{C}^M$ (\mathcal{C} is the field of complex numbers), and

$$\mathbf{s}(k) = [s_1(k) \quad s_2(k) \quad \cdots \quad s_M(k)]^T$$

$\mathbf{y}(k)$ and $\mathbf{v}(k)$ are $L \times 1$ vector of observation and additive noise in sensors with

$$\begin{aligned} \mathbf{y}(k) &= [y_1(k) \quad y_2(k) \quad \cdots \quad y_L(k)]^T \\ \mathbf{v}(k) &= [v_1(k) \quad v_2(k) \quad \cdots \quad v_L(k)]^T. \end{aligned}$$

It is also assumed that the signals and additive noise are stationary and zero-mean ergodic complex-valued random processes such that $E[v_i(k)v_j^*(k)] = \sigma_v^2\delta_{i-j}$ for $i, j = 1, \dots, L$. Here, $E[\cdot]$ and $*$ denote the expectation and conjugate transpose operators, respectively.

The harmonic retrieval problem can be described as follows. Consider the sum of M complex sinusoids in additive zero-mean complex white Gaussian noise $v(k)$ having variance σ_v^2 :

$$y(k) = x(k) + v(k) \quad (1c)$$

where

$$x(k) = \sum_{i=1}^M d_i e^{j(w_i k + \phi_i)} \quad (1d)$$

where $w_i, i = 1, \dots, M$ are sinusoidal frequencies, and d_i are positive amplitudes. In addition, it is assumed that the following conditions hold: 1) The ϕ_i are independent random variables uniformly distributed over $[-\pi, \pi]$, and 2) $v(k)$ is white with zero mean and independent of ϕ_i .

The covariance matrix R_y for the array data is given by

$$R_y = E\{\mathbf{y}(k)\mathbf{y}(k)^*\} = AR_S A^* + \sigma_v^2 I_L$$

where R_S is the covariance matrix for $s(k)$, and σ_n^2 is an unknown noise variance. Similarly, in the harmonic retrieval problem, $R_y = E\{\mathbf{y}(k)\mathbf{y}(k)^H\} = R_x + \sigma_v^2 I_L$. The sample covariance matrix will be denoted by \hat{R}_y and is estimated as in Section II-B.

Note that the minimum eigenvalue of R_y is equal to σ_v^2 with multiplicity $L-M$. If the θ 's are all distinct, the unknown matrix $A \in \mathcal{C}^{L \times M}$ is of rank M . In general, only the covariance matrix R_y is known. The actual value M and the dimension of both A and S are unknown, as is the value of σ_v^2 . Given the matrix R_y , our goal is to determine the number of signals and A from the noisy data $\{y(k)\}_{k=1}^N$.

It is shown in [23] that the sample covariance matrix $\hat{R}_y = R_y + O_p(1/\sqrt{N})$, where R_y is the true covariance matrix of the signal $y(k)$. The additive term decreases to zero in probability as $N \rightarrow \infty$. Therefore, any of the methods applied here using sample covariance matrices yields consistent estimates of the amplitudes and frequencies.

B. Preliminaries

Generally, modern high-resolution sub-space estimation schemes are of the following types:

- extrema searching techniques like spectral MUSIC [5];
- polynomial rooting techniques such as Root-MUSIC and Pisarenko methods [3], [24];
- matrix shifting methods such as ESPRIT [12]–[14], matrix pencils and unitary ESPRIT [11].

The subspace decomposition can be performed on the data matrix Y using a singular value decomposition (SVD)

$$X = U\Sigma V^* = [U_s \quad U_0] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_s^* \\ V_0^* \end{bmatrix} \quad (2)$$

where the vectors in U_s , associated with the M nonzero singular values, span the signal subspace, whereas the vectors in U_0 , associated with the zero singular values, span the noise subspace for which

$$a(\theta_k)^* U_0 = 0, \quad k = 1, 2, \dots, M. \quad (3)$$

For a more self-contained description, let us first give the basis for some subspace methods.

- 1) *Pisarenko Method* [3]: It is shown by Pisarenko that a positive definite Toeplitz matrix of order $M+1$ can be modeled as the covariance matrix of a stationary stochastic signal consisted of at most n sinusoids and a white noise process. The energy of the white noise signal equals the smallest eigenvalue γ_0 of the Toeplitz matrix R_y . In the case that γ_0 is simple, the frequencies of the sinusoids w_1, \dots, w_M are defined as the zeros $z_i = \exp(jw_i)$ of the eigenpolynomial $p(z) =$

$[1, z, \dots, z^M]\xi$, which are the amplitudes of the waves associated with those zeros, where ξ is the eigenvector corresponding to γ_0 .

- 2) *MUSIC* [5] searches over angle θ of M global minima of the null-spectrum function

$$P(\theta) = a(\theta)^* U_0 U_0^* a(\theta). \quad (4)$$

- 3) *ROOT-MUSIC algorithm* [24] forms the roots of the null-spectrum polynomial

$$P(z) = a\left(\frac{1}{z}\right)^* U_0 U_0^* a(z) \quad (5)$$

where $\{z_i = e^{j(2d/\lambda)\sin(\theta_i)}\}_{i=1}^M$ are the polynomial roots. The M roots with largest amplitudes inside the unit circle are chosen as the signal roots.

It is known that the theoretical covariance matrix is Toeplitz and centro-symmetric, i.e., $R_y = J R_y^* J$, where J is the permutation matrix with ones along the cross diagonal. To effectively use the structure of the data, the sample correlation matrix \hat{R}_y is estimated using the forward-backward method so that $\hat{R}_y = 1/(2(N-L+1)) Y_L^* Y_L$, where

$$Y_L = \begin{bmatrix} y_L & y_{L-1} & \cdots & y_1 \\ y_{L+1} & y_L & \cdots & y_2 \\ \cdots & \cdots & \cdots & \cdots \\ y_N & y_{N-1} & \cdots & y_{N-L+1} \\ y_1^* & y_2^* & \cdots & y_L^* \\ y_2^* & y_3^* & \cdots & y_{L+1}^* \\ \cdots & \cdots & \cdots & \cdots \\ y_{N-L+1}^* & y_{N-L+2}^* & \cdots & y_N^* \end{bmatrix} \quad (6)$$

where the notation y_k denotes $y(k)$. Thus, one may expect that the forward-backward method in (6) yields a better estimate of R_y than the lower or upper part of Y_L .

In the SVD-based methods, the correlation matrix \hat{R}_y is decomposed as $\hat{R}_y = U_1 \Lambda_1 V_1^* + U_2 \Lambda_2 V_2^*$, where $U_i^* U_j = \delta_{i-j} I$, $V_i^* V_j = \delta_{i-j} I$, $U_i^* V_j = 0$, and Λ_i is diagonal matrix for $i, j = 1, 2$. Here

- δ_{i-j} Kronecker delta function;
- Λ_1 matrix of the most significant singular values;
- Λ_2 diagonal matrix whose diagonal holds the least significant singular values.

Several techniques are available in the literature to compute the SVD or solve the eigenvalue problem in general. Well-established methods can be found in EISPACK [25] and [26]. The computational complexity of these algorithms is of order $O(L^3)$, where L is the size of the matrix. In the next section, we utilize the idea that for high SNR, signal singular values are generally larger than noise singular values, and thus, powering would widen the separation of the noise and signal eigenvectors.

III. DOMINANT SUBSPACES OF THE SAMPLE COVARIANCE MATRIX

Since R_y is Hermitian, it has the eigendecomposition $R_y = \sum_{i=1}^L \lambda_i u_i u_i^*$, where λ_i and u_i are the i th eigenvalue and i th corresponding eigenvector. For convenience, it is assumed that the eigenvalues are sorted in decreasing order so that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M > \lambda_{M+1} = \dots = \lambda_L = \sigma_v^2$

with corresponding eigenvectors $\{u_i\}_{i=1}^L$. The eigenvectors $\{u_i\}_{i=1}^M$ are usually called the signal vectors, and the eigenvectors $\{u_i\}_{i=M+1}^L$ are called the noise vectors. Let $P_s = [u_1 \ u_2 \ \cdots \ u_M]$, $P_0 = [u_{M+1} \ u_{M+2} \ \cdots \ u_L]$. Then, $U_s = P_s P_s^* = \sum_{i=1}^M u_i u_i^*$ and $U_0 = P_0 P_0^* = \sum_{i=M+1}^L u_i u_i^*$ are projections onto the signal and noise subspaces, respectively. In presence of white noise $v(k)$, $A(\theta) R_s A^*(\theta) = U_s \Lambda^2 U_s^*$, where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_M)$, and hence, $A(\theta) R_s^{1/2} = U_s \Lambda V$ for some orthogonal matrix V . Therefore, $U_0^* A(\theta) R_s^{1/2} = U_0^* U_s \Lambda V = 0$ from which $U_0^* A(\theta) = 0$ or equivalently $\{u_i\}_{i=M+1}^L$ are orthogonal to $\{a(\theta_i)\}_{i=1}^M$. As stated earlier, $\hat{R}_y = R_y + O_p(1/\sqrt{N})$ and therefore $\hat{\lambda}_i = \lambda_i + O_p(1/\sqrt{N})$, $\hat{u}_i = u_i + O_p(1/\sqrt{N})$ which imply that $\text{span}\{\hat{u}_i\}_{i=1}^M = \text{span}\{u_i\}_{i=1}^M + O_p(1/\sqrt{N})$. Note that $\text{Span}\{\hat{u}_k\}$ is the maximum likelihood estimate of the signal subspace. This fact is the basis for the use of MUSIC, ESPRIT, and others. The eigendecompositions are expensive to compute, requiring $O(L^3)$ flops for reasonably good estimates. In the next result, we derive a method for splitting the eigenspace into several invariant subspaces.

In the next theorem, we generalize the above idea to obtain a rational approximation of the signal subspace.

Theorem 1: Let $R_y = \sum_{i=1}^L \lambda_i u_i u_i^*$, where λ_i and u_i are the i th eigenvalue and i th corresponding eigenvector. Assume that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M > \lambda_{M+1} = \dots = \lambda_L = \sigma_v^2$ and that the corresponding eigenvectors $\{u_i\}_{i=1}^L$ are chosen to be orthogonal. Let $S \geq 1$ and $\{b_i\}_{i=1}^S$ be real positive numbers such that $0 < b_1 < b_2 < \dots < b_S$. Define $U_0 = \sum_{i:b_1 > \lambda_i} u_i u_i^*$ and $U_1 = \sum_{i:b_1 < \lambda_i < b_2} u_i u_i^*$, \dots , $U_r = \sum_{i:b_r < \lambda_i < b_{r+1}} u_i u_i^*$, for $r = 1, \dots, S-1$. Then

$$\lim_{n \rightarrow \infty} \prod_{r=1}^S (b_r^n I_L - R_y^n) (b_r^n I_L + R_y^n)^{-1} = U_0 - U_1 + U_2 + \dots + (-1)^S U_S. \quad (7)$$

Proof: The proof is given in Appendix A.

Remark 1: If w is a primitive $(S+1)$ th root of unity, one can apply Theorem 1 several times to compute

$$\lim_{n \rightarrow \infty} \prod_{r=1}^S (b_r^n I_L + w^r R_y^n) (b_r^n I_L + R_y^n)^{-1} = \prod_{i=0}^S \left(\sum_{i=0}^r U_i + w^i \sum_{i=r+1}^S U_i \right) = \sum_{i=0}^S (w^i)^i U_i \quad (8)$$

for $l = 1, 2, \dots, S$. These S equations can be solved for the U_i 's as follows.

$$U_r = \frac{1}{S} \sum_{i=1}^S (w^{-r})^i W_i, \quad \text{for } r = 0, \dots, S$$

where

$$W_l = \lim_{n \rightarrow \infty} \prod_{r=1}^S (b_r^n I_L + w^l R_y^n) (b_r^n I_L + R_y^n)^{-1}.$$

The significance of Theorem 1 is that it provides a way of splitting \mathcal{C}^L into several complementary invariant subspaces of the matrix R_y without actually computing any eigenvalues. These subspaces correspond to eigenvalues inside the strips $\text{Real}(z) < b_1, b_r < \text{Real}(z) < b_{r+1}$ for $r = 1, \dots, S-1$ and $b_S < \text{Real}(z)$. In the presence of white noise, only two

subspaces are required for deriving subspace methods. Thus, in the last result, if $r = 1$, we obtain the following.

Theorem 2: Let $R_y = \sum_{i=1}^L \lambda_i u_i u_i^*$, where λ_i and u_i are the i th eigenvalue and corresponding eigenvector. Let $b \geq 0$ such that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M > b > \lambda_{M+1} = \dots = \lambda_L = \sigma_v^2$ and that the corresponding eigenvectors $\{u_i\}_{i=1}^L$ are chosen to be orthogonal. Let $P_s = [u_1 \ u_2 \ \dots \ u_M]$ and $P_0 = [u_{M+1} \ u_{M+2} \ \dots \ u_L]$, and let $U_S = P_s P_s^*$, $U_0 = P_0 P_0^*$. Then, we have the following.

- i) $(b^n I_L - R_y^n)(b^n I_L + R_y^n)^{-1}$ converges to $U := U_0 - U_S$ (as $n \rightarrow \infty$), and therefore, $U_S = (I_L - U)/2$ and $U_0 = (I_L + U)/2$.
- ii) $R_y^n (b^n I_L + R_y^n)^{-1}$ converges to U_S (as $n \rightarrow \infty$).
- iii) $b^n (b^n I_L + R_y^n)^{-1}$ converges to U_0 (as $n \rightarrow \infty$).

Proof: The proof follows directly from Theorem 1.

A rough estimate of the threshold $\lambda_{\text{thr}} = b$ can be obtained from the equation $\lambda_{\text{thr}} = (\text{Tr}(R_y)/L)$, provided that L is sufficiently large. One can see the validity of this equation from noting that $\lambda_i \geq \sigma_v$ for $i \leq L$, and hence, $\text{Tr}(R_y) = \sum_{i=1}^L \lambda_i > L\sigma_v$. Here, $\text{Tr}(A)$ denotes the trace of A .

IV. APPROXIMATED SIGNAL SUBSPACE ALGORITHMS

In [9], an approximation of the noise subspace basis is derived using rational functions. This algorithm required *a priori* knowledge of the number of impinging signals and a threshold λ_{thr} , which separates signal and noise eigenvalues. The approximated noise subspace was then given by $\{(R_y^n/\lambda_{\text{thr}}^n) + I_L\}^{-1}$. Using this approximation, the minimum norm solution was derived. However, no procedure was provided for choosing λ_{thr} .

In the next few subsections, we apply Theorem 2 to obtain approximate minimum norm, MUSIC and ESPRIT estimators.

A. Approximated Minimum Norm Algorithm

The minimum norm solution of the DOA problem as formulated in [9] is to find $w \in \mathcal{C}^M$ that solves the minimization problem

$$\text{Minimize } w^* w \text{ subject to } U_s w = 0 \text{ and } w^T e_1 = 1 \quad (9)$$

where e_1 is the first column of the $L \times L$ identity matrix, and U_s is the projection of the signal subspace. Then, the minimum norm estimator determines the M highest peaks of the function $P_{MN}(\theta)$ given by

$$P_{MN}(\theta) = \frac{1}{|a^*(\theta)U_0 e_1|^2}. \quad (10a)$$

Note that the vector e_1 in (9) can be replaced by any vector c that is not in the signal sub-space with $c^* e_1 \neq 0$ to obtain

$$P_{MN}(\theta) = \frac{1}{|a^*(\theta)U_0 c|^2}. \quad (10b)$$

Thus, by approximating U_0 as in Theorem 2, we obtain

$$P_{MN}^{(n)}(\theta) = \frac{1}{|a^*(\theta)(I_L - R_y^n(b^n I_L + R_y^n)^{-1})e_1|^2}. \quad (11)$$

Clearly, $P_{MN}^{(n)}(\theta)$ converges to $P_{MN}(\theta)$ as $n \rightarrow \infty$.

B. Approximated MUSIC

Given the projection U_s onto the signal subspace, the MUSIC estimator is expressed as

$$\begin{aligned} P_{\text{MUSIC}}(\theta) &= \frac{1}{|a^*(\theta)(I_L - U_s)a(\theta)|} \\ &= \frac{1}{|L - a^*(\theta)U_s a(\theta)|} \end{aligned} \quad (12a)$$

where the frequencies are estimated as the location of the peaks of $P_{\text{MUSIC}}(\theta)$. The signal subspace approximation of Theorem 2 ii) yields the following approximated MUSIC estimator

$$P_{b,\text{MUSIC}}^{(n)}(\theta) = \frac{1}{|L - a^*(\theta)R_y^n\{b^n I_L + R_y^n\}^{-1}a(\theta)|}. \quad (12b)$$

From applying Theorem 2, we conclude that $P_{b,\text{MUSIC}}^{(n)}(\theta)$ converges to $P_{\text{MUSIC}}(\theta)$. Note that this estimator is dependent on the parameter b . An approximated threshold b can be provided as in the following algorithm.

Algorithm 1 (Rational-MUSIC)

- (i) Choose $n \geq 1$ (normally $n \leq 5$ suffices), and compute

$$\begin{aligned} F_n^{(1)} &= \left(\left(\frac{\text{Tr}(\hat{R}_y)}{L} \right)^n I_L - \hat{R}_y^n \right) \\ &\quad \cdot \left(\left(\frac{\text{Tr}(\hat{R}_y)}{L} \right)^n I_L + \hat{R}_y^n \right)^{-1} \end{aligned}$$

and let $\bar{F}_n = (F_n^{(1)} + F_n^{(1)*})/2$.

- (ii) Compute the approximated noise subspace $\hat{U}_0 = (I_L + \bar{F}_n)/2$.
- (iii) Compute $P_{b,\text{MUSIC}}^{(n)}(\theta)$ using $P_{b,\text{MUSIC}}^{(n)}(\theta) = (1/|a^*(\theta)\hat{U}_0 a(\theta)|)$, and locate the peaks. The frequencies are estimated as the angular positions of the peaks.

C. Approximated Rational-ESPRIT

The computational cost associated with the search of the peaks of MUSIC-type algorithms is usually very demanding, particularly for large dimensions. Several methods were proposed in the literature to lower the computational requirements of the traditional subspace methods such as ESPRIT [12], UNITARY-ESPRIT [11], ROOT-MUSIC [24], and many others. In these methods, the computation is focused on estimating the frequencies by solving an eigenvalue problem or a polynomial rather than searching the whole plane or a circle. Then, the frequencies are estimated as the angular positions of the eigenvalues of a matrix formed from the powers of the correlation matrix. Normally, all MUSIC, Root-MUSIC, and Minimum norm estimators require the estimation of the noise subspace, whereas ESPRIT-type methods require the knowledge of the signal subspace. The main idea behind ESPRIT can be explained as follows. Let $A(\theta)$ be as defined in (1). Then, the signal subspace is the column space of $A(\theta)$. Generally, $A(\theta)$ is unknown; however, a basis of the signal subspace can be obtained from the most significant eigenvectors of the correlation matrix. It can easily be shown that $E_2 A(\theta) = E_1 A(\theta) S$,

where $E_i \in R^{(L-1) \times L}$ with E_1 is the identity matrix with the last row removed, whereas E_2 is the identity matrix with the first row removed, and $S = \text{diag}(z_1, z_2, \dots, z_M)$, where $z_i = e^{j(2d/\lambda) \sin(\theta_i)}$. Hence, $(E_1 A(\theta))^+ E_2 A(\theta) = S$. Here, B^+ denotes the generalized inverse of B nonsingular matrix P . This property also holds for any matrix Q whose column space is the signal subspace, in which case, $Q = A(\theta)P$ for some nonsingular matrix P . In this case, if $Q_i = E_i Q$, and $i = 1, 2$, then $Q_1^+ Q_2 = P^{-1}(E_1 A(\theta))^+ E_2 A(\theta) P = P^{-1} S P$, which is similar to S . To develop an ESPRIT based on the rational approximation of Theorem 1, let b be a threshold separating noise and signal eigenvalues; then, from the above discussion, it can be shown that

$$\begin{aligned} \lim_{n \rightarrow \infty} (E_1 R_y^n (b^n I_L + R_y^n)^{-1})^+ E_2 R_y^n (b^n I_L + R_y^n)^{-1} \\ = P_1^{-1} S P_1 \end{aligned} \quad (13)$$

for some nonsingular matrix P_1 . Thus, the directions of arrival can be estimated from the eigenvalues of the above matrix for large n . Simulations showed that $n \leq 5$ normally produced reasonable results.

V. POWER-LIKE METHODS

In the previous section, a threshold that separates the noise and signal subspaces is assumed to be known *a priori*. In this section, we derive an approximation of the signal subspace using only powers of R_y on the assumption that the number of signals is known. It will be demonstrated in the next result that if the number M is known, then the signal subspace can be approximated to any desired degree of accuracy.

Theorem 3: Let E_M be an $L \times M$ full-rank matrix, where $E_M^T U_s$ is nonsingular. Let $R_M^{(n)} = R_y^n E_M$, and set $Q_n = R_M^{(n)} (R_M^{(n)*} R_M^{(n)})^{-1} R_M^{(n)*}$. Then, Q_n converges to U_s , where U_s is the projection onto the signal subspace. Additionally, the rate of convergence is asymptotically proportional to $(\lambda_{M+1}/\lambda_M)^n$.

Proof: See Appendix B.

Theorem 2 can be considered to be a basis of many subspace algorithms, which is shown next.

A. Power-Like Estimators

Let $R_M^{(n)}$ be as in Theorem 2; then, a second approximated MUSIC is given by

$$\begin{aligned} P_{\text{MUSIC}}^{(n)}(\theta) \\ = \frac{1}{|\alpha^*(\theta) (I_L - R_M^{(n)} (R_M^{(n)*} R_M^{(n)})^{-1} R_M^{(n)*}) \alpha(\theta)|} \end{aligned} \quad (14a)$$

and one can show that

$$\lim_{n \rightarrow \infty} P_{\text{MUSIC}}^{(n)}(\theta) = P_{\text{MUSIC}}(\theta).$$

Similarly, the minimum norm estimator can be written as

$$P_{MN}^{(n)}(\theta) = \frac{1}{|\alpha^*(\theta) (I_L - R_M^{(n)} (R_M^{(n)*} R_M^{(n)})^{-1} R_M^{(n)*}) e_1|} \quad (14b)$$

which converges to $P_{MN}(\theta)$ in (10a).

The MUSIC and Min-Norm frequency estimates were obtained as the frequencies at which the M highest peaks in the peak spectrum (14) occurred.

A ROOT-MUSIC algorithm based on Theorem 3 can be described as follows:

$$P^{(n)}(z) = a \left(\frac{1}{z} \right)^* R_M^{(n)} (R_M^{(n)*} R_M^{(n)})^{-1} R_M^{(n)*} a(z) \quad (14c)$$

where $z = e^{j(2d/\lambda) \sin(\theta)}$. The M roots with largest amplitudes inside the unit circle are chosen as the signal roots.

A more efficient method that can be viewed as a modification of Theorem 3 is presented in the following MUSIC algorithm. Note that this algorithm requires a rough estimate of the number of sources.

Algorithm 2 (Power-MUSIC)

- (i) Choose n sufficiently large (usually $n \leq 5$), and let $R_y^n = [c_1^{(n)} \ c_2^{(n)} \ \dots \ c_L^{(n)}]$, where $c_i^{(n)}$ is the i th column of R_y^n .
- (ii) Choose the M largest (in magnitude) columns, and form the matrix $C_M^{(n)} = [c_{i_1}^{(n)} \ c_{i_2}^{(n)} \ \dots \ c_{i_M}^{(n)}]$, ($\|c_{i_1}\| \geq \|c_{i_2}\| \geq \dots \geq \|c_{i_M}\|$).
- (iii) Compute $Q_M^{(n)} = C_M^{(n)} (C_M^{(n)*} C_M^{(n)})^{-1} C_M^{(n)*}$.
- (iv) Compute $P_{\text{MUSIC}}^{(n)}(\theta) = 1/(L - a^*(\theta) Q_M^{(n)} a(\theta))$, and locate the peaks.

B. Power-ESPRIT Algorithm

An approximated ESPRIT algorithm can be developed based on the following theorem.

Theorem 4: Let E_1, E_2 , and S be as defined in Section IV-C. Let E_M be an $L \times M$ full rank matrix, where $E_M^T U_s$ is nonsingular; then, $(E_1 R_y^n E_M)^+ E_2 R_y^n E_M$ converges to a matrix that is similar to S .

Proof: See Appendix C.

The matrix E_M in this theorem can be replaced by the selection matrix in Algorithm 2. In this case, the limits

$$\lim_{n \rightarrow \infty} (E_1 Q_M^{(n)})^+ E_2 Q_M^{(n)} \quad (15a)$$

$$\lim_{n \rightarrow \infty} (E_1 C_M^{(n)})^+ E_2 C_M^{(n)} \quad (15b)$$

exist and are similar to S . Another version of approximated ESPRIT-type can be derived from Theorem 3 and has the form

$$\lim_{n \rightarrow \infty} (E_1 Q_n)^+ (E_2 Q_n). \quad (15c)$$

This limit exists and is similar to S .

C. Estimation Based on Lagged Covariance Matrix

As shown in [27] and [28], subspace methods can be derived based on lagged covariances. These lagged covariance matrices can be obtained by partitioning R_y . This means that if a covariance matrix of large dimension ($L \geq 2M$) is constructed so that

$$R_y(0) = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \quad (16)$$

where R_{11} is of dimension $L/2$, assuming, for simplicity, that L is even, then $R_{21} = R_{12}^*$ and is less biased by noise than R_{11} or R_{22} , especially if the noise process is an MA process, e.g., white

noise. If R_{11} is of dimension K with $L = 2K$ and $K > M$, then $R_{12} = R_y(K)$, where $R_y(K)$ is Toeplitz, which is defined as

$$R_y(K) = \begin{bmatrix} r_y(K) & r_y(K+1) & \cdots & r_y(L-1) \\ r_y(K-1) & r_y(K) & \cdots & r_y(L-2) \\ \cdots & \cdots & \cdots & \cdots \\ r_y(1) & r_y(2) & \cdots & r_y(K) \end{bmatrix} \quad (17)$$

where $r_y(m) = E\{y(k+m)y^*(k)\}$. Thus, the matrix R_{12} can be utilized to derive subspace methods for frequency and bearing information. All methods derived so far can be replicated by replacing R_y with $R_y(K)$ for some $K > M$.

D. Fast Squaring Algorithm

As can be observed from the last few sections, approximating the signal subspace requires only powers of R_y . Thus, a numerically efficient method of powering is required. The simplest method of achieving this goal is the squaring procedure in which $R_y, R_y^2, R_y^4, \dots, R_y^{2^{p_0}}$ are computed, where p_0 is a sufficiently large integer. The only problem in computing these power matrices is that R_y^n become large (overflow) if $\lambda_1 > 1$ or small (underflow) if $\lambda_1 < 1$. To alleviate this numerical problem, scaling may be applied. A stable method of generating scaled powers of complex matrices can be described as follows:

$$B_0 = \rho \frac{R_y}{\text{Tr}(R_y)} \\ B_{n+1} = \rho \frac{B_n^2}{\text{Tr}(B_n)}, \quad n = 0, 1, \dots \quad (18)$$

where ρ is a positive number slightly less than unity, and $\text{Tr}(A)$ denotes the trace of A . By a suitable choice of ρ , one can be sure of staying within the range from -1 to 1 , even with round-off error. Fortunately, it is noticed from several simulations that $p_0 = 2$ or 3 yields reasonable separation of noise and signal subspaces. The rapidity of convergence to the limit depends on the ratios $|\lambda_{M+1}/\lambda_M|$. This ratio can be made smaller in two ways: One is by initially considering R_y^2 or R_y^4 , and the other is by considering $R_y + \tau I_L$, where τ is judiciously selected. It can be shown that the best choice of τ is $\tau = (\lambda_2 + \lambda_L)/2$. One can further improve the above results by choosing a Chebyshev polynomial $S_r(\lambda)$ of degree r , which satisfies $-1 \leq S_r(\lambda) \leq 1$ for $-1 \leq \lambda \leq 1$. This implies that $-1 \leq \lambda(S_r(R_y)) \leq 1$ for $-1 \leq \lambda \leq 1$, thus guaranteeing a numerical stability.

E. Minor Subspace Computation

The results of Section IV and V can be modified so that the noise subspace is approximated rather than the signal subspace. This can be established as follows. Let $\mu > \lambda_1$, and consider the matrix $R'_y = \mu I_L - R_y$. This new matrix is also Hermitian, Toeplitz, and centro-symmetric positive definite, and its eigenvalues, in decreasing order, are $\mu - \lambda_L \geq \mu - \lambda_{L-1} \cdots \geq \mu - \lambda_1$, i.e., $\mu - \lambda_L$ is the largest eigenvalue of R'_y . Hence, we can apply the same process above to find the noise subspace, in which

case, the MUSIC and minimum norm estimators have the forms

$$f_{\text{MUSIC}}^{(n)}(\theta) = \frac{1}{|a^*(\theta)(R_y'^n(b^n I_L + R_y'^n)^{-1})a(\theta)|} \quad (19)$$

and

$$f_{MN}^{(n)}(\theta) = \frac{1}{|a^*(\theta)(R_y'^n(b^n I_L + R_y'^n)^{-1})e_1|}. \quad (20)$$

The parameter b is as defined in Theorem 2. The number μ should be chosen carefully to keep appreciable separation between noise and signal subspaces to allow the rational and power methods to extract one of these subspaces. One way to choose μ is $\mu = \text{trace}(R_y)$. A more accurate approximation can be obtained by solving $\mu^n = \text{trace}(R_y^n)$. A sharper estimate of μ comes from $\mu^n = \text{trace}(R_y^n)/L$.

The approximated power-ESPRIT based on R'_y is obtained from the following limit:

$$\lim_{n \rightarrow \infty} E_1(b^n(b^n I_L + R_y'^n)^{-1})^{-1} E_2(b^n(b^n I_L + R_y'^n)^{-1}) \quad (21)$$

which exists and is similar to S .

VI. OPERATION COUNT

The methods presented in the previous sections are multiplication rich in that for a given n , \hat{R}_y^n is required and is followed by a matrix inversion. Matrix multiplication can be obtained very efficiently by applying the Strassen algorithm [26]. In this algorithm, if $A \in R^{L \times L}$ and $B \in R^{L \times L}$ with L is a power of 2, then $C = AB$ can be obtained with $s \approx L^{2.807}$ multiplications. Thus, asymptotically, the number of multiplications in the Strassen algorithm is $O(L^{2.807})$ compared with $O(L^3)$ in the conventional method. It should be mentioned that in [29], Bailey implemented a Strassen approach that required only 60% of the time needed by the conventional multiplication.

The number of flops in computing $(b^n I_L - \hat{R}_y^n)^{-1}(b^n I_L + \hat{R}_y^n)$ consists of approximately the number of flops in computing \hat{R}_y^n and the matrix inverse. Assuming that $n = 2^r$, both of these processes cost about $rL^{2.807} + (2L^3/3)$.

The number of flops required to compute the SVD of \hat{R}_y by the Golub-Reinsch algorithm is $21L^3$ [30]. For example, if we choose r to be 4, which corresponds to $n = 16$, which is a value that is very high in most applications, the number of flops required in the rational MUSIC is $4L^{2.807} + (2L^3/3)$, which is still much less than $21L^3$ using the Golub-Reinsch algorithm [30].

Efficient matrix inversion can be computed using the LU decomposition. Once the LU factorization of A is known, the inverse matrix A^{-1} can be computed in $L(L-1)(2L-1)/3$ flops [26]. Thus, the total number of flops involved in computing $(b^n I_L - \hat{R}_y^n)^{-1}(b^n I_L + \hat{R}_y^n)$ is about $(4L^3/3) + 2rL^3 = (2r + 1.333)L^3$. This number is still far less than the flop count for computing the SVD, which is about $21L^3$, for $r < 10$. Note that $r = 9$ corresponds to $n = 512$, which is extremely large for most applications. Thus, for all practical purposes, these algorithms, which are based on Theorem 2, are less costly than the truncated SVD-based methods. In addition, in many applications, the matrices under consideration have other structures, e.g., Toeplitz, Hankel, block Toeplitz, or block Hankel, and further saving of computation can be achieved.

VII. SIMULATION RESULTS

In this section, the frequency estimators described in Sections IV and V were examined on several data sets generated by the equation

$$y(n) = d_1 e^{j(2\pi f_1 n + \phi_1)} + d_2 e^{j(2\pi f_2 n + \phi_2)} + v(n) \quad (22)$$

where $d_1 = 1.0$, $d_2 = 1.0$, $f_1 = 0.5$, $f_2 = 0.52$, and $n = 1, 2, \dots, N = 25$. The ϕ_i are independent random variables uniformly distributed over the interval $[-\pi, \pi]$. The noise $v(n)$ is assumed to be white and uncorrelated with the signal. Note that $f_2 - f_1 < (1/N)$. The SNR for either sinusoid is defined as $10 \log_{10}(\sigma_x^2/\sigma_v^2)$, where $x(n) = d_1 e^{j(2\pi f_1 n + \phi_1)} + d_2 e^{j(2\pi f_2 n + \phi_2)}$, and σ_x^2 and σ_v^2 are the variances of $x(n)$ and $v(n)$, respectively. The covariance matrix is constructed using forward-backward method described in Section II to increase robustness. The size of the covariance matrix is chosen to be $L = 10$, which, in the absence of noise, has effective rank two. We performed experiments to compare the proposed methods versus the truncated SVD-based MUSIC. The SVD routine in MATLAB is used for the computation of the signal subspace eigenvectors and eigenvalues required to implement an SVD-based method for comparison. We varied SNR from 10 to 20 in 5-dB steps and estimated the frequencies for data length 25. For each experiment (with data length and SNR fixed), we performed 100 independent trials to estimate the frequencies. We use the following performance criterion (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N_e} \sum_{i=1}^{N_e} (\hat{f}_i - f_{\text{true}})^2}$$

to compare the results. Here, N_e is the number of independent realizations, and \hat{f}_i is the estimate provided from the i th realization. Several experiments were conducted to test the performance of Algorithms 1 and 2 and the SVD-based MUSIC. The mean values of estimated frequencies and their RMSE of the SVD-based MUSIC are given in Table II. The simulation results of applying Algorithms 1 and 2 are summarized as follows. First, Algorithm 1 is applied for a set of 100 random experiments for different (SNR = 20, 15, and 10 dB) using R_y^n with $n = 3$. The mean values and standard deviations of the estimated frequencies are given in Table I. The threshold b in these simulations is estimated by $b = (\text{Tr}(R_y)/L)$. The peak spectrum in each trial was computed using 1000 frequency bins covering a normalized frequency range of 0 to 1. Each spectrum distribution is scaled so that its maximum equals 1. The frequency estimates were obtained as the frequencies at which the two highest peaks in the peak spectrum occurred. The mean and RMSE are taken only over realizations where two peaks have occurred. Table I shows the estimated frequencies resulting from applying Algorithm 1 for the case in which SNR = 15 dB and using R_y^n for $n = 3$. In this table, the spectrum distribution of 100 trials were included. As can be seen from Tables I and II, both Algorithm 1 and the standard MUSIC have virtually identical performance.

Next, we repeated the experiments in part 1 using Algorithm 2. The results of testing this algorithm for different SNR were

TABLE I
MEAN AND RMSE OF FREQUENCIES FOR DATA OF TWO COMPLEX SINUSOIDS AT FREQUENCIES 0.50 AND 0.52 IN NOISE WITH SNR = 20, 15, AND 10, dB, DIMENSION OF DATA VECTORS $L = 10$. ALGORITHM 1 IS USED

SNR	f_1	f_2	$RMSE_{f_1}$	$RMSE_{f_2}$
20 dB	0.500556	0.521225	0.00563	0.012522
15 dB	0.500831	0.521748	0.00652	0.018531
10 dB	0.501025	0.523093	0.01126	0.032547

TABLE II
MEAN AND RMSE OF FREQUENCIES FOR DATA OF TWO COMPLEX SINUSOIDS AT FREQUENCIES 0.50 AND 0.52 IN NOISE WITH SNR = 20, 15, AND 10 dB, DIMENSION OF DATA VECTORS $L = 10$, $M = 2$. SVD-BASED METHOD IS USED

SNR	f_1	f_2	$RMSE_{f_1}$	$RMSE_{f_2}$
20 dB	0.500226	0.52133	0.004420	0.004142
15 dB	0.500762	0.52182	0.007814	0.008461
10 dB	0.501239	0.52214	0.013078	0.010458

TABLE III
MEAN AND RMSE OF FREQUENCIES FOR DATA OF TWO COMPLEX SINUSOIDS AT FREQUENCIES 0.50 AND 0.52 IN NOISE WITH SNR = 20, 15, AND 10 dB, DIMENSION OF DATA VECTORS $L = 10$, $M = 2$. ALGORITHM 2 IS USED

SNR	f_1	f_2	$RMSE_{f_1}$	$RMSE_{f_2}$
20 dB	0.50115	0.52098	0.00763	0.00883
15 dB	0.50182	0.52143	0.00827	0.01445
10 dB	0.50195	0.52192	0.01281	0.01205

averaged over 100 trials, and the mean and RMSE of each frequency was presented in Table III. At high SNR, it has been observed that each of the 100 spectra has two peaks near the true frequencies; however, at low SNR, some of the peak spectra displayed only one peak near $(f_1 + f_2)/2$. It is also noticed that the rate of joint detection of the two frequencies is identical to the SVD-based method. Clearly, Table III shows very good resolution of the two frequencies using a lower power of $n = 3$.

In the same experiments, the projection onto the dominant signal subspace is approximated for the 100 random experiments using Algorithm 2. Comparison between signal subspaces are made by using the eigenvalues of the matrix $Q_M^{(n)*} Q_M$. Here, $Q_M^{(n)}$ is obtained as in Algorithm 2, and Q_M is obtained from the SVD of the covariance matrix. The eigenvalues of this matrix are the cosine of the angles between the vectors in the two subspaces. These eigenvalues are found to be close to 1.

We also performed experiments to investigate the effect of the power n on the performance of Algorithm 2. It is observed that increasing n does not significantly effect the RMSE as long as $n \geq 3$.

Finally, we would like to comment on the performance of the developed algorithms with respect to overestimation of M . When n is small, it is observed that overestimation of M leads to better estimation of the frequencies. The robustness of the methods against overestimation of the number of sources M can be explained as follows. Overestimation of M means that additional vectors are included in the basis of the signal subspace. In our approximation, these vectors are not purely noise but contain some signal component. In the standard MUSIC, if additional vector is added to the basis, a purely noise vector is included in the signal subspace causing spurious peaks. Note

that when n is small, the first few signal vectors contain noise components since the noise and signal vectors are not well separated. Thus, we have to consider more vectors to capture the signal subspace. However, this separation becomes more prominent as n increases.

VIII. CONCLUSION

Several approaches for approximating the dominant and subdominant subspaces have been developed that avoid the costly eigendecomposition or SVD. More generally, we provided a way of splitting an L -dimensional space into several complementary invariant subspaces of the covariance matrix R_y without actually computing any eigenvalues. Frequency estimators such as MUSIC-, Minimum-Norm-, and ESPRIT-type are then derived using these approximated subspaces. The computation of obtaining these approximate subspaces and estimators are shown to be less than the standard techniques. Even though these methods are introduced only as approximations, they perform well even at low SNR. The good performance of these approximate methods is due to the fact that the estimated signal and noise vectors obtained using the proposed algorithms are not purely signal or noise vectors, whereas methods based on exact eigendecomposition treat some of these vectors as purely signal or purely noise vectors, which is not true especially at low SNR, resulting in undesirable effects. The main features of these methods are that they are computationally simple and easy to implement in that only matrix inversion and powers of the sample covariance matrix are required; nonetheless, a comparable performance to high-resolution exact eigenvector methods has been achieved.

APPENDIX A

PROOF OF THEROEM 1

From the eigendecoposition of R_y , we have

$$(b_r^n I_L - R_y^n)(b_r^n I_L + R_y^n)^{-1} = \sum_{i=1}^L \frac{b_r^n - \lambda_i^n}{b_r^n + \lambda_i^n} u_i u_i^*$$

and

$$\frac{b_r^n - \lambda_i^n}{b_r^n + \lambda_i^n} = \begin{cases} 1, & \text{if } \lambda_i < b_r \\ -1, & \text{if } \lambda_i > b_r. \end{cases} \quad (\text{A1})$$

Note that the matrices $\{U_i\}_{i=1}^S$ defined in Theorem 1 are projections and that $U_i U_j = U_i \delta(i - j)$, where δ is the Kronecker delta function. Hence, by virtue of $U_i U_j = U_i \delta(i - j)$

$$\begin{aligned} & \Pi_{r=1}^S (b_r^n I - R^n)(b_r^n I + R^n)^{-1} \\ & \rightarrow \Pi_{r=0}^S \left(\sum_{i=0}^r U_i - \sum_{i=r+1}^S U_i \right) \\ & = \left(U_0 - \sum_{i=1}^S U_i \right) (\cdots) \left(\sum_{i=0}^r U_i - \sum_{i=r+1}^S U_i \right) (\cdots) \\ & \left(\sum_{i=0}^S U_i - U_S \right) = U_0 - U_1 + U_2 + \cdots + (-1)^S U_S. \end{aligned}$$

Q.E.D.

APPENDIX B

PROOFS OF THEOREMS 3 AND 4

The proof of Theorems 3 and 4 follow directly from the following two results.

Lemma 5: Let $A, C, B \in \mathcal{C}^{L \times L}$ be diagonalizable matrices such that $|\lambda_{\min}(A)| > |\lambda_{\max}(B)|$; then, $(A^n)^{-1} C B^n \rightarrow 0$ as $n \rightarrow \infty$. ($\lambda_{\min}(A)$, and $\lambda_{\max}(A)$ denote the minimum and maximum eigenvalues of A).

Proof: We will only prove this Lemma for diagonalizable matrices. Let $A = \sum_{i=1}^L \lambda_i \mathbf{q}_i \mathbf{p}_i^*$ and $B = \sum_{j=1}^L \mu_j \mathbf{u}_j \mathbf{v}_j^*$ with $\mathbf{p}_i^* \mathbf{q}_j = \mathbf{u}_j^* \mathbf{v}_j = \delta_{i-j}$. Then

$$\begin{aligned} (A^n)^{-1} C B^n &= \left\{ \sum_{i=1}^L \lambda_i^{-n} \mathbf{q}_i \mathbf{p}_i^* \right\} C \left\{ \sum_{j=1}^L \mu_j^n \mathbf{u}_j \mathbf{v}_j^* \right\} \\ &= \sum_{i=1}^L \sum_{j=1}^L \left(\frac{\mu_j}{\lambda_i} \right)^n \mathbf{q}_i \mathbf{p}_i^* C \mathbf{u}_j \mathbf{v}_j^* \end{aligned} \quad (\text{B1})$$

which converges to 0 since $|\mu_j| < |\lambda_i|$ for $i, j = 1, \dots, L$. Q.E.D.

The last result holds for nondiagonalizable matrices and can be shown using a continuity argument.

Theorem 6: Let \hat{R}_y be a positive definite matrix such that $\hat{R}_y = Q \Lambda Q^*$, where $Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}$ is an orthogonal matrix, and $\Lambda = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix}$ is diagonal. Here, Q_{11} , Q_{12} , Q_{21} , and Q_{22} are $M \times M$, $M \times (L - M)$, $(L - M) \times M$, and $(L - M) \times (L - M)$ complex matrices, respectively. The matrices Λ_1 and Λ_2 are diagonal matrices such that $\Lambda_1 = \text{diag}(\lambda_1, \dots, \lambda_M)$ and $\Lambda_2 = \text{diag}(\lambda_{M+1}, \dots, \lambda_L)$, with $|\lambda_{\min}(\Lambda_1)| > |\lambda_{\max}(\Lambda_2)|$. Assume, without loss of generality, that the matrix Q_{11} is nonsingular. Let $\hat{R}_y = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$, and for each $n \geq 1$, let $\hat{R}_y^n = \begin{bmatrix} R_{11}(n) & R_{12}(n) \\ R_{21}(n) & R_{22}(n) \end{bmatrix}$, where R_{ij} and Q_{ij} have the same dimensions; then

$$\lim_{n \rightarrow \infty} R_{21}(n) R_{11}(n)^{-1} = Q_{21} Q_{11}^{-1} \quad (\text{B2})$$

and

$$\lim_{n \rightarrow \infty} R_{11}(n)^{-1} R_{12}(n) = Q_{11}^{-*} Q_{12}^*. \quad (\text{B3})$$

Proof: The decomposition

$$\hat{R}_y = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}^* \quad (\text{B4})$$

is possible since \hat{R}_y is Hermitian. From the relation B4, we obtain

$$\begin{aligned} \hat{R}_y^n &= \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} \Lambda_1^n & 0 \\ 0 & \Lambda_2^n \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}^* \\ &= \begin{bmatrix} Q_{11} \Lambda_1^n Q_{11}^* + Q_{12} \Lambda_2^n Q_{12}^* & Q_{11} \Lambda_1^n Q_{21}^* + Q_{12} \Lambda_2^n Q_{22}^* \\ Q_{21} \Lambda_1^n Q_{11}^* + Q_{22} \Lambda_2^n Q_{12}^* & Q_{21} \Lambda_1^n Q_{21}^* + Q_{22} \Lambda_2^n Q_{22}^* \end{bmatrix}. \end{aligned}$$

Since $|\lambda_{\min}(\Lambda_1)| > |\lambda_{\max}(\Lambda_2)|$, it follows from Lemma 4 that

$$\begin{aligned} & \lim_{n \rightarrow \infty} R_{21}(n)R_{11}(n)^{-1} \\ &= \lim_{n \rightarrow \infty} (Q_{21}\Lambda_1^n Q_{11}^* + Q_{22}\Lambda_2^n Q_{12}^*) \\ & \quad \cdot (Q_{11}\Lambda_1^n Q_{11}^* + Q_{12}\Lambda_2^n Q_{12}^*)^{-1} \\ &= \lim_{n \rightarrow \infty} (Q_{21}\Lambda_1^n Q_{11}^*) (Q_{11}\Lambda_1^n Q_{11}^*)^{-1} = Q_{21}Q_{11}^{-1}. \end{aligned}$$

Similarly, to prove (B3), we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} R_{11}(n)^{-1}R_{12}(n) \\ &= \lim_{n \rightarrow \infty} (Q_{11}\Lambda_1^n Q_{11}^* + Q_{12}\Lambda_2^n Q_{12}^*)^{-1} \\ & \quad \cdot (Q_{11}\Lambda_1^n Q_{21}^* + Q_{12}\Lambda_2^n Q_{22}^*) \\ &= \lim_{n \rightarrow \infty} (Q_{11}\Lambda_1^n Q_{11}^*)^{-1} (Q_{11}\Lambda_1^n Q_{21}^*) = Q_{11}^{-1}Q_{21}^*. \end{aligned}$$

Q.E.D.

Proof of Theorem 3: Let $E_M = [W_1^T, W_2^T]^T$ with $W_1 \in \mathcal{C}^{M \times M}$ and $W_2 \in \mathcal{C}^{(L-M) \times M}$ such that $Q_{11}W_1$ is nonsingular, and set $\hat{R}_y^n W = \begin{bmatrix} R_1(n) \\ R_2(n) \end{bmatrix}$, where $R_1(n) \in \mathcal{C}^{M \times M}$, and $R_2(n) \in \mathcal{C}^{(L-M) \times M}$. Then, applying the same ideas of the Proof of Theorem 5, we obtain

$$\lim_{n \rightarrow \infty} R_1(n)^{-1}R_2(n) = Q_{21}Q_{11}^{-1}.$$

By simple algebraic manipulation, we obtain

$$\begin{aligned} & \lim_{n \rightarrow \infty} \begin{bmatrix} R_1(n) \\ R_2(n) \end{bmatrix} (R_1^*(n)R_1(n) + R_2^*(n)R_2(n))^{-1} \\ & \quad \cdot \begin{bmatrix} R_1^*(n) & R_2^*(n) \end{bmatrix} \\ &= \lim_{n \rightarrow \infty} \begin{bmatrix} R_1(n) \\ R_2(n) \end{bmatrix} \{R_1^{-1}(n) \{I_L + R_1^{-*}(n)R_2^*(n) \\ & \quad \cdot R_2(n)R_1(n)^{-1}\} R_1(n)\}^{-1} \begin{bmatrix} R_1^*(n) & R_2^*(n) \end{bmatrix} \\ &= \lim_{n \rightarrow \infty} \begin{bmatrix} I_L \\ R_2(n)R_1^{-1}(n) \end{bmatrix} (I_L + R_1(n)^{-*}R_2^*R_2(n) \\ & \quad \cdot R_1^{-1}(n))^{-1} \begin{bmatrix} I_L & R_1^{-*}(n)R_2^*(n) \end{bmatrix} \\ &= \begin{bmatrix} I_L \\ Q_{21}Q_{11}^{-1} \end{bmatrix} (I_L + Q_{11}^{-*}Q_{21}^*Q_{21}Q_{11}^{-1})^{-1} \begin{bmatrix} I_L & Q_{11}^{-*}Q_{21}^* \end{bmatrix} \\ &= \begin{bmatrix} Q_{11} \\ Q_{21} \end{bmatrix} (Q_{11}^*Q_{11} + Q_{21}^*Q_{21})^{-1} \begin{bmatrix} Q_{11}^* & Q_{21}^* \end{bmatrix} = Q_1^*Q_1 \end{aligned}$$

where $Q_1 = \begin{bmatrix} Q_{11} \\ Q_{21} \end{bmatrix}$ is defined in Theorem 5. Note that the third equality follows from Theorem 6.

Q.E.D.

Proof of Theorem 4: Since the columns of each of $A(\theta)$ and Q_1 span the signal subspace, it follows that $\begin{bmatrix} Q_{11} \\ Q_{21} \end{bmatrix} = A(\theta)P$ for some nonsingular matrix P . Therefore, for large m , the matrix $\hat{R}_y^n W$ can be approximated as

$$\begin{aligned} \hat{R}_y^n W &\approx \begin{bmatrix} Q_{11} \\ Q_{21} \end{bmatrix} \Lambda_1^n (Q_{11}^* W_1 + Q_{21}^* W_1) \approx Q_1 \Lambda_1^n Q_1^* W \\ &\approx A(\theta)P \Lambda_1^n Q_1^* W. \end{aligned} \quad (\text{B5})$$

The nonsingularity of $Q_1^* W$ guarantees that the matrix $R_1(n)$ is nonsingular for all sufficiently large m (see Section IV-C). Hence, from (B5), we also obtain

$$E_2 A(\theta) P_n (E_1 A(\theta) P_n)^+ = E_2 A(\theta) (E_1 A(\theta))^+$$

which is similar to S.

Q.E.D.

Remark: One of the other important aspects of Theorem 6 is that the dominant eigenvectors can be extracted using only the elements of \hat{R}_y^n for sufficiently large n . The only problem is that there is no guarantee that Q_{11} is invertible. This can be overcome by permuting the rows of \hat{R}_y^n . Another interpretation of Theorem 5 is that a basis of the signal subspace can be approximated by applying the Gram–Schmidt process on the columns of $\begin{bmatrix} I_L \\ R_{21}(n)R_{11}(n)^{-1} \end{bmatrix}$.

REFERENCES

- [1] D. W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids; making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, pp. 975–989, Sept. 1982.
- [2] P. Stoica, T. Soderstrom, and F. Ti, "Asymptotic properties of the high-order yule-walker estimates of sinusoidal frequencies," *IEEE Trans. Signal Processing*, vol. 37, pp. 1721–1734, Nov. 1989.
- [3] V. F. Pisarenko, "On the estimation of spectra by means of nonlinear functions of the covariance matrix," *Geophys. J. Roy. Astron. Soc.*, vol. 28, pp. 511–531, 1972.
- [4] J. A. Cadzow, "Spectral estimation: An overdetermined rational model equation approach," *Proc. IEEE*, vol. 70, pp. 907–939, Sept. 1982.
- [5] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 276–280, 1986.
- [6] D. H. Johnson and S. R. DeGraaf, "Improving the resolution of bearing in passive sonar arrays by eigenvalue analysis," *IEEE Trans. Signal Processing*, vol. ASSP-30, pp. 638–647, Aug. 1982.
- [7] J. A. Cadzow and O. M. Solomon, "Algebraic approach to system identification," *IEEE Trans. Signal Processing*, vol. ASSP-34, pp. 462–468, June 1986.
- [8] J. A. Cadzow, "Signal enhancement—A composite property mapping algorithm," *IEEE Trans. Signal Processing*, vol. 36, pp. 49–62, Jan. 1988.
- [9] V. T. Ermolaev and A. B. Gershman, "Fast algorithm for minimum-norm direction-of-arrival estimation," *IEEE Trans. Signal Processing*, vol. 42, pp. 2389–2394, Sept. 1994.
- [10] A. B. Gershman, "Direction of arrival estimation using generalized minimum norm approach," *Electron. Lett.*, vol. 27, no. 16, pp. 1485–1486, Aug. 1991.
- [11] M. Haardt and J. A. Nossek, "Unitary ESPRIT: How to obtain increased estimation accuracy with a reduced computational burden," *IEEE Trans. Signal Processing*, vol. 43, pp. 1232–1242, May 1995.
- [12] R. H. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoustic, Speech, Signal Processing*, vol. ASSP-37, pp. 984–995, July 1989.
- [13] R. H. Roy, A. Paulraj, and T. Kailath, "ESPRIT—A subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1340–1342, 1986.
- [14] H. Park, "ESPRIT direction-of-arrival estimation in the presence of spatially correlated noise," *SIAM J. Matrix Anal. Appl.*, vol. 15, no. 1, pp. 185–193, Jan. 1994.
- [15] Y. Hua and T. K. Sarkar, "On SVD for estimating generalized eigenvalues of singular matrix pencils in noise," *IEEE Trans. Signal Processing*, vol. 39, pp. 892–899, Apr. 1991.
- [16] A. Eriksson, P. Stoica, and T. Soderstrom, "Markov-based eigenanalysis method for frequency estimation," *IEEE Trans. Signal Processing*, vol. 42, pp. 586–594, Mar. 1994.
- [17] S. M. Kay and A. K. Shaw, "Frequency estimation by principal component AR spectral estimation method without eigendecomposition," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 95–101, Jan. 1988.

- [18] J. T. Karhunen and J. Joutsenalo, "Sinusoidal frequency estimation by signal subspace approximation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 40, pp. 2961–2972, Dec. 1992.
- [19] C. E. Davila and M. Azmoodeh, "Efficient estimation of the signal subspace without eigendecomposition," *IEEE Trans. Signal Processing*, vol. 42, pp. 236–239, Jan. 1994.
- [20] D. Tufts and C. D. Melissinos, "Simple, effective computation of principal eigenvectors and their eigenvalues and application to high-resolution estimation of frequencies," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1046–1053, Oct. 1986.
- [21] H. B. Lee and M. S. Wengrovitz, "Resolution threshold of beamspace MUSIC for two closely spaced emitters," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 38, pp. 1545–1559, Sept. 1990.
- [22] K. M. Buckley and X.-L. Xu, "Spatial-spectrum estimation in a location sector," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 38, pp. 1842–1852, Nov. 1990.
- [23] B. Porat, *Digital Processing of Random Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
- [24] D. B. Rao and K. V. S. Hari, "Performance analysis of root-music," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 37, pp. 1939–1949, Dec. 1989.
- [25] B. S. Garbow, J. M. Boyle, J. J. Dongarra, and C. B. Moler, *Matrix Eigensystem Routines—EISPACK Guide Extension*, G. Goos and J. Hortmanis, Eds. New York: Springer-Verlag, 1977.
- [26] G. H. Golub and C. G. Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: John Hopkins Univ. Press, 1989.
- [27] M. A. Hasan and A. A. Hasan, "Hankel matrices of finite rank with applications to signal processing and polynomials," *J. Math. Anal. Appl.*, vol. 208, pp. 218–242, 1997.
- [28] M. A. Hasan and M. R. Azimi-Sadjadi, "Separation of multiple time delays using new spectral estimation schemes with applications to underwater target detection," *IEEE Trans. Signal Processing*, vol. 46, pp. 1580–1590, June 1998.
- [29] D. Bailey, "Extra high speed matrix multiplication on Cray-2," *SIAM J. Sci. Statist. Comput.*, vol. 9, pp. 603–607, 1988.
- [30] G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," *Numer. Math.*, vol. 14, pp. 403–420, 1970.



Mahmood R. Azimi-Sadjadi (SM'89) received the M.Sc. and Ph.D. degrees from the Imperial College of Science and Technology, University of London, London, U.K., in 1978 and 1982, respectively, both in electrical engineering with specialization in digital signal/image processing.

He is currently a Full Professor with the Department of Electrical and Computer Engineering, Colorado State University (CSU), Fort Collins. He is also serving as the Director of the Multisensory Computing Laboratory (MUSCL), CSU. His main areas of interest include digital signal and image processing; target detection, classification, and tracking; adaptive filtering and system identification; and neural networks. His research interests in these areas have resulted in more than 100 journal and refereed conference publications. He is co-author of the book *Digital Filtering in One and Two Dimensions* (New York: Plenum, 1989). He served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and is currently serving as an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS.

Dr. Azimi-Sadjadi was the recipient of the 1999 ABELL Teaching Award, the 1993 ASEE-Navy Senior Faculty Fellowship Award, the 1991 CSU Dean's Council Award, and the 1984 Dow Chemical Outstanding Young Faculty Award.



Ali A. Hasan received the B.S. and M.Sc. degrees from the University of Technology, Baghdad, Iraq, in 1981 and 1985, respectively, both in electrical engineering.

He is currently with the Faculty of College of Electronic Engineering, University of Nasser, Bani Waleed, Libya, where he teaches optimal control, multivariable control systems, electromechanical systems, and system identification and adaptive control. His research interests include control theory, artificial intelligence, fuzzy logic, adaptive systems,

and estimation theory.



Mohammed A. Hasan (M'98) received the Ph.D. degrees in mathematics and electrical engineering from Colorado State University, Fort Collins, in 1991, and 1997, respectively.

He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, University of Minnesota, Duluth. His research interests include adaptive systems, signal/image processing, estimation theory, control theory, numerical analysis, optimization, numerical linear algebra, and computational and applied mathematics. His current

research includes high-resolution methods, fast algorithms, parallelizable inverse-free algorithms, subspace decompositions and their applications in signal processing, and sinusoidal frequency estimation. One of his educational interests is the use of microprocessor DSP chips in teaching signal processing and signals and systems courses.