Computationally Efficient Maximum Likelihood Estimation of Structured Covariance Matrices

Hongbin Li, *Student Member, IEEE*, Petre Stoica, *Fellow, IEEE*, and Jian Li, *Senior Member, IEEE*

***Abstract*— By invoking the extended invariance principle (EXIP), we present herein a computationally efficient method that provides asymptotic (for large samples) maximum likelihood (AML) estimation for structured covariance matrices and will be referred to as the AML algorithm. A closed-form formula for estimating Hermitian Toeplitz covariance matrices that makes AML computationally simpler than most existing Hermitian Toeplitz matrix estimation algorithms is derived. Although the AML covariance matrix estimator can be used in a variety of applications, we focus on array processing in this paper. Our simulation study shows that AML enhances the performances of angle estimation algorithms, such as MUSIC, by making them very close to the corresponding Cramér–Rao bound (CRB) for uncorrelated signals. Numerical comparisons with several structured and unstructured covariance matrix estimators are also presented.**

I. Introduction

THE COVARIANCE matrix of a stationary signal is Hermitian and Toeplitz. However, the conventional sample covariance matrix obtained from a finite number of observations seldom has this structure. Estimating structured covariance matrices is of particular interest in a variety of applications, including array processing and time series analysis. For example, many well-known algorithms in array processing are based on an estimate of the covariance matrix, and using these algorithms with structured covariance matrix estimates in lieu of the sample covariance matrix may yield better angle estimation performance.

The literature on structured covariance matrix estimation includes [1]–[9] (also see the references therein). An important technique is the maximum likelihood (ML) approach considered in [1]–[5], among which [2] appears to be the first to study the ML method in its full generality. Since the *exact* ML estimation of a Hermitian Toeplitz covariance matrix has no closed-form solution [10], the ML methods proposed in the previous studies are iterative and computationally involved, yet they are not guaranteed to yield the global optimal solution. To avoid this difficulty, other suboptimal approaches have been considered, a notable example being the iterated Toeplitz approximation method (ITAM). ITAM alternatively makes use of rank approximation and Toeplitzation techniques to ensure that the covariance matrix estimate has the desired structure [6]–[8]. However, in spite of the fact that the ITAM estimator produces a covariance matrix estimate that is, in general, closer in the Frobenius norm sense to the true covariance matrix than the sample covariance matrix, it is not guaranteed that better application-related performances, such as angle estimation accuracy in array processing, ensue. In fact, we have found that using the ITAM covariance matrix estimate with MUSIC [11], [12] (referred to as ITAM-MUSIC) provides inconsistent (in signal-to-noise ratio or SNR) angle estimates (see Section V for details). In addition, ITAM is an iterative algorithm that requires a sequence of eigendecompositions of intermediate covariance matrix estimates, and therefore, it is still computationally inefficient. A new covariance estimator was recently introduced in [9] by Fourier inverting the Capon power spectral density estimates [13]. Although the Capon covariance estimator was derived primarily for covariance sequence estimation, structured covariance matrices can be readily constructed from covariance sequence estimates. Although the Capon covariance sequence estimator usually gives lower mean squared errors (MSE's) than the sample covariance sequence estimator, the former is, in general, not an optimal method, nor is the structured covariance matrix estimate obtained therefrom.

In this paper, we present a computationally efficient method for structured covariance matrix estimation. The method provides an asymptotic (for large samples) maximum likelihood (AML) estimate of a structured covariance matrix and will be referred to as the AML algorithm. AML makes use of the extended invariance principle (EXIP) for parameter estimation, which was introduced in [14]. The key idea of EXIP is first reparameterizing the ML criterion to allow a simple solution and then refining that solution by using a weighted least squares (WLS) technique. It turns out that, by invoking EXIP, the simple solution obtained in the first step of AML is the unstructured sample covariance matrix, and that the optimal weighting matrix is the Fisher information matrix (FIM) corresponding to the unstructured ML criterion. As we will show, the AML approach yields a closed-form solution to the Hermitian Toeplitz covariance matrix estimation problem.

The quality of a covariance matrix estimate should be assessed by studying how it behaves in specific applications. We consider herein the impact of using structured covariance matrix estimates on angle estimation in array processing. In particular, we obtain the angle estimates by using MUSIC

Manuscript recevied December 5, 1997; revised May 28, 1998. This work was supoorted in part by the National Science Foundation under Grant MIP-9457388, the Office of Naval Research under Grant N00014-96-0817, and the Senior Individual Grant Program of the Swedish Foundation for Strategic Research. The associate editor coordinating the review of this paper and approving it for publication was Dr. Sergios Theodoridis.

H. Li and J. Li are with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: li@dsp.ufl.edu).

P. Stoica is with the Systems and Control Group, Uppsala University, Uppsala, Sweden.

Publisher Item Identifier S 1053-587X(99)03232-8.

with the AML covariance matrix estimate, and this approach is referred to as AML-MUSIC. By exploiting the Toeplitz structure of the covariance matrix in angle estimation, we assume implicitly the *a priori* knowledge that the incident signals are uncorrelated. With this additional knowledge, the corresponding Cramér–Rao bound (CRB), which is referred to as the structured CRB or S-CRB, should be lower than the CRB without this knowledge, which is referred to as the unstructured CRB or U-CRB. Simulation results show that the performance of AML-MUSIC is very close to the S-CRB as the number of observations increases, whereas, as is well known, using MUSIC with the unstructured sample covariance matrix (which is referred to as the standard MUSIC) can only approach the U-CRB [15].

A weighted subspace fitting (WSF) algorithm was proposed in [16] for estimating the arrival angles of uncorrelated signals. WSF obtains better angle estimates than the standard MUSIC by exploiting the *a priori* knowledge of signal correlation as well, but in a different way. It has been shown in [16] that WSF asymptotically (for large samples) achieves the S-CRB. Since the array covariance matrix can be parameterized by the signal and noise parameters (see, e.g., [17]), the WSF algorithm can be viewed as a structured covariance matrix estimator by using the WSF estimates of the signal and noise parameter estimates, although this was not observed in [16]. Numerical studies indicate that in terms of angle estimation performance, AML-MUSIC and WSF yield similar angle estimates for most cases, except that in some difficult scenarios, such as at relatively low SNR's or when the signal arrival angles are closely spaced, the former tends to outperform the latter slightly. Moreover, in Section V, we will show that AML-MUSIC is computationally more efficient than WSF.

An outline of this paper is as follows. Section II presents the problem formulation and a brief description of the exact ML approach. The derivation of the AML algorithm for Hermitian Toeplitz covariance matrix estimation is given in Section III. Section IV addresses the implementation issues of AML. Section V contains a numerical study of the AML estimator as well as comparisons with ITAM, WSF, and the unstructured sample covariance matrix estimator when applied to angle estimation. Finally, Section VI concludes this study.

II. Problem Formulation

Assume that $\mathbf{y}(n) \in \mathbb{C}^{M\times 1}$, $n = 1, 2, \ldots, N$, denote N independent samples of a circularly symmetric complex Gaussian stationary random process with zero-mean and Hermitian Toeplitz covariance matrix $\mathbf{R}(\boldsymbol{\phi})$ that is a known function of an unknown parameter vector $\boldsymbol{\phi} \in \mathbb{R}^{L\times 1}$. The problem of interest herein is to determine a Hermitian Toeplitz matrix estimate $\mathbf{R}(\hat{\boldsymbol{\phi}})$ of $\mathbf{R}(\boldsymbol{\phi})$ from $\{\mathbf{y}(n)\}$.

The previous situation occurs in many applications including array processing, in which $\{\mathbf{y}(n)\}_{n=1}^{N}$ denotes the array output vectors when i) the incoming signals are uncorrelated, and ii) a uniform linear array (ULA) is employed (see, e.g., [11], [15], and [17]). Usually, a Hermitian Toeplitz matrix is parameterized such that $\boldsymbol{\phi}$ consists of the real and imaginary parts of the first column (or row) of $\mathbf{R}$. However, for the covariance matrix $\mathbf{R}(\boldsymbol{\phi})$ in the array processing application, another natural way of parameterizing $\mathbf{R}$ is by means of the signal and noise parameters. Let K uncorrelated signals impinge on a ULA of M sensors, and assume that the additive noise is spatially white and independent of the signals. Then, the spatial covariance matrix has the form [11], [15], [18], [19]

$$\mathbf{R}(\boldsymbol{\phi}) = \mathbf{A}(\boldsymbol{\theta})\mathbf{S}\mathbf{A}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I}_M \tag{1}$$

where $\boldsymbol{\theta} = [\theta_1 \;\; \theta_2 \;\; \ldots \;\; \theta_K]^T \in \mathbb{R}^{K\times 1}$ denotes the vector consisting of the arrival angles relative to the array normal direction, $\mathbf{S} \in \mathbb{R}^{K\times K}$ denotes the diagonal signal covariance matrix, $(\cdot)^H$ denotes the conjugate transpose, σ^2 denotes the noise variance, $\mathbf{I}_M$ denotes the $M \times M$ identity matrix, and $\mathbf{A}(\boldsymbol{\theta}) \in \mathbb{C}^{M\times K}$ denotes the array manifold matrix and has the form of (2), shown at the bottom of the page, with λ_0, d and θ_k denoting the signal wavelength, the spacing between two adjacent sensors, and the arrival angle of the kth signal, respectively. The parameter vector is then given by

$$\boldsymbol{\phi} = [\boldsymbol{\theta}^T \quad \mathbf{s}^T \quad \sigma^2]^T \tag{3}$$

where $(\cdot)^T$ denotes the transpose, and $\mathbf{s} \in \mathbb{R}^{K\times 1}$ is the vector consisting of the diagonal elements of $\mathbf{S}$. The WSF approach in [16] can be viewed as an estimator of the spatial covariance matrix by choosing $\boldsymbol{\phi}$ as in (3). In our approach, however, the general covariance matrix estimation problem is considered, and hence, $\boldsymbol{\phi} \in \mathbb{R}^{(2M-1)\times 1}$ is the vector consisting of the real and imaginary parts of the first column or row of $\mathbf{R}$.

The exact ML estimate $\hat{\boldsymbol{\phi}}$ of $\boldsymbol{\phi}$ is obtained by maximizing the likelihood function, which is equivalent to

$$\hat{\boldsymbol{\phi}} = \arg\min_{\boldsymbol{\phi}\in D_\phi} L_\phi(\boldsymbol{\phi}) \tag{4}$$

where $D_\phi = \mathbb{R}^{(2M-1)\times 1}$, and

$$L_\phi(\boldsymbol{\phi}) = \ln|\mathbf{R}(\boldsymbol{\phi})| + \mathrm{tr}\left[\mathbf{R}^{-1}(\boldsymbol{\phi})\tilde{\mathbf{R}}\right] \tag{5}$$

with

- $|\cdot|$ determinant;
- $\mathrm{tr}(\cdot)$ trace;
- $\tilde{\mathbf{R}}$ sample covariace matrix:

$$\tilde{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{y}(n)\mathbf{y}^H(n). \tag{6}$$

$$\mathbf{A}(\boldsymbol{\theta}) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{-j2\pi d\sin\theta_1/\lambda_0} & e^{-j2\pi d\sin\theta_2/\lambda_0} & \cdots & e^{-j2\pi d\sin\theta_K/\lambda_0} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-j2\pi(M-1)d\sin\theta_1/\lambda_0} & e^{-j2\pi(M-1)d\sin\theta_2/\lambda_0} & \cdots & e^{-j2\pi(M-1)d\sin\theta_K/\lambda_0} \end{bmatrix} \tag{2}$$

If we impose no structure on $\mathbf{R}$ except for the Hermitian symmetry (i.e., $\mathbf{R}^H = \mathbf{R}$), then it is known that the ML estimate of $\mathbf{R}$ is given by $\tilde{\mathbf{R}}$ (see, e.g., [2], [20]), whereas if we observe the structure of $\mathbf{R}$ implied by the parameterization of $\mathbf{R}(\phi)$, the ML estimate of $\mathbf{R}$ is given by $\mathbf{R}(\hat{\phi})$. Asymptotically (in N), $\mathbf{R}(\hat{\phi})$ will have better accuracy than $\tilde{\mathbf{R}}$. Then, for example, in array processing applications, using $\mathbf{R}(\hat{\phi})$ with angle estimators such as MUSIC [11], [12] and ESPRIT [18] may yield better angle estimation accuracy than using $\tilde{\mathbf{R}}$. However, solving for the ML solution from (4) turns out to be very complicated because of the nonlinearity of the cost function. This limits the interest in using the exact ML structured covariance matrix estimate in practical applications.

III. Derivation of the AML Estimator

By making use of EXIP, an asymptotic ML structured covariance matrix estimate can be obtained as follows. Let $\mathbf{r} = \text{vec}(\mathbf{R}) \in \mathbb{C}^{M^2\times 1}$, where $\text{vec}(\cdot)$ denotes the operation of stacking the columns of a matrix on top of one another, and let $\gamma \in D_\gamma \subset \mathbb{R}^{M^2\times 1}$ denote the vector made from the real and imaginary parts of the elements of $\mathbf{R}$ both above and on the main diagonal. Evidently, there is an $M^2 \times M^2$ matrix $\mathbf{F}$ such that

$$\gamma = \mathbf{F}\mathbf{r}. \tag{7}$$

Furthermore, since the mapping from $\mathbf{r}$ to γ is one-to-one, $\mathbf{F}$ must be nonsingular. Note that γ represents a reparameterization of the covariance matrix $\mathbf{R}$ originally parameterized by ϕ. It follows that there exists a mapping from ϕ to γ, i.e.,

$$\gamma = f(\phi) \in D_\gamma, \quad \forall \phi \in D_\phi \tag{8}$$

such that

$$L_\gamma(f(\phi)) = L_\phi(\phi), \quad \forall \phi \in D_\phi. \tag{9}$$

We can see that $f(\cdot)$ is also a one-to-one mapping from D_ϕ to D_γ, which is similar to the mapping in (7). Consequently, (4) is equivalent to

$$\hat{\gamma} = \arg\min_{\gamma \in D_\gamma} L_\gamma(\gamma) \tag{10}$$

which in turn implies that

$$\hat{\gamma} = f(\hat{\phi}). \tag{11}$$

Equation (11) is the well-known *invariance principle* (IP) [21]. However, solving (10) instead of (4) yields no computational advantage at all. To achieve computational simplification, we make use of EXIP and enlarge the constraint set associated with (10). Specifically, we consider the unstructured optimization problem

$$\tilde{\gamma} = \arg\min_{\gamma \in \tilde{D}_\gamma} L_\gamma(\gamma), \quad D_\gamma \subset \tilde{D}_\gamma \tag{12}$$

where $\tilde{D}_\gamma = \mathbb{R}^{M^2\times 1}$, which means that the Toeplitz constraint on $\mathbf{R}$ is relaxed. We can see that the set D_γ is a zero-measure subset of $\tilde{D}_\gamma$, which implies that $\tilde{\gamma} \notin D_\gamma$ with probability one. The solution of (12) is, as we have mentioned in Section II, $\tilde{\gamma} = \mathbf{F}\text{vec}(\tilde{\mathbf{R}})$. Since $\tilde{\mathbf{R}}$ is a consistent estimate of $\mathbf{R}$, we have

$$\lim_{N\to\infty} \tilde{\gamma} = \lim_{N\to\infty} \hat{\gamma} = \gamma. \tag{13}$$

This consistency property enables us to obtain an asymptotically (in N) valid approximation of $\hat{\phi}$ from $\tilde{\gamma}$, as follows.

Theorem 1: Let $\tilde{\gamma}$ be given by (12) or, equivalently, $\tilde{\gamma} = \mathbf{F}\text{vec}(\tilde{\mathbf{R}})$. Then

$$\tilde{\phi} = \arg\min_{\phi \in D_\phi} [\tilde{\gamma} - f(\phi)]^T \mathbf{\Gamma}^{-1} [\tilde{\gamma} - f(\phi)] \tag{14}$$

is an asymptotically (in N) valid approximation of the ML estimate $\hat{\phi}$, where $\mathbf{\Gamma}$ is a consistent (in N) estimate of the inverse of the Fisher information matrix (FIM) corresponding to the unstructured ML criterion given in (12) or, equivalently, of the covariance matrix of $\tilde{\gamma}$.

Proof: See [14] and [20]. □

Let $\tilde{\mathbf{r}} = \text{vec}(\tilde{\mathbf{R}})$. Using (7) and the facts that $\tilde{\gamma} = \mathbf{F}\tilde{\mathbf{r}}$, and $\mathbf{\Gamma} = \mathbf{F}\mathbf{C}\mathbf{F}^H$, where $\mathbf{C} = \text{cov}(\tilde{\mathbf{r}})$, we can readily check that (14) is equivalent to

$$\tilde{\phi} = \arg\min_{\phi \in D_\phi} [\tilde{\mathbf{r}} - \mathbf{r}(\phi)]^H \mathbf{C}^{-1} [\tilde{\mathbf{r}} - \mathbf{r}(\phi)]. \tag{15}$$

It turns out to be more convenient to work with (15) than with (14) since we thus avoid the transformation from $\tilde{\mathbf{r}}$ to $\tilde{\gamma}$.

An expression for $\mathbf{C}$, which is needed in (15), is obtained as follows. Let $\mathbf{r}_m$ denote the mth column of $\mathbf{R}$, and let $\tilde{\mathbf{r}}_m$ denote the m-th column of $\tilde{\mathbf{R}}$. We have

$$E\{\tilde{\mathbf{r}}_m\} = \mathbf{r}_m \tag{16}$$

and

$$\begin{aligned} E\{\tilde{\mathbf{r}}_{m_1}\tilde{\mathbf{r}}_{m_2}^H\} &= \frac{1}{N^2}\sum_{n_1=1}^{N}\sum_{n_2=1}^{N} E\big\{\mathbf{y}(n_1)y_{m_1}^*(n_1)y_{m_2}(n_2)\mathbf{y}^H(n_2)\big\} \\ &= \frac{1}{N^2}\sum_{n_1=1}^{N}\sum_{n_2=1, n_2\neq n_1}^{N} E\big\{\mathbf{y}(n_1)y_{m_1}^*(n_1)\big\} \\ &\quad \times E\big\{y_{m_2}(n_2)\mathbf{y}^H(n_2)\big\} \\ &\quad + \frac{1}{N^2}\sum_{n_1=1}^{N} E\big\{\mathbf{y}(n_1)y_{m_1}^*(n_1)y_{m_2}(n_1)\mathbf{y}^H(n_1)\big\} \end{aligned} \tag{17}$$

where $(\cdot)^*$ denotes the complex conjugate, $y_m(n)$ denotes the mth element of $\mathbf{y}(n)$, and we have used the assumption that $\mathbf{y}(n_1)$ and $\mathbf{y}(n_2)$ are independent of one another for $n_1 \neq n_2$ (see Section II). The first term of the last equality of (17) is easily seen to be equal to $(1 - \frac{1}{N})\mathbf{r}_{m_1}\mathbf{r}_{m_2}^H$. By using the standard result on calculating the fourth-order moment of Gaussian random variables (see, e.g., [17]), along with the zero-mean and circular symmetry assumption of $\mathbf{y}(n)$, we can see that the klth element of the second term of (17) is $(R_{k,m_1}R_{m_2,l} + R_{k,l}R_{m_2,m_1})/N$, which is recognized as the klth element of $(\mathbf{r}_{m_1}\mathbf{r}_{m_2}^H + R_{m_2,m_1}\mathbf{R})/N$. Here, $R_{k,l}$ denotes the klth element of $\mathbf{R}$. It follows that

$$E\big\{\tilde{\mathbf{r}}_{m_1}\tilde{\mathbf{r}}_{m_2}^H\big\} = \mathbf{r}_{m_1}\mathbf{r}_{m_2}^H + \frac{1}{N}R_{m_2,m_1}\mathbf{R} \tag{18}$$

and, consequently

$$\mathbf{C} \triangleq E\{(\tilde{\mathbf{r}}-\mathbf{r})(\tilde{\mathbf{r}}-\mathbf{r})^H\} = \frac{1}{N}(\mathbf{R}^T \otimes \mathbf{R}) \tag{19}$$

where $\otimes$ denotes the matrix Kronecker product. A natural consistent (in N) estimate of $\mathbf{C}$ is given by

$$\tilde{\mathbf{C}} = \frac{1}{N}(\tilde{\mathbf{R}}^T \otimes \tilde{\mathbf{R}}). \tag{20}$$

Using (20) in (15), we obtain

$$\tilde{\phi} = \arg\min_{\phi \in D_\phi} [\tilde{\mathbf{r}} - \mathbf{r}(\phi)]^H (\tilde{\mathbf{R}}^{-T} \otimes \tilde{\mathbf{R}}^{-1})[\tilde{\mathbf{r}} - \mathbf{r}(\phi)]. \tag{21}$$

Consider next the function $\mathbf{r}(\phi)$ for the case that $\mathbf{R}(\phi)$ is Hermitian Toeplitz. Let

$$\mathbf{Q}_m = \begin{bmatrix} \mathbf{0}_{M-m,m} & \mathbf{I}_{M-m} \\ \mathbf{0}_{m,M-m} & \mathbf{0}_{m,m} \end{bmatrix}, \quad m = 0, 1, \ldots, M-1 \tag{22}$$

where $\mathbf{0}_{r,s}$ denotes the $r \times s$ matrix with zero elements everywhere. Note that $\mathbf{Q}_0 = \mathbf{I}_M$. It follows that

$$\mathbf{R} \triangleq \begin{bmatrix} \rho_0 & \rho_1 & \cdots & \rho_{M-1} \\ \rho_1^* & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_1 \\ \rho_{M-1}^* & \cdots & \rho_1^* & \rho_0 \end{bmatrix} = \rho_0 \mathbf{Q}_0 + \sum_{m=1}^{M-1} \left(\rho_m \mathbf{Q}_m + \rho_m^* \mathbf{Q}_m^T\right). \tag{23}$$

Let

$$\mathbf{\Sigma} = \left[\mathrm{vec}(\mathbf{Q}_0) \quad \mathrm{vec}(\mathbf{Q}_1) \quad \mathrm{vec}\left(\mathbf{Q}_1^T\right) \quad \ldots \quad \mathrm{vec}(\mathbf{Q}_{M-1}) \quad \mathrm{vec}\left(\mathbf{Q}_{M-1}^T\right)\right] \tag{24}$$

and

$$\boldsymbol{\xi} = [\rho_0 \quad \rho_1 \quad \rho_1^* \quad \ldots \quad \rho_{M-1} \quad \rho_{M-1}^*]^T. \tag{25}$$

It follows from (25) that

$$\mathbf{r}(\phi) = \mathbf{\Sigma}\boldsymbol{\xi} = \mathbf{\Sigma} \underbrace{\begin{bmatrix} 1 & & & & & 0 \\ & 1 & j & & & \\ & 1 & -j & & & \\ & & & \ddots & & \\ & & & & 1 & j \\ 0 & & & & 1 & -j \end{bmatrix}}_{\mathbf{\Omega}} \phi = \mathbf{\Psi}\phi \tag{26}$$

where

$$\mathbf{\Psi} = \mathbf{\Sigma}\mathbf{\Omega} \tag{27}$$

and

$$\phi = [\rho_0 \quad \mathrm{Re}(\rho_1) \quad \mathrm{Im}(\rho_1) \quad \cdots \quad \mathrm{Re}(\rho_{M-1}) \quad \mathrm{Im}(\rho_{M-1})]^T. \tag{28}$$

Using (26) in (21) and minimizing the so-obtained quadratic function yields the (asymptotic) ML estimate of ϕ

$$\tilde{\phi} = [\mathrm{Re}(\mathbf{\Psi}^H \tilde{\mathbf{C}}^{-1}\mathbf{\Psi})]^{-1}[\mathrm{Re}(\mathbf{\Psi}^H \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{r}})] \tag{29}$$

where $\tilde{\mathbf{C}}$ is given by (20). To see this, let

$$\bar{\mathbf{r}} = \tilde{\mathbf{C}}^{-1/2}\tilde{\mathbf{r}} \tag{30}$$

and

$$\bar{\mathbf{\Psi}} = \tilde{\mathbf{C}}^{-1/2}\mathbf{\Psi} \tag{31}$$

where $\tilde{\mathbf{C}}^{-1/2}$ denotes the Hermitian square root of $\tilde{\mathbf{C}}^{-1}$. Then

$$\begin{aligned}
&[\tilde{\mathbf{r}} - \mathbf{r}(\phi)]^H \tilde{\mathbf{C}}^{-1}[\tilde{\mathbf{r}} - \mathbf{r}(\phi)] \\
&\quad = \|\bar{\mathbf{r}} - \bar{\mathbf{\Psi}}\phi\|^2 \\
&\quad = \phi^T \bar{\mathbf{\Psi}}^H \bar{\mathbf{\Psi}}\phi - \phi^T \bar{\mathbf{\Psi}}^H \bar{\mathbf{r}} - \bar{\mathbf{r}}^H \bar{\mathbf{\Psi}}\phi + \bar{\mathbf{r}}^H \bar{\mathbf{r}} \\
&\quad = \phi^T [\mathrm{Re}(\bar{\mathbf{\Psi}}^H \bar{\mathbf{\Psi}})]\phi - 2\phi^T \mathrm{Re}(\bar{\mathbf{\Psi}}^H \bar{\mathbf{r}}) + \bar{\mathbf{r}}^H \bar{\mathbf{r}}
\end{aligned} \tag{32}$$

where we have used the fact that $\bar{\mathbf{\Psi}}^H \bar{\mathbf{\Psi}}$ is Hermitian, and hence

$$\phi^T [\mathrm{Im}(\bar{\mathbf{\Psi}}^H \bar{\mathbf{\Psi}})]\phi = 0. \tag{33}$$

Equation (29) immediately follows from (32).

Next, we note that the $\mathrm{Re}(\cdot)$ in (29) can be dropped since both $\mathbf{\Psi}^H \tilde{\mathbf{C}}^{-1}\mathbf{\Psi}$ and $\mathbf{\Psi}^H \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{r}}$ can be shown to be real-valued due to the special structure of $\mathbf{\Psi}$ [22]. With this observation, we obtain the final formula of AML for estimating a Hermitian Toeplitz covariance matrix

$$\tilde{\phi} = (\mathbf{\Psi}^H \tilde{\mathbf{C}}^{-1}\mathbf{\Psi})^{-1}(\mathbf{\Psi}^H \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{r}}). \tag{34}$$

Remark 1: If we relax the Gaussian assumption, then the estimate given by (4) is no longer the ML estimate. In such a case, it seems that the use of the *covariance matching* criterion (15) [or (14)] makes more sense than using (4). Note, however, that the circular symmetry assumption should be maintained; otherwise, the expression for $\mathbf{C} = \mathrm{cov}(\tilde{\mathbf{r}})$ is more complicated.

Remark 2: $\mathbf{R}(\tilde{\phi})$, as given by (34), is not guaranteed to be positive semidefinite. However, this may occur only if $\mathbf{R}$ is close to singular, and N is relatively small. If $N \gg 1$, then by the consistency of $\tilde{\phi}$, the matrix $\mathbf{R}(\tilde{\phi})$ must be positive semidefinite. Our experimental experience suggests that for a number of data samples as small as, for example, $N = 15$, the estimated covariance matrix is always positive semidefinite, even when $\mathbf{R}$ is nearly singular.

IV. Implementation

For large M, directly applying (34) for implementation is computationally intensive and is not recommended because it involves multiplications and inversions of matrices of large dimensions (recall $\tilde{\mathbf{C}} \in \mathbb{C}^{M^2 \times M^2}$ and $\mathbf{\Sigma} \in \mathbb{R}^{M^2 \times (2M-1)}$). However, both $\tilde{\mathbf{C}}$ and $\mathbf{\Psi}$ have very rich structures that can be exploited for efficient implementation.

Redefine $\tilde{\mathbf{C}} \triangleq \tilde{\mathbf{R}}^T \otimes \tilde{\mathbf{R}}$ since dropping $1/N$ in (20) does not affect our solution. We can compute $\tilde{\mathbf{C}}^{-1}\mathbf{\Sigma}$ as

$$\begin{aligned}
&\tilde{\mathbf{C}}^{-1}\mathbf{\Sigma} = \\
&[\mathrm{vec}(\tilde{\mathbf{R}}^{-1}\mathbf{Q}_0\tilde{\mathbf{R}}^{-1}) \; \mathrm{vec}(\tilde{\mathbf{R}}^{-1}\mathbf{Q}_1\tilde{\mathbf{R}}^{-1}) \; \mathrm{vec}[(\tilde{\mathbf{R}}^{-1}\mathbf{Q}_1\tilde{\mathbf{R}}^{-1})^H] \\
&\ldots \; \mathrm{vec}(\tilde{\mathbf{R}}^{-1}\mathbf{Q}_{M-1}\tilde{\mathbf{R}}^{-1}) \; \mathrm{vec}[(\tilde{\mathbf{R}}^{-1}\mathbf{Q}_{M-1}\tilde{\mathbf{R}}^{-1})^H]]
\end{aligned} \tag{35}$$

where we have used the matrix Kronecker product result [23]

$$\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B}). \tag{36}$$

Hence, we need only to compute $\{\tilde{\mathbf{R}}^{-1}\mathbf{Q}_m\tilde{\mathbf{R}}^{-1}\}_{m=0}^{M-1}$. Let $\mathbf{E}_{m_1,m_2} \in \mathbb{R}^{M\times M}$ denote the elementary matrix with unit m_1m_2th element and zero elsewhere. We then have

$$\mathbf{Q}_m = \sum_{i=1}^{M-m} \mathbf{E}_{i,i+m}, \quad m = 0,1,\ldots,M-1 \tag{37}$$

and

$$\mathbf{Q}_m^T = \sum_{i=1}^{M-m} \mathbf{E}_{i+m,i}, \quad m = 0,1,\ldots,M-1. \tag{38}$$

Note that

$$\mathbf{A}\mathbf{E}_{m_1,m_2}\mathbf{B} = \mathbf{A}_{:,m_1}\mathbf{B}_{m_2,:} \tag{39}$$

where $\mathbf{A}_{:,m_1}$ denotes the column vector that is the m_1th column of $\mathbf{A}$, and $\mathbf{B}_{m_2,:}$ denotes the row vector made from the m_2th row of $\mathbf{B}$. Hence, $\tilde{\mathbf{R}}^{-1}\mathbf{Q}_m\tilde{\mathbf{R}}^{-1}$ is easily calculated as a sum of vector outer products

$$\tilde{\mathbf{R}}^{-1}\mathbf{Q}_m\tilde{\mathbf{R}}^{-1} = \sum_{i=1}^{M-m} (\tilde{\mathbf{R}}^{-1})_{:,i}(\tilde{\mathbf{R}}^{-1})_{i+m,:}. \tag{40}$$

We next notice that some elements of $\mathbf{\Sigma}^T\tilde{\mathbf{C}}^{-1}\mathbf{\Sigma}$ have the form

$$\text{vec}^T(\mathbf{Q}_{m_1})\text{vec}(\tilde{\mathbf{R}}^{-1}\mathbf{Q}_{m_2}\tilde{\mathbf{R}}^{-1}), \quad m_1, m_2 = 0,1,\ldots,M-1. \tag{41}$$

The other elements of $\mathbf{\Sigma}^T\tilde{\mathbf{C}}^{-1}\mathbf{\Sigma}$ have similar forms but with the argument of the first vec$(\cdot)$ changed to $\mathbf{Q}_{m_1}^T$ and/or the argument of the second vec$(\cdot)$ changed to $(\tilde{\mathbf{R}}^{-1}\mathbf{Q}_{m_2}\tilde{\mathbf{R}}^{-1})^H$. We can compute (41) as

$$\begin{aligned}\text{vec}^T&(\mathbf{Q}_{m_1})\text{vec}(\tilde{\mathbf{R}}^{-1}\mathbf{Q}_{m_2}\tilde{\mathbf{R}}^{-1})\\ &= \text{tr}\left[\mathbf{Q}_{m_1}^T\tilde{\mathbf{R}}^{-1}\mathbf{Q}_{m_2}\tilde{\mathbf{R}}^{-1}\right]\\ &= \text{tr}\left[\sum_{i=1}^{M-m_1}\mathbf{E}_{i+m_1,i}(\tilde{\mathbf{R}}^{-1}\mathbf{Q}_{m_2}\tilde{\mathbf{R}}^{-1})\right]\\ &= \text{tr}\left[\sum_{i=1}^{M-m_1}\mathbf{e}_{i+m_1}(\tilde{\mathbf{R}}^{-1}\mathbf{Q}_{m_2}\tilde{\mathbf{R}}^{-1})_{i,:}\right]\\ &= \sum_{i=1}^{M-m_1}(\tilde{\mathbf{R}}^{-1}\mathbf{Q}_{m_2}\tilde{\mathbf{R}}^{-1})_{i,i+m_1}\end{aligned} \tag{42}$$

where $\mathbf{e}_i \in \mathbb{R}^{M\times 1}$ denotes the vector with the unit ith element and zero elsewhere, and we have made use of the facts that

$$\text{vec}^T(\mathbf{A})\text{vec}(\mathbf{B}) = \text{tr}(\mathbf{A}^T\mathbf{B}) \tag{43}$$

and

$$\mathbf{E}_{m_1,m_2}\mathbf{A} = \mathbf{e}_{m_1}\mathbf{A}_{m_2,:}. \tag{44}$$

The other elements of $\mathbf{\Sigma}^T\tilde{\mathbf{C}}^{-1}\mathbf{\Sigma}$ can be similarly calculated. Note that (44) requires no computation since $\mathbf{e}_{m_1}\mathbf{A}_{m_2,:}$ simply has zero elements everywhere, except that its m_1th row is the same as the m_2th row of $\mathbf{A}$, which explains how we arrived at the last equality of (42). Thus, once we have obtained $\tilde{\mathbf{C}}^{-1}\mathbf{\Sigma}$, the calculation of $\mathbf{\Sigma}^T\tilde{\mathbf{C}}^{-1}\mathbf{\Sigma}$ requires only a few additions.

Similarly, we decompose $\mathbf{\Omega}$ into a sum of elementary matrices

$$\mathbf{\Omega} = \mathbf{\Omega}_r + j\mathbf{\Omega}_i \tag{45}$$

where

$$\mathbf{\Omega}_r \triangleq \mathbf{E}_{1,1} + \sum_{m=1}^{M-1}(\mathbf{E}_{2m,2m} + \mathbf{E}_{2m+1,2m}) \tag{46}$$

and

$$\mathbf{\Omega}_i \triangleq \sum_{m=1}^{M-1}(\mathbf{E}_{2m,2m+1} - \mathbf{E}_{2m+1,2m+1}). \tag{47}$$

(Note that in (46) and (47) $\mathbf{E}_{m_1,m_2} \in \mathbb{R}^{(2M-1)\times(2M-1)}$.) Consequently, we have

$$\begin{aligned}\mathbf{\Psi}^H\tilde{\mathbf{C}}^{-1}\mathbf{\Psi} &= \mathbf{\Omega}^H(\mathbf{\Sigma}^T\tilde{\mathbf{C}}^{-1}\mathbf{\Sigma})\mathbf{\Omega}\\ &= \mathbf{\Omega}_r^T(\mathbf{\Sigma}^T\tilde{\mathbf{C}}^{-1}\mathbf{\Sigma})\mathbf{\Omega}_r + \mathbf{\Omega}_i^T(\mathbf{\Sigma}^T\tilde{\mathbf{C}}^{-1}\mathbf{\Sigma})\mathbf{\Omega}_i\\ &\quad + j\big[\mathbf{\Omega}_r^T(\mathbf{\Sigma}^T\tilde{\mathbf{C}}^{-1}\mathbf{\Sigma})\mathbf{\Omega}_i\\ &\quad - \big(\mathbf{\Omega}_r^T(\mathbf{\Sigma}^T\tilde{\mathbf{C}}^{-1}\mathbf{\Sigma})\mathbf{\Omega}_i\big)^H\big].\end{aligned} \tag{48}$$

We can see that computing $\mathbf{\Omega}^H(\mathbf{\Sigma}^T\tilde{\mathbf{C}}^{-1}\mathbf{\Sigma})\mathbf{\Omega}$ from $\mathbf{\Sigma}^T\tilde{\mathbf{C}}^{-1}\mathbf{\Sigma}$ reduces to a few additions. Specifically, to calculate the premultiplication with $\mathbf{\Omega}_r^T$ or $\mathbf{\Omega}_i^T$, we use (44), whereas to calculate the postmultiplication with $\mathbf{\Omega}_r$ or $\mathbf{\Omega}_i$, we use

$$\mathbf{A}\mathbf{E}_{m_1,m_2} = \mathbf{A}_{:,m_1}\mathbf{e}_{m_2}^T. \tag{49}$$

Again, $\mathbf{A}_{:,m_1}\mathbf{e}_{m_2}^T$ costs no computation and has zero elements everywhere, except that its m_2th column is the same as the m_1th column of $\mathbf{A}$.

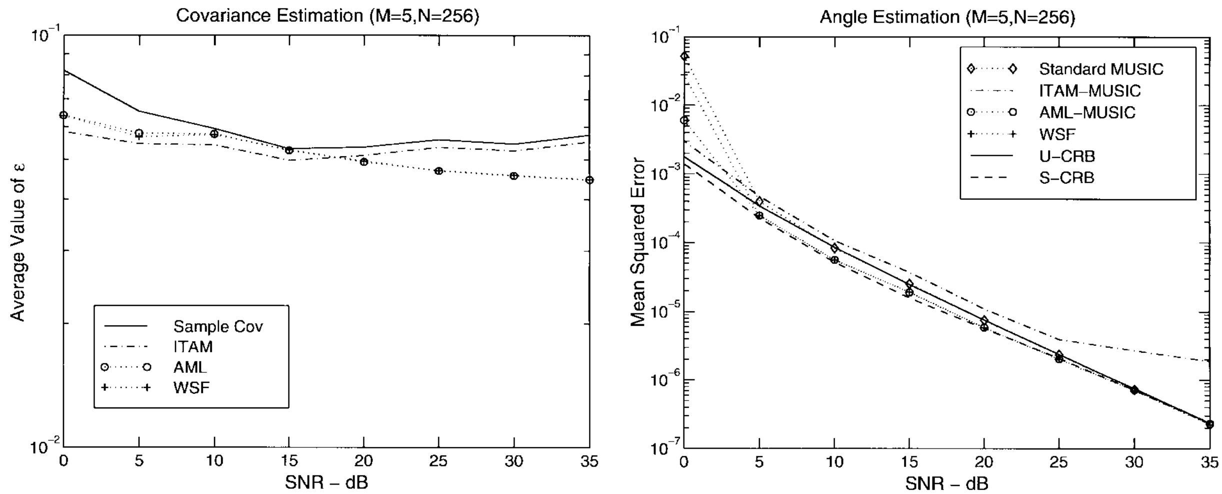
By using the fact that

$$\begin{aligned}\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{r}} &= (\tilde{\mathbf{R}}^{-T}\otimes\tilde{\mathbf{R}}^{-1})\text{vec}(\tilde{\mathbf{R}}) = \text{vec}(\tilde{\mathbf{R}}^{-1}\tilde{\mathbf{R}}\tilde{\mathbf{R}}^{-1})\\ &= \text{vec}(\tilde{\mathbf{R}}^{-1})\end{aligned} \tag{50}$$

we can see that the previous techniques used for calculating $\mathbf{\Psi}^H\tilde{\mathbf{C}}^{-1}\mathbf{\Psi}$ can also be used to calculate $\mathbf{\Psi}^H\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{r}} = \mathbf{\Omega}^H\mathbf{\Sigma}^T\text{vec}(\tilde{\mathbf{R}}^{-1})$.

In summary, the AML estimator can be implemented as follows.

1) Obtain the sample covariance matrix $\tilde{\mathbf{R}}$ by (6) and its inverse $\tilde{\mathbf{R}}^{-1}$.
2) Compute $\tilde{\mathbf{C}}^{-1}\mathbf{\Sigma}$ and $\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{r}}$ by using (35) and (50), respectively.
3) Calculate $\mathbf{\Sigma}^T\tilde{\mathbf{C}}^{-1}\mathbf{\Sigma}$ in an elementwise manner by using (42), and compute $\mathbf{\Sigma}^T\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{r}}$ similarly.
4) Decompose $\mathbf{\Omega}^H(\mathbf{\Sigma}^T\tilde{\mathbf{C}}^{-1}\mathbf{\Sigma})\mathbf{\Omega}$ by using (48) as well as (44) and (49). In addition, compute $\mathbf{\Omega}^H\mathbf{\Sigma}^T\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{r}}$ similarly.
5) Finally, calculate $\tilde{\phi}$ by (34), and construct the AML covariance matrix estimate $\mathbf{R}(\tilde{\phi})$ from $\tilde{\phi}$.



(a) (b)

Fig. 1. (a) Average values of ϵ, the normalized Frobenius distance between the estimated covariance matrix and the true one, versus SNR when $N = 256$ and $M = 5$. (b) MSE's of the estimates of θ_1 and the corresponding U-CRB and S-CRB versus SNR when $N = 256$ and $M = 5$.

Assume that the data is complex-valued and $N \gg M$. Then, calculating $\tilde{\mathbf{R}}$ by (6) and $\tilde{\mathbf{C}}^{-1}\mathbf{\Sigma}$ by (35) requires approximately $4M^2N$ and $10M^4$ flops,[1] respectively, which represent the most involved steps in the AML algorithm. Computing the inverse of $\mathbf{\Psi}^H\tilde{\mathbf{C}}^{-1}\mathbf{\Psi}$ needs an additional number of flops of $O((2M-1)^3)$, which is approximately $70M^3$ flops in MATLAB. Computations needed by the remaining steps of AML are negligible. This gives a rough estimate of the number of MATLAB flops required by AML to be $4M^2N + 10M^4 + 70M^3$. Our simulations show that this number tends to slightly overestimate the true number of flops required by AML when M is moderate or large.

V. Numerical Examples

In this section, we provide a numerical study of the AML covariance matrix estimator. Comparisons are made with several typical unstructured and structured covariance matrix estimators, namely, the sample covariance matrix estimator, ITAM, and the WSF algorithm. Our primary interest herein is to compare the performances of using the various covariance matrix estimators in angle estimation. Whereas the WSF algorithm is essentially an angle estimator, the other three estimators are used with the root-MUSIC algorithm [12] for angle estimation. (Even though they are not shown here, we find that using ESPRIT with the ITAM or AML covariance matrices yields similar results as using root-MUSIC.) In what follows, we also compare covariance matrix estimation in terms of the Frobenius norms of the differences between the estimated covariance matrices and the true $\mathbf{R}$. (As [16] does not address the covariance estimation problem, we briefly explain in Appendix A how to use the WSF algorithm to estimate the structured covariance matrix.) When a covariance matrix estimator is used in a specific application, the Frobenius distance may not be an appropriate measure to assess the quality of that covariance matrix estimator since the Frobenius norm measure ignores the fine structures, such as the eigenstructure, of the covariance matrix, which may be of particular interest in that application.

The results shown below are all based on 500 independent realizations.

A. Performance versus SNR

Consider the problem of estimating the arrival angles $\theta_1 = 0°$ and $\theta_2 = 5°$ of two uncorrelated signals with equal power impinging on a ULA of $M = 5$ sensors separated by a half wavelength. Let $N = 256$, and define the SNR for the kth incoming signal as

$$\mathrm{SNR}_k \triangleq 10\log_{10}\frac{s_k}{\sigma^2} \tag{51}$$

where s_k denotes the kth element of the $\mathbf{s}$ in (3), which is the variance of the kth signal. Let

$$\epsilon \triangleq \frac{\|\mathbf{R}-\hat{\mathbf{R}}\|_F}{\|\mathbf{R}\|_F} \tag{52}$$

where $\hat{\mathbf{R}}$ denotes any of the four covariance matrix estimates, and $\|\cdot\|_F$ denotes the matrix Frobenius norm. The average values of ϵ versus SNR for the four covariance matrix estimators are shown in Fig. 1(a). In terms of the Frobenius distance to the true covariance matrix $\mathbf{R}$, the three structured covariance matrix estimators yield better covariance matrix estimates than the unstructured sample covariance matrix. In addition, the ITAM covariance matrix is worse than both the AML and the WSF covariance matrices when the SNR is moderately high, whereas the latter two appear to perform similarly.

By the Carathéodory parameterization, any positive semidefinite Hermitian Toeplitz matrix can be expressed as in (1) with $K < M$ [17], thereby reducing the covariance matrix estimation problem to that of estimating the *equivalent* parameter vector $\boldsymbol{\phi}$ in (3), which is a methodology adopted by WSF when used for covariance matrix estimation. Note that WSF uses $2K+1$ unknowns, whereas AML has $2M-1$

[1] A flop denotes either a floating-point addition or a floating-point multiplication, as adopted by MATLAB.

unknowns. If $K < M - 1$, then, according to the *parsimony principle* [20], WSF is likely to provide more accurate covariance matrix estimates than AML because the former uses fewer number of unknowns than the latter. Yet, WSF has its own problems, which may affect its performance. For example, WSF assumes the knowledge of the exact number of incoming "signals" K, whereas AML does not. In scenarios where the "SNR" is low or some "signals" are close to one another, an accurate estimate of K is usually difficult to obtain. In such cases, it may be desirable to use AML instead of WSF because using incorrect value of K in general degrades WSF significantly. Another reason (perhaps, the more important one) to prefer AML over WSF is AML's computational simplicity, as illustrated later.

Next, we investigate the angle estimation problem. Fig. 1(b) shows the mean-squared errors (MSE's) of the various angle estimates of θ_1 and the corresponding U-CRB and S-CRB versus SNR. We note the following.

- The standard MUSIC approaches the U-CRB as the SNR increases, which is a well-known fact [15], [19].
- AML-MUSIC and WSF are asymptotically very close to the S-CRB (this happens for SNR ≥ 5 dB in the present case), with AML-MUSIC having a lower threshold SNR than WSF.
- The difference between the U-CRB and the S-CRB is small for high SNR's.
- ITAM-MUSIC never attains the S-CRB and performs worse than the standard MUSIC when the SNR increases.

The ITAM estimator was originally proposed as an algorithm that can be used to enhance the performance of such algorithms as MUSIC and ESPRIT when the SNR is relatively low [7]. As also indicated by Fig. 1(b), ITAM's performance is indeed quite good at low SNR values. However, ITAM is not an optimal method, and therefore, there is no surprise that ITAM-MUSIC never achieves the S-CRB. On the other hand, the inconsistency (in SNR) of ITAM appears surprising at first sight. To explain it briefly, note that as the SNR goes to infinity, we have

$$\lim_{\sigma^2 \to 0} \tilde{\mathbf{R}} = \mathbf{A}(\boldsymbol{\theta})\tilde{\mathbf{S}}\mathbf{A}^H(\boldsymbol{\theta}) \tag{53}$$

where $\tilde{\mathbf{S}}$ is the sample signal covariance matrix that is *not* diagonal for finite N. In spite of the fact that $\tilde{\mathbf{R}}$ is not Toeplitz in this case, the *signal* and *noise subspaces* can be obtained *exactly* from $\tilde{\mathbf{R}}$ when the SNR goes to infinity. For a subspace-based algorithm like root-MUSIC, perfect angle estimates can be obtained if exact subspace estimation is available. However, ITAM attempts to find a Toeplitz matrix that is as close to $\tilde{\mathbf{R}}$ as possible but makes no effort to ensure appropriate subspace approximation. As a result, the subspaces of $\tilde{\mathbf{R}}$ are distorted by the sequences of approximations introduced by ITAM. Hence, ITAM-MUSIC is inconsistent in SNR.

It is interesting to note that, even though AML assumes a Toeplitz structure as ITAM, it does not suffer from the inconsistency problem of ITAM. In Appendix B, we show that using the AML criterion (21) in array processing when the SNR is high is equivalent to seeking a Toeplitz matrix that is the closest to the range space of $\mathbf{A}(\boldsymbol{\theta})$. Consequently, the AML covariance matrix estimate provides consistent (in SNR) subspace estimates, and AML-MUSIC in turn yields consistent angle estimates as the SNR increases.

B. Performance versus Snapshot Number

In this example, we study the effect of the snapshot number N on the performance of the various covariance matrix and angle estimators. The parameters are the same as in the previous example, except that we fix SNR $= 10$ dB and vary N from 10 to 1000. Fig. 2(a) and (b), respectively, shows the average values of ϵ and the MSE's of the estimates of θ_1 versus N. As we can see from Fig. 2(b), the standard MUSIC approaches the U-CRB for sufficiently large N, whereas AML-MUSIC has similar performance as WSF and is very close to the S-CRB when $N \geq 40$ for the present case. In addition, we note that AML has the lowest threshold N in this example. On the other hand, ITAM-MUSIC cannot achieve the S-CRB (or even the U-CRB) for all N values considered. Increasing N does not help ITAM-MUSIC much, even though the ITAM covariance matrices are always closer to the true $\mathbf{R}$ in the Frobenius norm sense than the sample covariance matrices, as shown in Fig. 2(a). Fig. 2(a) also shows that as in the previous example, AML and WSF perform quite similarly for covariance estimation, and both are asymptotic methods and provide better estimates than the sample covariance matrix when N is large enough.

In the next two examples, we will no longer show the covariance matrix estimation results in the form of the ϵ-error since we found that little additional insight can be gained from them.

C. Performance versus Angle Separation

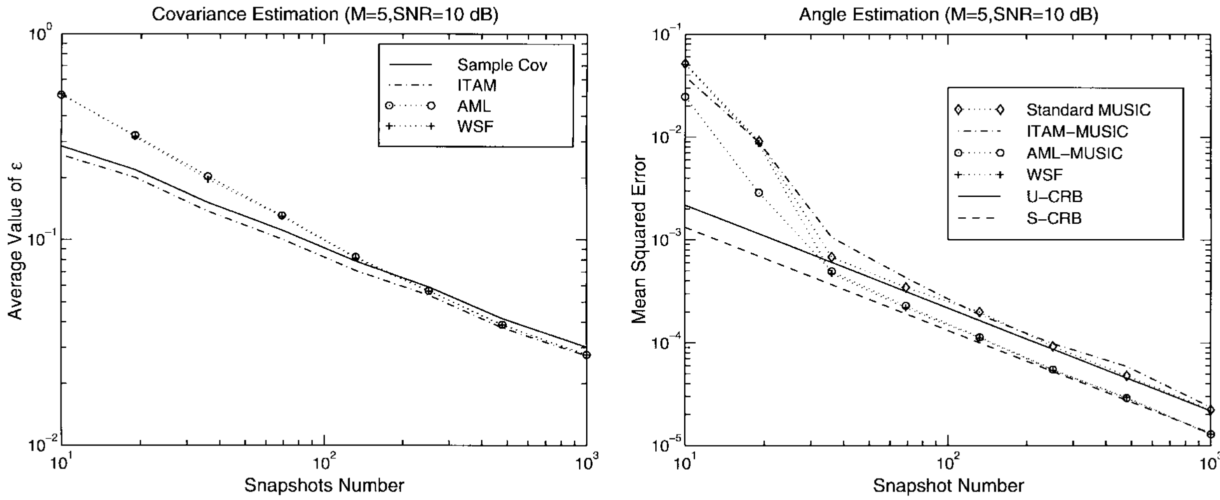
The parameters used for this example are the same as in the first example, except that SNR is fixed at SNR $= 10$ dB, and the signals are from $\theta_1 = 0^\circ$ and $\theta_2 = \Delta\theta$, with $\Delta\theta$ varying from 1° to 10°. Fig. 3 shows the MSE's of the estimates of θ_1 versus $\Delta\theta$. It can be seen for the present case that when $\Delta\theta > 2^\circ$, AML-MUSIC and WSF approaches the S-CRB, and the standard MUSIC approaches the U-CRB. Again, AML-MUSIC has the lowest threshold $\Delta\theta$, whereas ITAM-MUSIC gives the worst angle estimates for most $\Delta\theta$ considered.

D. Performance versus Sensor Number

Next, we study the impact of varying the sensor number M on angle estimation. We fix SNR $= 10$ dB, $N = 256$, and vary M from 3 to 10. All the other parameters are the same as in the first example. The MSE's of θ_1 versus M are shown in Fig. 4. Like in the previous examples, AML-MUSIC and WSF have similar performances and are very close to the S-CRB, the standard MUSIC approaches the U-CRB, and ITAM-MUSIC is the worst for most of the sensor number choices.

E. Computational Complexities

The previous examples have shown that in terms of angle estimation, AML-MUSIC and WSF in general perform



(a) (b)

Fig. 2. (a) Average values of ϵ, the normalized Frobenius distance between the estimated covariance matrix and the true one, versus N when SNR $= 10$ dB and $M = 5$. (b) MSE's of the estimates of θ_1 and the corresponding U-CRB and S-CRB versus N when SNR $= 10$ dB and $M = 5$.

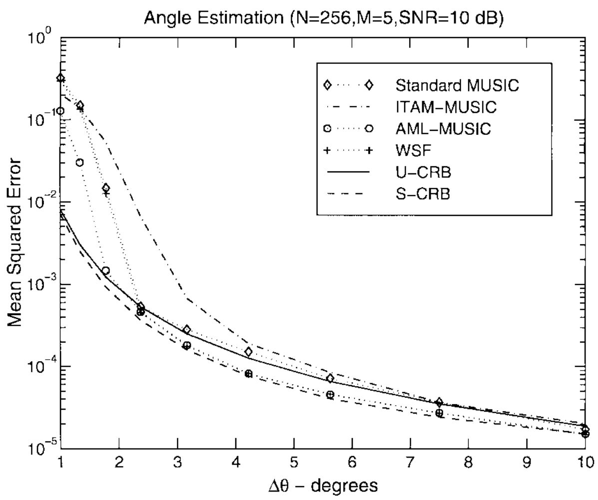


Fig. 3. MSE's of the estimates of θ_1 and the corresponding U-CRB and S-CRB versus $\Delta\theta$ when SNR $= 10$ dB and $M = 5$.

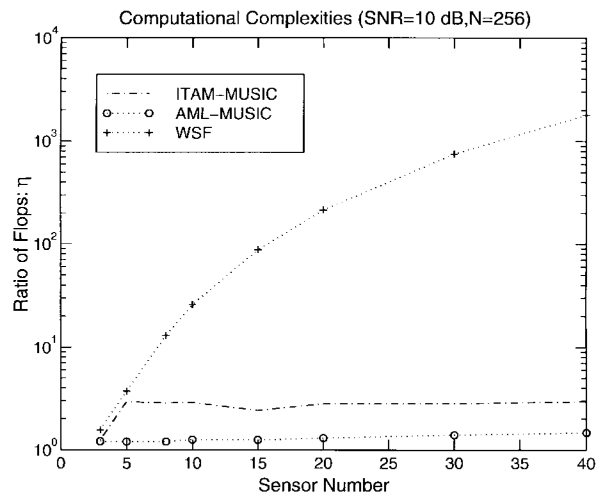


Fig. 5. Flop ratio η versus M when SNR $= 10$ dB and $N = 256$.

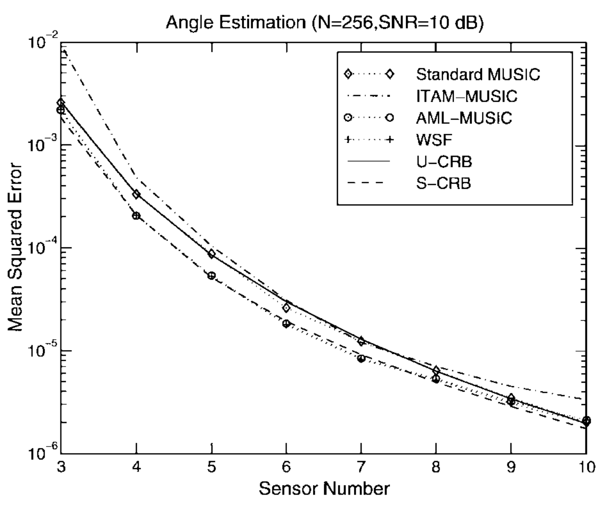


Fig. 4. MSE's of the estimates of θ_1 and the corresponding U-CRB and S-CRB versus M when SNR $= 10$ dB and $N = 256$.

similarly, except that in some difficult scenarios, such as at low SNR's or when the arrival angles of some signals are closely spaced, AML-MUSIC is slightly better than WSF. We shall emphasize here that AML and WSF are quite different algorithms in that AML is derived specifically for covariance matrix estimation, whereas WSF is primarily for angle (or frequency) estimation. In the following example, we show that using AML in angle estimation introduces very modest additional computations. Define η as the ratio of the number of MATLAB flops needed by ITAM-MUSIC, AML-MUSIC, or WSF to that by the standard MUSIC. The parameters are the same as in the previous example, except that M is varied from 3 to 40. Fig. 5 shows the curves of η versus M. It is seen that AML-MUSIC is computationally the most efficient algorithm. It should be mentioned that for WSF, we used a MATLAB code provided by one of the authors of [16]. The code was provided without any computational optimality claim, and hence, faster implementation of WSF may be possible.

VI. Conclusion

We have presented an asymptotic maximum likelihood method, referred to as AML, for structured covariance matrix estimation. A closed-form formula for Toeplitz covariance matrix estimation has been derived that facilitates computationally efficient implementation of the AML algorithm. We have shown that using the AML covariance matrix estimate with MUSIC improves the angle estimation accuracy of MUSIC. Comparisons with other typical covariance matrix estimators have also been made, and the superiority of AML has been established by examining how the covariance matrix estimators influence the angle estimation accuracy.

We have also shown that the Frobenius distance is, in general, not an appropriate measure to assess the quality of covariance matrix estimates in applications where the fine structure of the estimated covariance matrix plays an important role. For subspace-based algorithms, such as MUSIC, the performance is critically dependent on the accuracy of subspace approximation. The ITAM estimator cannot provide consistent (in SNR) estimates of the relevant subspaces, and hence, ITAM-MUSIC is an inconsistent angle estimator, despite the fact that the ITAM covariance matrix estimate is, in general, better than the sample covariance matrix in terms of the Frobenius distance.

Finally, we shall stress that even though we only considered Hermitian Toeplitz matrix estimation in this paper, it is straightforward to extend the proposed technique to estimate any other matrix that has a linear structure.

Appendix A
Using WSF for Covariance Estimation

After the WSF estimate of $\boldsymbol{\theta}$ is obtained, we need to find the estimates of the signal and noise power, i.e., $\mathbf{s}$ and σ^2, in order to get a structured covariance matrix estimate by using (1). In this appendix, we describe a method that provides the asymptotic ML estimates of $\mathbf{s}$ and σ^2, again by exploiting the *covariance matching* criterion.

Let $\boldsymbol{\gamma} = [\mathbf{s}^T \ \ \sigma^2]^T$. We first rewrite (1) as

$$\mathbf{r}(\boldsymbol{\gamma}) \triangleq \text{vec}(\mathbf{R}) = [\mathbf{A}^*(\boldsymbol{\theta}) \otimes \mathbf{A}(\boldsymbol{\theta})]\text{vec}(\mathbf{S}) + \text{vec}(\mathbf{I}_M)\sigma^2. \tag{54}$$

Observe that

$$\text{vec}(\mathbf{S}) = \mathbf{H}\mathbf{s} \tag{55}$$

where $\mathbf{H} \in \mathbb{R}^{K^2 \times K}$ is the selection matrix that has the form

$$\mathbf{H} = \sum_{k=1}^{K} \mathbf{E}_{(k-1)K+k,k}. \tag{56}$$

Here, $\mathbf{E}_{k,l} \in \mathbb{R}^{K^2 \times K}$ denotes the elementary matrix with unit klth element and zero elsewhere. By substituting (55) into (54), we have

$$\mathbf{r}(\boldsymbol{\gamma}) = [\{\mathbf{A}^*(\boldsymbol{\theta}) \otimes \mathbf{A}(\boldsymbol{\theta})\}\mathbf{H} \quad \text{vec}(\mathbf{I}_M)]\begin{bmatrix}\mathbf{s}\\ \sigma^2\end{bmatrix} \triangleq \mathbf{G}\boldsymbol{\gamma}. \tag{57}$$

Next, we invoke the covariance matching criterion

$$\tilde{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}} [\tilde{\mathbf{r}} - \mathbf{r}(\boldsymbol{\gamma})]^H \tilde{\mathbf{C}}^{-1} [\tilde{\mathbf{r}} - \mathbf{r}(\boldsymbol{\gamma})]. \tag{58}$$

According to EXIP, $\tilde{\boldsymbol{\gamma}}$ is an asymptotic ML estimate of $\boldsymbol{\gamma}$. To solve the above WLS problem, we use the WSF estimate of $\boldsymbol{\theta}$ to replace the $\boldsymbol{\theta}$ in $\mathbf{G}$. The solution is then obtained as [see (30)–(34)]

$$\tilde{\boldsymbol{\gamma}} = (\mathbf{G}^H \tilde{\mathbf{C}}^{-1} \mathbf{G})^{-1} (\mathbf{G}^H \tilde{\mathbf{C}}^{-1} \tilde{\mathbf{r}}). \tag{59}$$

Appendix B
Analysis of AML at High SNR

In this appendix, we show that the AML estimate of the spatial covariance matrix $\mathbf{R}$ provides accurate estimate of the signal subspace of $\mathbf{R}$ at high SNR. Let the eigendecomposition of $\tilde{\mathbf{R}}$ be

$$\tilde{\mathbf{R}} = \tilde{\mathbf{E}}_s \tilde{\boldsymbol{\Lambda}}_s \tilde{\mathbf{E}}_s^H + \tilde{\mathbf{E}}_n \tilde{\boldsymbol{\Lambda}}_n \tilde{\mathbf{E}}_n^H \tag{60}$$

where $\tilde{\boldsymbol{\Lambda}}_s$ is the diagonal matrix containing the K largest eigenvalues, with the columns of $\tilde{\mathbf{E}}_s$ being the associated eigenvectors, and $\tilde{\boldsymbol{\Lambda}}_n$ is the diagonal matrix containing the remaining eigenvalues with the columns of $\tilde{\mathbf{E}}_n$ being the corresponding eigenvectors. Since, for sufficiently small σ^2, we have

$$\tilde{\boldsymbol{\Lambda}}_n = O(\sigma^2) \tag{61}$$

it follows that

$$\tilde{\mathbf{R}}^{-1} = \tilde{\mathbf{E}}_s \tilde{\boldsymbol{\Lambda}}_s^{-1} \tilde{\mathbf{E}}_s^H + \tilde{\mathbf{E}}_n \tilde{\boldsymbol{\Lambda}}_n^{-1} \tilde{\mathbf{E}}_n^H \approx \tilde{\mathbf{E}}_n \tilde{\boldsymbol{\Lambda}}_n^{-1} \tilde{\mathbf{E}}_n^H. \tag{62}$$

Hence, we can rewrite the cost function in (21) at high SNR as

$$\begin{aligned} &[\text{vec}(\tilde{\mathbf{R}}^T - \mathbf{R}^T(\phi))]^T (\tilde{\mathbf{R}}^{-T} \otimes \tilde{\mathbf{R}}^{-1}) [\text{vec}(\tilde{\mathbf{R}} - \mathbf{R}(\phi))] \\ &\quad = \text{tr}[(\tilde{\mathbf{R}} - \mathbf{R}(\phi))\tilde{\mathbf{R}}^{-1}(\tilde{\mathbf{R}} - \mathbf{R}(\phi))\tilde{\mathbf{R}}^{-1}] \\ &\quad \approx \text{tr}\left[\left(\mathbf{I}_M - \mathbf{R}(\phi)\tilde{\mathbf{E}}_n \tilde{\boldsymbol{\Lambda}}_n^{-1} \tilde{\mathbf{E}}_n^H\right)^2\right] \end{aligned} \tag{63}$$

where we have used the fact that [23]

$$\text{tr}(\mathbf{ABCD}) = \text{vec}^T(\mathbf{A}^T)(\mathbf{D}^T \otimes \mathbf{B})\text{vec}(\mathbf{C}). \tag{64}$$

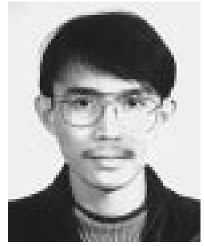
The second term in (63) is of the order $O(\sigma^{-2})$ and, hence, is the dominant one. To minimize the criterion function in (63), the AML estimation $\mathbf{R}(\tilde{\phi})$ of $\mathbf{R}$ must minimize this term (as $\sigma^2 \to 0$). However, this is only possible if the null space of $\mathbf{R}(\tilde{\phi})$ is close to $\tilde{\mathbf{E}}_n$, which implies that the range space of $\mathbf{R}(\tilde{\phi})$ is close to that of $\tilde{\mathbf{E}}_s$, which, in turn, approaches the range space of $\mathbf{A}(\boldsymbol{\theta})$ (the so-called signal subspace) as $\sigma^2 \to 0$.

Acknowledgment

The authors would like to express their gratitude to Dr. M. Jansson for sharing his MATLAB code of the WSF algorithm and for the helpful discussions with us. They also thank the anonymous referees, whose comments led to a significant improvement of the manuscript.

REFERENCES

[1] T. W. Anderson, "Asymptotically efficient estimation of covariance matrices with linear structure," *Ann. Stat.*, vol, 1, no. 1, pp. 135–141, 1973.
[2] J. P. Burg and D. G. Luenberger, and D. L. Wenger, "Estimation of structured covariance matrices," *Proc. IEEE*, vol. 70, pp. 963–974, Sept. 1982.
[3] S. Degerine, "Maximum likelihood estimation of autocovariance matrices from replicated short time series," *J. Time Series Anal.*, vol. 8, pp. 135–146, 1987.
[4] D. R. Fuhrmann and M. I. Miller, "On the existence of positive-definite maximum-likelihood estimates of structured covariance matrices," *IEEE Trans. Inform. Theory*, vol. 34, pp. 722–729, July 1988.
[5] D. B. Williams and D. H. Johnson, "Robust estimation of structured covariance matrices," *IEEE Trans. Signal Processing*, vol. 41, pp. 2891–2906, Sept. 1993.
[6] H. Wang and G. H. Wakefield, "Signal-subspace approximation for line spectrum estimation," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Dallas, TX, Apr. 1987, pp. 2054–2057.
[7] J. A. Cadzow, "Signal enhancement—A composite property mapping algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 49–62, Jan. 1988.
[8] D. M. Wilkes and M. H. Hayes, "Iterated Toeplitz approximation of covariance matrices," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, New York, Apr. 1988, pp. 1663–1666.
[9] H. Li, P. Stoica, and J. Li, "Capon estimation of covariance sequences," *Circuits, Syst., Signal Process.*, vol. 17, no. 1, pp. 29–49, Jan. 1998.
[10] D. R. Fuhrmann, "Application of Toeplitz covariance estimation to adaptive beamforming and detection," *IEEE Trans. Signal Processing*, vol. 39, pp. 2194–2198, Oct. 1991.
[11] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. AP-34, pp. 276–280, Mar. 1985.
[12] A. J. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Boston, MA, Apr. 1983, pp. 336–339.
[13] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.
[14] P. Stoica and T. Söderström, "On reparameterization of loss functions used in estimation and the invariance principle," *Signal Process.*, vol. 17, pp. 383–387, Aug. 1989.
[15] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér-Rao bound," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 720–741, May 1989.
[16] M. Jansson, B. Göransson, and B. Ottersten, "Analysis of a subspace-based spatial frequency estimator," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Munich, Germany, Apr. 1997.
[17] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1997.
[18] R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 984–995, July 1989.
[19] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér-Rao bound: Further results and comparisons," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 2140–2150, May 1990.
[20] T. Söderström and P. Stoica, *System Identification*. London, U.K.: Prentice-Hall, 1989.
[21] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," to appear in *Digital Signal Process.*
[22] S. Zacks, *Parametric Statistical Inference: Basic Theory and Modern Approaches*. Oxford, U.K.: Pergamon, 1981.
[23] A. Graham, *Kronecker Products and Matrix Calculus with Applications*. Chichester, U.K.: Ellis Horwood, 1981.



Hongbin Li (S'98) received the B.S. and M.S. degrees in electrical engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1991 and 1994, respectively. He is currently pursuing the Ph.D. degree in electrical engineering at the University of Florida, Gainesville.

From 1994 to 1995, he was an Assistant Lecturer at UESTC. Since 1996, he has been a Research Assistant with the Department of Electrical and Computer Engineering, University of Florida. His current research interests include digital signal processing, detection and estimation theory, and wireless communications.

Mr. Li is a member of Tau Beta Pi and Phi Kappa Phi. He received the 1999 Graduate Research Award from the University of Florida Chapter of Sigma Xi.

Petre Stoica (F'94) received the M.Sc. and D.Sc. degrees, both in automatic control, from the Bucharest Polytechnic Institute, Bucharest, Romania, in 1972 and 1979, respectively. In 1993, he received an Honorary Doctorate degree from Uppsala University, Sweden.

His scientific interests include system identification, time series analysis and prediction, statistical signal and array processing, spectral analysis, and radar signal processing. He has published some 350 papers in journals and conference records on these topics. He has also published eight book chapters and seven books, of which the most recent one is *Introduction to Spectral Analysis* [Englewood Cliffs, NJ: Prentice-Hall, 1997 (coauthored with R. Moses)]. He was associated with the Department of Automatic Control and Computers, Polytechnic Institute of Bucharest, recently as Professor of System Identification and Signal Processing. He spent 1992, 1993, and the first half of 1994 with the Systems and Control Group, Uppsala University, Uppsala, Sweden, as a Guest Professor. In the second half of 1994, he held a 150th Anniversary Visiting Professorship with the Applied Electronics Department at Chalmers University of Technology, Gothenburg, Sweden. Currently, he is affiliated with the Systems and Control Group, Uppsala University, where he holds the position of Professor of System Modeling.

Dr. Stoica was co-recipient (with A. Nehorai) of the 1989 IEEE ASSP Senior Award for contributions to array processing and recipient of the 1996 Technical Achievement Award of the IEEE Signal Processing Society for fundamental contributions to statistical signal processing, as well as of several other awards and prizes. He is listed in *Who's Who in the World*. He is also on the editorial boards of four journals in the field: *Signal Processing*, *Circuits, Systems and Signal Processing*, *Multidimensional Systems and Signal Processing*, and *Journal of Forecasting*. He is a Corresponding Member of the Romanian Academy and the Royal Statistical Society.

Jian Li (S'87–M'91–SM'97), for photograph and biography, see this issue, p. 1271.