Min-Norm Interpretations and Consistency of MUSIC, MODE and ML

Yariv Ephraim, *Fellow, IEEE,* Neri Merhav, *Senior Member, IEEE,* and Harry L. Van Trees, *Life Fellow, IEEE*

***Abstract*— The multiple signal characterization (MUSIC) approach, its generalization to correlated signals known as the method of direction estimation (MODE), and the deterministic maximum likelihood (ML) approach for bearing estimation in array processing are shown to be signal subspace fitting approaches in a minimum norm sense. MODE, for example, is shown to be an approach in which the array manifold is linearly estimated from principal empirical eigenvectors in a minimum weighted Frobenius norm sense. Using the min-norm interpretations, a unified proof for strong consistency of the three approaches is provided for stationary and ergodic signals.**

I. Introduction

ESTIMATION of bearing parameters of a set of signals impinging on an array of sensors has attracted much attention in the past two decades. Two estimation approaches have dominated the field. The first is the multiple signal characterization (MUSIC) approach developed in [1]. The second is the classical maximum likelihood (ML) parameter estimation approach under Gaussian assumptions considered in [2] and [3]. A third approach, known as the method of direction estimation (MODE) [4, eq. (3.6)], constitutes a generalization of MUSIC to bearing estimation of correlated signals. The three approaches have been extensively studied, see, e.g., [1]–[14] and the references therein. They are also the subject of this paper.

The theoretical basis of MUSIC is that under ideal conditions, when the exact covariance matrix of the noisy signals is known, and the noise is white, each steering vector is orthogonal to the noise subspace spanned by the nonprincipal eigenvectors of the covariance of the noisy signals. MUSIC is implemented by replacing the unknown covariance matrix of the noisy signals with its empirical estimate. The deterministic ML approach assumes that the signals are deterministic unknown and the noise is Gaussian. Hence, joint ML estimation of the signals and their bearings is performed. Since the signals are nuisance "parameters" in this formulation, deterministic ML has been controversial. Indeed, it leads to inefficient bearing estimates for any finite number of sensors [6, theorem 5.2], [8, corollary 1]. The cost function of the MODE approach was originally derived in [5, theorem 6.1] as an asymptotic version of the deterministic likelihood function. Nevertheless, a version of the MODE approach was shown in [7] to be asymptotically efficient relative to the stochastic Cramer–Rao bound [14].

In this study new nonasymptotic interpretations of MODE and ML as signal subspace fitting approaches in a minimum norm sense are provided. These interpretations reveal the estimation criterion under which each approach is optimal. For example, MODE is shown to be an approach in which the array manifold is *linearly* estimated from principal empirical eigenvectors in the minimum weighted Frobenius norm sense. The weighting matrix is the covariance of the signals.

Our signal subspace fitting approach differs from that of Viberg and Ottersten [8]. Their approach was applied to the deterministic ML, ESPRIT, and multidimensional MUSIC (MD-MUSIC). In this approach the empirical data is linearly fitted by the array manifold in the minimum Frobenius norm sense. The definition of empirical data depends on the approach being interpreted. For example, in interpreting MD-MUSIC the data is represented by the matrix of empirical principal eigenvectors. Our signal subspace fitting approach does the opposite in interpreting MODE. While they fit the data by the model we estimate the model from the data. Note that unlike MODE, MD-MUSIC does not reduce to MUSIC when the signals are uncorrelated. MD-MUSIC degenerates to MUSIC when there is a single source and the norm of the steering vector is independent of the bearing parameters.

Using the min-norm signal subspace fitting interpretations, a unified proof for strong consistency of MUSIC, MODE and ML is developed. This proof is applicable to any stationary and ergodic correlated signals. The proof uses elements from existing proofs given in [5] and [8], but it appears to be more complete. Note that the proof could not be inferred as a particular case of the analysis in [6] since that analysis was performed under Gaussian assumptions. We do not make this assumption. Furthermore, the proof could not be inferred from the results in [8] which are concerned only with asymptotic versions of MD-MUSIC and ML. These criteria do not coincide with MODE or ML.

II. Preliminaries

Assume M plane wave sources impinging on an array of L sensors with arbitrary but known geometry. Let θ_m denote the vector of bearing parameters of the mth source. Let

Manuscript received May 18, 1993; revised June 19, 1995. The associate editor coordinating review of this paper and approving it for publication was Prof. Douglas Williams.

Y. Ephraim was with the Speech Research Department, AT&T Bell Laboratories, Murray Hill, NJ. He is now with the Department of Electrical and Computer Engineering and the C^3I Center of Excellence, George Mason University, Fairfax, VA 22030 USA.

H. L. Van Trees is with the Department of Electrical and Computer Engineering and the C^3I Center of Excellence, George Mason University, Fairfax, VA 22030 USA.

N. Merhav is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa, Israel.

IEEE Log Number 9415279.

$\Theta \triangleq (\theta_1, \cdots, \theta_M)$ denote the matrix of bearing parameters from the different sources. Let $S(t) \triangleq (s_1(t), \cdots, s_M(t))^T$ denote the $M \times 1$ vector of signals from the different sources at time t where $(\cdot)^T$ denotes vector transpose. Similarly, let $V(t) \triangleq (v_1(t), \cdots, v_L(t))^T$ and $Y(t) \triangleq (y_1(t), \cdots, y_L(t))^T$ denote the $L \times 1$ vectors of noise and received signals at the L sensors, respectively. For narrow band signals, $S(t), V(t)$ and $Y(t)$ represent complex envelopes and

$$Y(t) = D(\Theta)S(t) + V(t) \tag{1}$$

where $D(\Theta)$ represents an $L \times M$ matrix of steering vectors $\{d(\theta_m)\}$ or the array manifold. It is assumed that $M < L$ and that any set of M distinct steering vectors are linearly independent. Thus, the rank of $D(\Theta)$ is M. Furthermore, as usual, the noise is assumed spatially and temporarily white with variance σ_v^2.

When the signals are assumed random, such as in the MUSIC and MODE approaches, the covariance matrix of $Y(t)$ is given by

$$R(\Theta) = D(\Theta)PD^{\#}(\Theta) + \sigma_v^2 I \tag{2}$$

where P denotes the covariance matrix of the signal vector $S(t)$ and it is assumed positive definite, and $(\cdot)^{\#}$ denotes conjugate transpose. Since the noise is assumed spatially white, $R(\Theta)$, $D(\Theta)PD^{\#}(\Theta)$, and $\sigma_v^2 I$ have the same set of eigenvectors. Since $\mathrm{rank}\{D(\Theta)\} = M < L$, the covariance matrix $D(\Theta)PD^{\#}(\Theta)$ has $L - M$ zero eigenvalues. In addition, since all eigenvalues of $\sigma_v^2 I$ equal σ_v^2, the covariance matrix $R(\Theta)$ has $L - M$ eigenvalues such that each equals to σ_v^2.

Let $U(\Theta) = [u_1(\Theta), \cdots, u_L(\Theta)]$ and $\Lambda(\Theta) = \mathrm{diag}(\lambda_1(\Theta), \cdots, \lambda_L(\Theta))$ denote the matrices of eigenvectors and eigenvalues of $R(\Theta)$, respectively. Assume that $\{\lambda_l(\Theta)\}$ are given in a descending order. Let $U(\Theta) = [U_1(\Theta), U_2(\Theta)]$ where $U_1(\Theta) = [u_1(\Theta), \cdots, u_M(\Theta)]$ denotes the $L \times M$ matrix of eigenvectors of $R(\Theta)$ whose corresponding eigenvalues are greater than σ_v^2. These eigenvectors are referred to as the principal eigenvectors. It is well known that the columns of $U_1(\Theta)$ span the signal subspace defined as the span of the steering vectors $D(\Theta)$. The columns of $U_2(\Theta)$ span the complementary orthogonal noise subspace.

Since $R(\Theta)$ is not available, it is commonly replaced by its empirical covariance

$$\hat{R} \triangleq \frac{1}{N} \sum_{n=1}^{N} Y(t_n) Y^{\#}(t_n). \tag{3}$$

In this case, $\hat{U} = [\hat{u}_1, \cdots, \hat{u}_L]$ and $\hat{\Lambda} = \mathrm{diag}(\hat{\lambda}_1, \cdots, \hat{\lambda}_L)$ denote the eigenvectors and eigenvalues of $\hat{R}$, respectively, where the eigenvalues are assumed to be in a descending order. In addition, $\hat{U} = [\hat{U}_1, \hat{U}_2]$ where $\hat{U}_1 = [\hat{u}_1, \cdots, \hat{u}_M]$ denotes the set of principal empirical eigenvectors which correspond to the M largest empirical eigenvalues.

III. Min-Norm Music, Mode and ML

A. Min-Norm MUSIC

We first show that MUSIC is an approach in which each steering vector is linearly estimated in the least squares sense from principal empirical eigenvectors, i.e., the MUSIC estimate of θ_m is obtained from

$$\min_{\theta_m} \min_{c_m} \left\| d(\theta_m) - \sum_{i=1}^{M} c_m(i) \hat{u}_i \right\|^2 \tag{4}$$

where $c_m = (c_m(1), \cdots, c_m(M))^T$ denotes a vector of coefficients and $\|\cdot\|$ denotes the Euclidean norm. This interpretation was originally given by Schmidt in [1, p. 62] but it has not received much attention in the literature.

The minimizing coefficients in (4) are given by $\hat{c}_m(i) = \hat{u}_i^{\#} d(\theta_m)$. Hence,

$$\begin{aligned} \min_{\theta_m} \min_{c_m} & \left\| d(\theta_m) - \sum_{i=1}^{M} c_m(i) \hat{u}_i \right\|^2 \\ &= \min_{\theta_m} \left\| d(\theta_m) - \sum_{i=1}^{M} \hat{u}_i \hat{u}_i^{\#} d(\theta_m) \right\|^2 \\ &= \min_{\theta_m} \| d(\theta_m) - \hat{U}_1 \hat{U}_1^{\#} d(\theta_m) \|^2 \\ &= \min_{\theta_m} \| (I - \hat{U}_1 \hat{U}_1^{\#}) d(\theta_m) \|^2 \\ &= \min_{\theta_m} \| \hat{U}_2 \hat{U}_2^{\#} d(\theta_m) \|^2 \\ &= \min_{\theta_m} d^{\#}(\theta_m) \hat{U}_2 \hat{U}_2^{\#} d(\theta_m) \end{aligned} \tag{5}$$

which is identical to the MUSIC approach in [1, p. 96].

The above formulation provides a min-norm interpretation of MUSIC in estimating a single bearing parameter θ_m. MUSIC can also be interpreted as a min-norm approach for simultaneous estimation of all bearing parameters Θ. Consider estimation of Θ from $\hat{U}_1$ through linear estimation of the array manifold $D(\Theta)$ using the weak or Frobenius norm [15]. Specifically, let $C_M \triangleq [c_1, \cdots, c_M]$ be an $M \times M$ matrix of coefficients and estimate Θ from

$$\min_{\Theta} \min_{C_M} |D(\Theta) - \hat{U}_1 C_M|^2 \tag{6}$$

where the weak norm is defined by

$$|B|^2 \triangleq \frac{1}{L} \mathrm{tr}\{B^{\#} B\} = \frac{1}{L} \sum_{m=1}^{M} \|b_m\|^2 \tag{7}$$

and b_m is the mth column of B. Substituting $B = D(\Theta) - \hat{U}_1 C_M$, (6) becomes

$$\begin{aligned} \min_{\Theta} \min_{C_M} & \frac{1}{L} \sum_{m=1}^{M} \| d(\theta_m) - \hat{U}_1 c_m \|^2 \\ &= \frac{1}{L} \sum_{m=1}^{M} \min_{\theta_m} \min_{c_m} \| d(\theta_m) - \hat{U}_1 c_m \|^2 \end{aligned} \tag{8}$$

where each summand on the right hand side of (8) is identical to (4).

B. Min-Norm MODE

We now show that the MODE estimate of Θ can be obtained from linear estimation of the array manifold $D(\Theta)$ from the principal empirical eigenvectors $\hat{U}_1$ by minimizing a weighted Frobenius norm, i.e.,

$$\min_{\Theta}\min_{C_M}|D(\Theta)-\hat{U}_1C_M|_P^2 \tag{9}$$

where

$$|B|_P^2 \triangleq \frac{1}{L}\operatorname{tr}\{BPB^{\#}\}=\frac{1}{L}\sum_{i,j=1}^{M}\bar{P}_{i,j}b_i^{\#}b_j \tag{10}$$

and $\bar{P}_{i,j}$ denotes the complex conjugate of $P_{i,j}$. Since P is Hermitian positive definite, $|B|_P$ is a norm.

The minimization of (9) over C_M gives $C_M=\hat{U}_1^{\#}D(\Theta)$. Hence,

$$\begin{aligned}
&\min_{\Theta}\min_{C_M}|D(\Theta)-\hat{U}_1C_M|_P^2\\
&\quad=\min_{\Theta}|D(\Theta)-\hat{U}_1\hat{U}_1^{\#}D(\Theta)|_P^2\\
&\quad=\min_{\Theta}|\hat{U}_2\hat{U}_2^{\#}D(\Theta)|_P^2\\
&\quad=\min_{\Theta}\frac{1}{L}\operatorname{tr}\{D^{\#}(\Theta)\hat{U}_2\hat{U}_2^{\#}D(\Theta)P\},
\end{aligned} \tag{11}$$

which is identical to the MODE approach in [4, eq. (3.6)]. Continuing the development of (11) using (2) and $\hat{U}_2\hat{U}_2^{\#}=I-\hat{G}$, where $\hat{G}=\hat{U}_1\hat{U}_1^{\#}$ is the projection matrix onto the column space of $\hat{U}_1$ [15], gives

$$\begin{aligned}
&\min_{\Theta}\min_{C_M}|D(\Theta)-\hat{U}_1C_M|_P^2\\
&\quad=\min_{\Theta}\frac{1}{L}\operatorname{tr}\{\hat{U}_2\hat{U}_2^{\#}D(\Theta)PD^{\#}(\Theta)\}\\
&\quad=\min_{\Theta}\frac{1}{L}\operatorname{tr}\{(I-\hat{G})(R(\Theta)-\sigma_v^2I)\}\\
&\quad=\min_{\Theta}\frac{1}{L}\operatorname{tr}\{(I-\hat{G})R(\Theta)\}-\frac{L-M}{L}\sigma_v^2.
\end{aligned} \tag{12}$$

Hence, the MODE estimate of Θ can be obtained from minimization of $\operatorname{tr}\{(I-\hat{G})R(\Theta)\}$.

Note that for uncorrelated sources P is a diagonal matrix and minimization of either $|B|^2$ or $|B|_P^2$ results in the same estimate of Θ. In this case the bearing of each source can be independently estimated using the standard MUSIC approach. When the sources are correlated, however, all bearings must be simultaneously estimated and the two norms provide different estimates of Θ.

Note also that the proposed signal subspace fitting approach can lead to new bearing estimation approaches if other than the Frobenius norm are considered in (9). A possible norm is the strong norm $\|B\|^2=\lambda_{\max}\{B^{\#}B\}$ where $\lambda_{\max}\{\cdot\}$ denotes the largest eigenvalue.

C. Min-Norm ML

The deterministic ML estimation approach is now interpreted as a min-norm approach for estimating Θ. It is shown that the ML estimate of Θ is obtained from

$$\min_{\Theta}\min_{C_L}|\hat{U}-U_1(\Theta)C_L|_{\hat{\Lambda}}^2 \tag{13}$$

where C_L is an $M\times L$ matrix of estimation coefficients.

To prove (13) note that

$$|\hat{U}-U_1(\Theta)C_L|_{\hat{\Lambda}}^2=\frac{1}{L}\sum_{l=1}^{L}\hat{\lambda}_l\|\hat{u}_l-U_1(\Theta)c_l\|^2. \tag{14}$$

Minimization of (14) over C_L gives $\hat{c}_l=U_1^{\#}(\Theta)\hat{u}_l$, $l=1,\cdots,L$. Hence,

$$\begin{aligned}
&\min_{\Theta}\min_{C_L}\sum_{l=1}^{L}\hat{\lambda}_l\|\hat{u}_l-U_1(\Theta)c_l\|^2\\
&\quad=\min_{\Theta}\sum_{l=1}^{L}\hat{\lambda}_l\|\hat{u}_l-U_1(\Theta)U_1^{\#}(\Theta)\hat{u}_l\|^2\\
&\quad=\min_{\Theta}\sum_{l=1}^{L}\hat{\lambda}_l\hat{u}_l^{\#}(I-U_1(\Theta)U_1^{\#}(\Theta))\hat{u}_l\\
&\quad\overset{\text{(i)}}{=}\min_{\Theta}\sum_{l=1}^{L}\hat{\lambda}_l\hat{u}_l^{\#}(I-G(\Theta))\hat{u}_l\\
&\quad=\min_{\Theta}\operatorname{tr}\left\{(I-G(\Theta))\sum_{l=1}^{L}\hat{\lambda}_l\hat{u}_l\hat{u}_l^{\#}\right\}\\
&\quad=\min_{\Theta}\operatorname{tr}\{(I-G(\Theta))\hat{R}\}\\
&\quad=\operatorname{tr}\{\hat{R}\}-\max_{\Theta}\operatorname{tr}\{G(\Theta)\hat{R}\}
\end{aligned} \tag{15}$$

which is equivalent to the ML approach of [3]. Equality (i) results from the relation $G(\Theta)=U_1(\Theta)U_1^{\#}(\Theta)$ for the projection matrix $G(\Theta)$ onto the column space of $U_1(\Theta)$ [15].

Note that in the min-norm interpretation of ML the empirical eigenvectors $\{\hat{u}_l\}$ are linearly fitted by principal eigenvectors of the theoretical covariance $R(\Theta)$. In the min-norm interpretation of MODE the steering vectors were linearly estimated from the principal empirical eigenvectors. Since each eigenvector of $R(\Theta)$ depends on all bearing parameters, all $\{\theta_m\}$ must be simultaneously estimated in the ML approach. Another difference between the two approaches is that MODE estimation involves M empirical eigenvectors while ML estimation involves all L empirical eigenvectors. This observation was first made by Ziskind and Wax [3] when they compared ML with MUSIC.

D. Consistency of MUSIC, MODE, and ML

We have seen in (12) and (15) that the MODE estimate and the deterministic ML estimate of the bearings Θ of possibly correlated signals can respectively be obtained from minimization of the empirical cost functions

$$\mathcal{L}_{\mathrm{MODE}}(\Theta)\triangleq\operatorname{tr}\{(I-\hat{G})R(\Theta)\} \tag{16}$$

and

$$\mathcal{L}_{\mathrm{ML}}(\Theta)\triangleq\operatorname{tr}\{(I-G(\Theta))\hat{R}\}. \tag{17}$$

We now present a unified general proof for strong consistency of the two algorithms when the noisy signals are only assumed stationary and ergodic. The proof applies to MUSIC as a particular case of MODE when the sources are uncorrelated.

As mentioned earlier, consistency of the two algorithms could not be inferred from the analysis in [5]–[8]. We further note that the proof of consistency given in [8], and the independent proof for consistency of the ML approach given in [5] without the Gaussian assumption, are incomplete as we elaborate below.

We prove consistency in three steps. First we show that as $N \to \infty$, both $\mathcal{L}_{\mathrm{MODE}}(\Theta)$ and $\mathcal{L}_{\mathrm{ML}}(\Theta)$ converge uniformly w.p. 1. to the following limit functions:

$$\mathcal{L}_{\mathrm{MODE}}(\Theta) \xrightarrow{\mathrm{U}} \mathrm{tr}\{(I - G(\Theta^*))R(\Theta)\} \tag{18}$$

$$\mathcal{L}_{\mathrm{ML}}(\Theta) \xrightarrow{\mathrm{U}} \mathrm{tr}\{(I - G(\Theta))R(\Theta^*)\} \tag{19}$$

where Θ^* denotes the true value of Θ. Second, we show that both limit functions in (18) and (19) have unique minimizers at $\Theta = \Theta^*$. Third, we show that convergence of the empirical cost functions in (18) and (19) imply convergence of the MODE estimator $\hat{\Theta}_{\mathrm{MODE}}$ and the ML estimator $\hat{\Theta}_{\mathrm{ML}}$ to Θ^*.

The third step in our proof is crucial, since uniform convergence of the empirical cost function of an estimator does not necessarily imply convergence of the estimator sequence itself to the minimizer of the limit function nor it implies convergence of that sequence at all. Unfortunately, this step is missing from the proofs presented in [5] and [8]. The proof given in [5] for the consistency of the ML estimator is only concerned with the second step outlined above. The proof given in [8] for the consistency of a family of signal subspace fitting approaches (which includes an asymptotic version of the ML estimator) is only concerned with the first two steps outlined above. The proof given in [13] for consistency of the ML approach contains all three steps. The second step, however, is somewhat implicit in that proof, and the third step relies heavily on the setup of the array processing problem. The second step is important, since the existence of a unique minimizer of the limit function is a necessary and sufficient condition for convergence of the estimator sequence.

Consistency of MODE and ML is proven here under the following assumptions: A1) The source signals and noise are stationary and ergodic; A2) The principal eigenvalues of $R(\Theta)$ are distinct; A3) The covariance matrix P of the sources is positive definite; A4) $M < L$; A5) Any set of $M+1$ distinct steering vectors are linearly independent; A6) The parameter space of Θ is compact; A7) The empirical cost functions $\mathcal{L}_{\mathrm{MODE}}(\Theta)$ and $\mathcal{L}_{\mathrm{ML}}(\Theta)$ are continuous in Θ.

Assumption A1 implies that $\hat{R} \to R(\Theta^*)$ w.p. 1. as $N \to \infty$. Assumptions A1–A2 imply that $\hat{U}_1 \to U_1(\Theta^*)$ w.p. 1. [8, lemma 3]. Assumption A3 guarantees that the rank of the covariance matrix of the received clean signals $D(\Theta)PD(\Theta)$ equals the number of sources M. This assumption together with A4) were used in the derivation of MUSIC. In addition, A3 and A5 will be important in determining uniqueness of the estimator. Assumption A6 is automatically satisfied since we are considering bearing estimation in which each component of $\theta_m \in [0, 2\pi]$. Assumption A7 is satisfied if the steering vectors $\{d(\theta_m)\}$ are continuous functions of $\{\theta_m\}$. This assumption is usually met in practice.

We provide here a proof for consistency of both MODE and ML in which each step is explicitly shown. In addition, the proof of the third crucial step is given for a general setup. This proof may be useful for other problems besides the array processing problem. We begin our proof by showing the validity of (18). Note that for any Θ,

$$\begin{aligned}
|\hat{G} - G(\Theta)| &\overset{\text{(i)}}{\leq} \|G(\Theta) - \hat{G}\| \\
&= \|U_1(\Theta)U_1^{\#}(\Theta) - \hat{U}_1\hat{U}_1^{\#}\| \\
&= \|U_1(\Theta)U_1^{\#}(\Theta) - U_1(\Theta)\hat{U}_1^{\#} \\
&\quad + U_1(\Theta)\hat{U}_1^{\#} - \hat{U}_1\hat{U}_1^{\#}\| \\
&= \|U_1(\Theta)(U_1(\Theta) - \hat{U}_1)^{\#} \\
&\quad + (U_1(\Theta) - \hat{U}_1)\hat{U}_1^{\#}\| \\
&\overset{\text{(ii)}}{\leq} \|U_1(\Theta)(U_1(\Theta) - \hat{U}_1)^{\#}\| \\
&\quad + \|(U_1(\Theta) - \hat{U}_1)\hat{U}_1^{\#}\| \\
&\overset{\text{(iii)}}{\leq} \|U_1(\Theta)\| \, \|(U_1(\Theta) - \hat{U}_1)^{\#}\| \\
&\quad + \|\hat{U}_1\| \, \|(U_1(\Theta) - \hat{U}_1)\| \\
&\overset{\text{(iv)}}{=} \|U_1(\Theta) - \hat{U}_1\| + \|U_1(\Theta) - \hat{U}_1\| \\
&= 2\|U_1(\Theta) - \hat{U}_1\|
\end{aligned} \tag{20}$$

where (i) results from the fact that the strong norm dominates the weak norm, i.e., for any two matrices A and B, $|A| \leq \|A\|$; (ii) results from the triangle inequality for the strong norm: $\|A + B\| \leq \|A\| + \|B\|$; Inequality (iii) results from [16, p. 364]: $\|AB\| \leq \|A\|\|B\|$; And (iv) results from $\|\hat{U}_1\| = \|U_1(\Theta)\| = 1$. Now,

$$\begin{aligned}
&\sup_{\Theta} |\mathcal{L}_{\mathrm{MODE}}(\Theta) - \mathrm{tr}\{(I - G(\Theta^*))R(\Theta)\}| \\
&\overset{\text{(i)}}{=} \sup_{\Theta} |\mathrm{tr}\{(\hat{G} - G(\Theta^*))R(\Theta)\}| \\
&\overset{\text{(ii)}}{\leq} L \sup_{\Theta}[|R(\Theta)|] \, |\hat{G} - G(\Theta^*)| \\
&\overset{\text{(iii)}}{\leq} 2L \sup_{\Theta}[|R(\Theta)|] \, \|\hat{U}_1 - U_1(\Theta^*)\|
\end{aligned} \tag{21}$$

where (i) results from (16); (ii) is a consequence of the inequality $\frac{1}{L}|\mathrm{tr}\{AB\}| \leq |A||B|$ which can be easily shown using the Schwarz inequality [17]; and (iii) results from (20). Since $R(\Theta)$ is uniformly bounded, and by Assumptions A1–A2 $\hat{U}_1 \to U_1(\Theta^*)$ w.p. 1. as $N \to \infty$, we have that

$$\begin{aligned}
&\sup_{\Theta} |\mathcal{L}_{\mathrm{MODE}}(\Theta) - \mathrm{tr}\{(I - G(\Theta^*))R(\Theta)\}| \\
&\qquad \to 0 \text{ w.p. 1. as } N \to \infty.
\end{aligned} \tag{22}$$

Thus

$$\mathcal{L}_{\mathrm{MODE}}(\Theta) \xrightarrow{\mathrm{U}} \mathrm{tr}\{(I - G(\Theta^*))R(\Theta)\} \text{ w.p. 1.} \tag{23}$$

The theoretical cost function $\mathrm{tr}\{(I - G(\Theta^*))R(\Theta)\}$ corresponding to the MODE algorithm is minimized by a *unique* $\Theta = \Theta^*$. To show this result notice that

$$\begin{aligned}
&\mathrm{tr}\{(I - G(\Theta^*))R(\Theta)\} \\
&\quad = \mathrm{tr}\{(I - G(\Theta^*))(R(\Theta) - \sigma_v^2 I)\} + \sigma_v^2 \mathrm{tr}\{(I - G(\Theta^*))\} \\
&\quad = \mathrm{tr}\{(I - G(\Theta^*))D(\Theta)PD^{\#}(\Theta)\} \\
&\qquad + \sigma_v^2(L - M) \geq \sigma_v^2(L - M)
\end{aligned} \tag{24}$$

since

$$\begin{aligned}\operatorname{tr}\{(I-G(\Theta^*))D(\Theta)PD^{\#}(\Theta)\} \\ = \operatorname{tr}\{(I-G(\Theta^*))D(\Theta)PD^{\#}(\Theta)(I-G(\Theta^*))\} \geq 0\end{aligned} \tag{25}$$

due to the positive definiteness of P assumed in A3. If $\Theta = \Theta^*$ then $(I-G(\Theta^*))D(\Theta) = 0$ since $G(\Theta^*)$ is the projector onto the column space of $D(\Theta^*)$. Hence, (25) holds with equality. If $\Theta \neq \Theta^*$ then $(I-G(\Theta^*))D(\Theta) \neq 0$ since if not, then $(I-G(\Theta^*))D(\Theta) = 0$, and as before, $(I-G(\Theta^*))D(\Theta^*) = 0$. This means that $G(\Theta^*)$ is the projector onto the column space of $D(\Theta^*)$ and $D(\Theta)$. By Assumption A5, the dimension of this column space is at least $M+1$. This, however, contradicts the fact that $G(\Theta^*)$ is a projector on an M-dimensional subspace. Since, in addition, P is positive definite by Assumption A3, we have from (25) that $\operatorname{tr}\{(I-G(\Theta^*))D(\Theta)PD^{\#}(\Theta)\} > 0$ when $\Theta \neq \Theta^*$. Thus, equality in (24) holds iff $\Theta = \Theta^*$ and the minimizer of $\operatorname{tr}\{(I-G(\Theta^*))R(\Theta)\}$ is unique.

Using similar arguments it can be shown that $\mathcal{L}_{\mathrm{ML}}(\Theta)$ converges uniformly to $\operatorname{tr}\{(I-G(\Theta))R(\Theta^*)\}$ w.p. 1., and that the minimizer of $\operatorname{tr}\{(I-G(\Theta))R(\Theta^*)\}$ is unique. Specifically,

$$\begin{aligned}&\sup_{\Theta} |\mathcal{L}_{\mathrm{ML}}(\Theta) - \operatorname{tr}\{(I-G(\Theta))R(\Theta^*)\}| \\ &\overset{\text{(i)}}{=} \sup_{\Theta} |\operatorname{tr}\{(I-G(\Theta))(\hat{R}-R(\Theta^*))\}| \\ &\overset{\text{(ii)}}{\leq} L \sup_{\Theta} [|I-G(\Theta)|]\, |\hat{R}-R(\Theta^*)|\end{aligned} \tag{26}$$

where (i) results from (17) and (ii) results from the Schwarz inequality. Since $|I-G(\Theta)|$ is uniformly bounded, and $\hat{R} \to R(\Theta^*)$ w.p. 1. as $N \to \infty$, we have shown uniform convergence of $\mathcal{L}_{\mathrm{ML}}(\Theta)$ to $\operatorname{tr}\{(I-G(\Theta))R(\Theta^*)\}$. Uniqueness of the minimizer of $\operatorname{tr}\{(I-G(\Theta))R(\Theta^*)\}$ can be shown using an identical argument to that given in (24) and (25).

In summary, we have shown that $\mathcal{L}_{\mathrm{MODE}}(\Theta)$ and $\mathcal{L}_{\mathrm{ML}}(\Theta)$ converge uniformly to $\operatorname{tr}\{(I-G(\Theta^*))R(\Theta)\}$ and $\operatorname{tr}\{(I-G(\Theta))R(\Theta^*)\}$, respectively, and both limit functions are minimized by the unique $\Theta = \Theta^*$. We now show that these facts together with Assumptions A6–A7 imply convergence of the estimator sequences $\hat{\Theta}_{\mathrm{ML}}$ and $\hat{\Theta}_{\mathrm{MODE}}$ to Θ^*.

Let $f_N(\Theta)$ denote either $\mathcal{L}_{\mathrm{ML}}(\Theta)$ or $\mathcal{L}_{\mathrm{MODE}}(\Theta)$. Let $f(\Theta)$ denote either $\operatorname{tr}\{(I-G(\Theta))R(\Theta^*)\}$ or $\operatorname{tr}\{(I-G(\Theta^*))R(\Theta)\}$. Let Θ_N^* denote the minimizer of $f_N(\Theta)$. Recall that Θ^* is the minimizer of $f(\Theta)$. We now show that

$$f_N(\Theta) \xrightarrow{\mathrm{U}} f(\Theta) \Rightarrow \Theta_N^* \to \Theta^*. \tag{27}$$

From the uniform convergence of $f_N(\Theta)$ to $f(\Theta)$ we have that

$$|f(\Theta_N^*) - f_N(\Theta_N^*)| \leq \epsilon_N \tag{28}$$

where $\epsilon_N \to 0$ as $N \to \infty$. Furthermore,

$$\begin{aligned}f(\Theta^*) - \epsilon_N \leq f(\Theta_N^*) - \epsilon_N \leq f_N(\Theta_N^*) \leq f_N(\Theta^*) \\ \leq f(\Theta^*) + \epsilon_N\end{aligned} \tag{29}$$

implies

$$|f_N(\Theta_N^*) - f(\Theta^*)| \leq \epsilon_N. \tag{30}$$

Hence, from (28) and (30)

$$|f(\Theta_N^*) - f(\Theta^*)| \leq 2\epsilon_N \tag{31}$$

and $f(\Theta_N^*) \to f(\Theta^*)$ as $N \to \infty$. Hence, it suffices to show that

$$f(\Theta_N^*) \to f(\Theta^*) \Rightarrow \Theta_N^* \to \Theta^*. \tag{32}$$

Let $\Theta_{N_i}^* \to \Theta'$, where $\Theta_{N_i}^*$ is a subsequence of Θ_N^*. The existence of such convergent subsequence results from the compactness of the parameter space. We have that

$$\begin{aligned}f(\Theta^*) &\overset{\text{(i)}}{=} \lim_{N\to\infty} f(\Theta_N^*) \overset{\text{(ii)}}{=} \lim_{i\to\infty} f(\Theta_{N_i}^*) \\ &\overset{\text{(iii)}}{=} f(\lim_{i\to\infty} \Theta_{N_i}^*) = f(\Theta')\end{aligned} \tag{33}$$

where i) results from the assumption in (32); ii) results from the fact that $f(\Theta_{N_i}^*)$ is a convergent sequence; and iii) results from continuity of $f(\cdot)$ which in turns results from our assumption that $f_N(\cdot)$ are continuous and convergence is uniform. Since the minimizer of $f(\cdot)$ is unique by assumption,

$$f(\Theta^*) = f(\Theta') \Rightarrow \Theta' = \Theta^*. \tag{34}$$

Thus, there exists a subsequence of Θ_N^* which converges to Θ^*.

Assume next that Θ_N^* does not converge to Θ^*. This means that there exist $\epsilon > 0$ and $\Theta_{N_i}^*$ such that $\|\Theta_{N_i}^* - \Theta^*\| > \epsilon$. Yet, the subsequence $f(\Theta_{N_i}^*)$ converges to $f(\Theta^*)$ since $f(\Theta_N^*)$ is a convergent sequence. From the previous discussion, $\Theta_{N_i}^*$ must have a subsequence $\Theta_{N_{ij}}^*$ such that $\Theta_{N_{ij}}^* \to \Theta^*$. This, however, can not happen since all elements of $\Theta_{N_{ij}}^*$ are at a minimum distance of ϵ from Θ^*. Hence, Θ_N^* must converge to Θ^*.

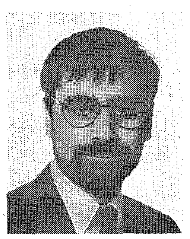
ACKNOWLEDGMENT

The authors are grateful to A. Orlitsky for his help in proving the consistency of MODE and ML.

REFERENCES

[1] R. O. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation," Ph.D. dissertation, Stanford Univ., Stanford, CA., 1981.
[2] M. Wax, "Detection and estimation of superimposed signals," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1985.
[3] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternate projection," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1553–1560, Oct. 1988.
[4] P. Stoica and K. Sharman, "Maximum likelihood methods for direction of arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1132–1143, July 1990.
[5] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer–Rao bound," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 720–741, May 1989.
[6] ——, "MUSIC, Maximum likelihood, and Cramer–Rao bound: Further results and comparisons," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 2140–2150, Dec. 1990.

[7] ——, "Performance study of conditional and unconditional direction of arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1783–1795, Oct. 1990.
[8] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, pp. 1110–1120, May 1991.
[9] B. Ottersten, M. Viberg, and T. Kailath, "Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data," *IEEE Trans. Signal Processing*, vol. 40, pp. 590–599, Mar. 1992.
[10] M. Kaveh and A. J. Barabell, "The statistical performance of the MUSIC and the minimum-norm algorithm in resolving plane waves in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 331–341, Apr. 1986.
[11] B. Porat and B. Friedlander, "Analysis of the asymptotic relative efficiency of the MUSIC algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 532–543, Apr. 1988.
[12] B. Friedlander, "A sensitivity analysis of the MUSIC algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1740–1751, Oct. 1990.
[13] Q. Wu and D. R. Fuhrmann, "A parametric method for determining the number of signals in narrow-band direction finding," *IEEE Trans. Signal Processing*, vol. 39, pp. 1848–1857, Aug. 1991.
[14] A. J. Weiss and B. Friedlander, "On the Cramer–Rao bound for direction finding of correlated signals," *IEEE Trans. Signal Processing*, vol. 41, pp. 495–499, Jan. 1993.
[15] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. New York: Addison-Wesley, 1990.
[16] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, 2nd ed. New York: Academic, 1985.
[17] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 251–266, July 1995.



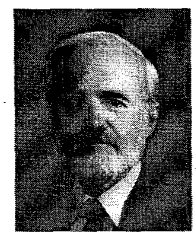
Yariv Ephraim (S'82–M'84–SM'90–F'94) received the D.Sc. degree in electrical engineering in 1984 from the Technion—Israel Institute of Technology, Haifa, Israel.

He was a Rothschild Post-Doctoral Fellow at the Information Systems Laboratory of Stanford University, Stanford, CA, during 1984–1985. He was a Member of the Technical Staff at the Information Principles Research Laboratory of AT&T Bell Laboratories, Murray Hill, NJ, during 1985–1993. In 1991, he joined George Mason University, Fairfax, VA, where he is currently an Associate Professor of Electrical and Computer Engineering.

Neri Merhav (M'87–SM'92) was born in Haifa, Israel, on March 16, 1957. He received the B.Sc., M.Sc., and D.Sc. degrees from the Technion—Israel Institute of Technology, Haifa, in 1982, 1985, and 1988, respectively, all in electrical engineering.

From 1982 to 1985, he was a Research Associate with the Israel IBM Scientific Center, Haifa, where he developed algorithms for speech coding, speech synthesis, and adaptive filtering of speech signals in array sensors. From 1988 to 1990, he was with the Speech Research Department of AT&T Bell Laboratories, Murray Hill, NJ, where he investigated and developed algorithms for speech recognition. He is currently with the Electrical Engineering Department of the Technion. His research interests are statistical signal processing, information theory, and statistical communications.

Dr. Merhav was a co-recipient of the 1993 Paper Award of the IEEE Information Theory Society.



Harry L. Van Trees (M'57–SM'73–F'74–LF'94) received the B.Sc. degree from the U.S. Military Academy at West Point in 1952 and the Sc.D. degree from the Massachusetts Institute of Technology (MIT), Cambridge, USA, in 1961.

From 1961 to 1975, he was with the Electrical Engineering Department at MIT, becoming a Full Professor in 1969. During this period, he was active in graduate course development and was the leader of a research group working in detection and estimation theory and radar/sonar theory. Since 1988, he has been with George Mason University, Fairfax, VA, where he is currently a Distinguished Professor of Information Technology, Electrical and Systems Engineering. In June of 1989, he founded the Center of Excellence in Command, Control, Communications, and Intelligence, which now has 18 faculty members associated with it. The Center has a research program that includes work in sensing and data fusion, command support systems, communications, modeling and simulation, C3 architectures, and information systems. He has served as Chief Scientist of the Defense Communication Agency and Chief Scientist of the U.S. Air Force. He was Principal Deputy Assistant Secretary of Defense (C^3I). He served as an Executive Vice President and President of M/A-COM Government Systems Division. He is the author of a three-volume set of books on detection, estimation, and modulation theory. These books contain a unified approach to communications, radar, sonar, and seismic applications. The first volume is a classic in its field and is used in graduate schools throughout the world; it is currently in its 27th printing.

Dr. Van Trees has received the Presidential Award for Meritorious Executive, the Distinguished Civilian Service Awards, and the AFCEA Gold Medal for Engineering.