ON SUBSPACE BASED SINUSOIDAL FREQUENCY ESTIMATION

Martin Kristensson , Magnus Jansson and Björn Ottersten

Department of Signals, Sensors, & Systems
Royal Institute of Technology (KTH)
Sweden

ABSTRACT

Subspace based methods for frequency estimation rely on a low-rank system model that is obtained by collecting the observed scalar valued data samples into vectors. Estimators such as MUSIC and ESPRIT have for some time been applied to this vector model. Also, a statistically attractive Markov-like procedure [1] for this class of methods has been proposed in the literature. Herein, the Markov estimator is re-investigated. Several results regarding rank, performance, and structure are given in a compact manner. The results are used to establish the large sample equivalence of the Markov estimator and the Approximate Maximum Likelihood (AML) algorithm proposed by Stoica et. al..

1. INTRODUCTION

Model based parameter estimation using subspace based methods can be an attractive alternative to maximum likelihood estimation. In order to apply subspace methods, a low rank signal model at hand must be available. In some cases, like in array signal processing, this structure is present directly in the received data vectors. In other cases, e.g., sinusoidal frequency estimation, subspace system identification and blind channel identification, the low rank vector valued data structure can be obtained by applying a temporal window to the received data. Vector valued data models obtained from an underlying scalar valued process are in this paper referred to as windowed data models.

Intuitively, the statistical properties of subspace methods when applied to windowed data models are different from models where the low rank structure is physically present in the system. In this paper, the statistical properties of subspace based estimators applied to windowed data models are examined using a subspace based sinusoidal frequency estimator as an example. The focus is thus *not* on obtaining a new estimator, but rather on gaining insight on the behavior of the studied class of methods. The estimator that is analyzed here is close to the algorithm presented in [1].

Here, we extend the analysis in [1] and show that, by carefully exploiting the structure, compact expressions for the estimation error covariance can in fact be obtained. In addition, these expressions enable further analysis of the rank properties of certain weighting and residual covariance matrices. These rank properties were left as an open question in [1], but they are in fact essential when determining optimal weighting matrices. The results presented here also make it possible to establish the large sample equivalence of the Markov estimator and the approximate maximum likelihood approach (AML) in [6]. The equivalence implies that the subspace approach provides the minimum asymptotic error covariance in the class of all estimators based on a given set of covariance estimates.

2. DATA MODEL AND DEFINITIONS

The N samples of the scalar-valued signal $y(t)$ are assumed to be the sum of d complex-valued sinusoids in additive zero-mean white Gaussian noise

$$x_k(t) = \alpha_k e^{\mathrm{i}(\omega_k t + \phi_k)}, \qquad k = 1, \ldots, d, \tag{1}$$

$$y(t) = \sum_{k=1}^{d} x_k(t) + n(t), \quad t = 1, \ldots, N. \tag{2}$$

Here, $\alpha_k > 0$ is the real-valued amplitude. The frequencies $\boldsymbol{\omega} = [\omega_1, \ldots, \omega_d]^T$ are assumed to be distinct deterministic parameters, and the phases $\{\phi_k\}$ are assumed to be uniformly distributed on $[0, 2\pi)$ and mutually independent. The noise, $n(t)$, is assumed to be independent of the phases and to satisfy

$$\mathrm{E}\{n(t)n^c(t-\tau)\} = \begin{cases} \sigma & \tau = 0, \\ 0 & \tau \neq 0, \end{cases} \tag{3}$$

$$\mathrm{E}\{n(t)n(t-\tau)\} = 0, \tag{4}$$

where $(\cdot)^c$ denotes complex conjugate. A low rank matrix representation of the problem is obtained by collecting $m > d$ received samples in a column vector

$$\mathbf{y}(t) = \begin{bmatrix} y(t) & y(t+1) & \ldots & y(t+m-1) \end{bmatrix}^T. \tag{5}$$

Here, $(\cdot)^T$ denotes the transpose and $(\cdot)^*$ will be used to denote the complex conjugate transpose. To establish the widely used matrix model for the vector valued system in (5) we introduce the notation

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) & x_2(t) & \ldots & x_d(t) \end{bmatrix}^T. \tag{6}$$

This results in the matrix formulation

$$\mathbf{y}(t) = \mathbf{A}_m(\boldsymbol{\omega})\mathbf{x}(t) + \mathbf{n}(t), \quad t = 1, \ldots, N-m+1, \tag{7}$$

where the additive noise vector, $\mathbf{n}(t)$, is defined similarly to $\mathbf{y}(t)$ in (5) and the $m \times d$ Vandermonde matrix $\mathbf{A}_m(\boldsymbol{\omega})$ is given by

$$\mathbf{A}_m(\boldsymbol{\omega}) = \begin{bmatrix} 1 & \ldots & 1 \\ e^{\mathrm{i}\omega_1} & \ldots & e^{\mathrm{i}\omega_d} \\ \vdots & & \vdots \\ e^{\mathrm{i}(m-1)\omega_1} & \ldots & e^{\mathrm{i}(m-1)\omega_d} \end{bmatrix}. \tag{8}$$

The argument $\boldsymbol{\omega}$ is omitted in the sequel when not required. The covariance matrix, $\mathbf{R}$, of the received windowed sequence is

$$\mathbf{R} = \mathrm{E}\{\mathbf{y}(t)\mathbf{y}^*(t)\} = \mathbf{A}_m\mathbf{S}\mathbf{A}_m^* + \sigma\mathbf{I}_m \tag{9}$$

where the covariance matrix $\mathbf{S}$ of $\mathbf{x}(t)$ is diagonal with the elements $\boldsymbol{\alpha} = [\alpha_1^2, \ldots, \alpha_d^2]^T$ on the main diagonal. The subscript on the identity matrix, $\mathbf{I}_m$, indicates the dimension of the matrix. The eigendecomposition of the covariance matrix is central in subspace based estimation methods and is given by

$$\mathbf{R} = \sum_{k=1}^{m} \lambda_k \mathbf{u}_k \mathbf{u}_k^* = \mathbf{U}_s \boldsymbol{\Lambda}_s \mathbf{U}_s^* + \mathbf{U}_n \boldsymbol{\Lambda}_n \mathbf{U}_n^* \tag{10}$$

where the eigenvalues are indexed in descending order and

$$\begin{aligned} &\mathbf{U}_s = \begin{bmatrix} \mathbf{u}_1 & \ldots & \mathbf{u}_d \end{bmatrix}, \quad \mathbf{U}_n = \begin{bmatrix} \mathbf{u}_{d+1} & \ldots & \mathbf{u}_m \end{bmatrix}, \\ &\boldsymbol{\Lambda}_s = \operatorname{diag}[\lambda_1 \ldots \lambda_d], \qquad \boldsymbol{\Lambda}_n = \sigma \mathbf{I}_{m-d}. \end{aligned} \tag{11}$$

The matrix $\mathbf{U}_s$ spans the same space as $\mathbf{A}_m$, which is often denoted the signal subspace. The matrix $\mathbf{U}_n$ spans the orthogonal complement, often referred to as the noise subspace. When working with subspace methods, it is in many cases advantageous to use a parameterization of the noise subspace instead of the signal subspace. In this problem, one noise subspace parameterization is given by the $m \times (m-d)$ matrix

$$\mathbf{G}_m(\mathbf{g}) = \begin{bmatrix} g_0 & \ldots & g_d & 0 & \ldots & 0 \\ 0 & g_0 & \ldots & g_d & \ddots & \vdots \\ \vdots & \ddots & & & & 0 \\ 0 & \ldots & 0 & g_0 & \ldots & g_d \end{bmatrix}^T, \tag{12}$$

where the parameters $\mathbf{g} = \begin{bmatrix} g_0, & \ldots, & g_d \end{bmatrix}^T$ are defined by

$$g_0 + g_1 z^1 + \ldots + g_d z^d = g_d \prod_{k=1}^{d} (z - e^{-\mathrm{i}\omega_k}). \tag{13}$$

That $\mathbf{G}_m^* \mathbf{A}_m = \mathbf{0}$ is easy to verify using (12) and (13). The mapping from $\boldsymbol{\omega}$ to $\mathbf{g}$ is unique up to a complex scalar multiplication. Since the roots of (13) lie on the unit circle, the polynomial can be written such that its coefficients satisfy the complex conjugate symmetry constraint $g_k = g_{d-k}^c$ for $k = 0, \ldots, d$.

3. SUBSPACE BASED ESTIMATOR

Subspace based estimators exploit the orthogonality between the noise and the signal subspaces. With a sample estimate, $\hat{\mathbf{U}}_s$, of $\mathbf{U}_s$, calculated from the collected data samples, the orthogonality can be expressed according to

$$\boldsymbol{\epsilon}(\boldsymbol{\omega}) = \operatorname{vec}[\hat{\mathbf{U}}_s^* \mathbf{G}_m(\boldsymbol{\omega})] \approx \mathbf{0}. \tag{14}$$

Here, vec[·] is the vectorization operator. The notation $\hat{\cdot}$ is used to denote estimated quantities. In the noiseless case, the matrices $\hat{\mathbf{U}}_s$ and $\mathbf{G}_m$ are exactly orthogonal, and setting the above residual to zero yields the true frequencies. When the estimate of $\mathbf{U}_s$ is not exact but computed from the received data samples, then the frequency estimates obtained by minimizing the norm of the residual vector are consistent. Note that, for simplicity, the noise subspace parameterization is considered to be a function of $\boldsymbol{\omega}$ rather than $\mathbf{g}$. The approach of [1] is to estimate $\boldsymbol{\omega}$ by minimizing a weighted norm of particular linear combinations of the real and imaginary parts of $\boldsymbol{\epsilon}(\boldsymbol{\omega})$ in (14). To describe this mathematically, we first define the real valued residual vector

$$\boldsymbol{\epsilon}_r(\boldsymbol{\omega}) = \begin{bmatrix} \operatorname{Re}\{\boldsymbol{\epsilon}(\boldsymbol{\omega})\} \\ \operatorname{Im}\{\boldsymbol{\epsilon}(\boldsymbol{\omega})\} \end{bmatrix} = \mathbf{L} \begin{bmatrix} \boldsymbol{\epsilon}(\boldsymbol{\omega}) \\ \boldsymbol{\epsilon}^c(\boldsymbol{\omega}) \end{bmatrix} \tag{15}$$

where $\mathbf{L}$ is a simple transformation matrix containing $\mathbf{I}$ and $\pm \mathrm{i}\,\mathbf{I}$. The investigated class of estimators can now be written

$$\hat{\boldsymbol{\omega}} = \arg\min_{\boldsymbol{\omega}} V_r(\boldsymbol{\omega}) \tag{16}$$

$$V_r(\boldsymbol{\omega}) = \boldsymbol{\epsilon}_r^T(\boldsymbol{\omega}) \mathbf{W}_r \boldsymbol{\epsilon}_r(\boldsymbol{\omega}), \tag{17}$$

where $\mathbf{W}_r \geq \mathbf{0}$ is a symmetric weighting matrix. The subscript r on a quantity indicates that it is real-valued. The method proposed in [1] is a member in the class described in this section.

4. PRELIMINARIES

In what follows we discuss different alternatives for how to estimate the covariance matrix $\mathbf{R}$ from which $\hat{\mathbf{U}}_s$ in (14) can be obtained. The most commonly used estimate of $\mathbf{R}$ is

$$\hat{\mathbf{R}} = \frac{1}{N-m+1} \sum_{k=1}^{N-m+1} \mathbf{y}(k) \mathbf{y}^*(k). \tag{18}$$

However, contrary to $\mathbf{R}$, the sample estimate $\hat{\mathbf{R}}$ is not Toeplitz. It turns out that from an analysis point of view it is beneficial to work with sample estimates that are Toeplitz. A sample covariance matrix that is Toeplitz can be obtained as follows. First, estimate the scalar-valued autocorrelation function by, e.g.,

$$\hat{r}(\tau) = \begin{cases} \frac{1}{N-\tau} \sum_{t=\tau+1}^{N} y(t) y^*(t-\tau) & \tau = 0, \ldots, m-1, \\ \hat{r}^*(-\tau) & \tau = -1, \ldots, -m+1, \end{cases} \tag{19}$$

then form the Toeplitz structured estimate

$$\hat{\mathbf{R}}_{\mathrm{T}} = \begin{bmatrix} \hat{r}(0) & \ldots & \hat{r}(-m+1) \\ \hat{r}(1) & \ldots & \hat{r}(-m+2) \\ \vdots & \vdots & \vdots \\ \hat{r}(m-1) & \ldots & \hat{r}(0) \end{bmatrix}. \tag{20}$$

The difference between the two sample covariance matrices introduced above is only due to "edge" effects. In particular, we have

$$\hat{\mathbf{R}} = \hat{\mathbf{R}}_{\mathrm{T}} + O(1/N) \tag{21}$$

where $O(1/N)$ is the statistical counterpart of the corresponding deterministic quantity. As is well known, the asymptotic covariance of the estimates is only dependent on $O(1/\sqrt{N})$ terms and the two sample covariance matrices in (21) thus yield estimates of the same accuracy when the number of samples is large. Observe that, in the finite sample case, $\hat{\mathbf{R}}$ may yield better estimates than $\hat{\mathbf{R}}_{\mathrm{T}}$, this despite they are asymptotically equivalent. Since the statistical analysis is simplified considerably if the Toeplitz structure can be used, we assume in the analysis that $\hat{\mathbf{R}}_{\mathrm{T}}$ is used in the estimator. However, note that the analysis is still valid for both $\hat{\mathbf{R}}_{\mathrm{T}}$ and $\hat{\mathbf{R}}$.

Toeplitz matrices are completely determined by their first row and column. These elements in the Toeplitz covariance matrix, $\mathbf{R}$, are for notational convenience collected in the vector

$$\mathbf{r} = \begin{bmatrix} r(-m+1) & r(-m+2) & \ldots & r(m-1) \end{bmatrix}^T. \tag{22}$$

The Toeplitz structure of $\mathbf{R}$ and the noise subspace parameterization matrix, $\mathbf{G}_m$, imply that also the product of these two matrices is Toeplitz with the (k,l)-element equal to

$$\delta_{k-l} = [\mathbf{R}\mathbf{G}_m]_{k,l} = \sum_{p=0}^{d} g_p r(k-l-p). \tag{23}$$

Thus, the product can be written

$$\mathbf{R}\mathbf{G}_m = \begin{bmatrix} \delta_0 & \delta_{-1} & \dots & \delta_{-(m-d-1)} \\ \delta_1 & \delta_0 & \dots & \delta_{-(m-d)} \\ \vdots & & & \vdots \\ \delta_{m-1} & & & \delta_d \end{bmatrix}. \tag{24}$$

Analogous to (22), the elements δ_l are collected in the vector

$$\boldsymbol{\delta} = \begin{bmatrix} \delta_{-(m-d-1)} & \delta_{-(m-d-2)} & \dots & \delta_{m-1} \end{bmatrix}^T. \tag{25}$$

The reason for collecting the elements in the Toeplitz matrices in vectors is that it facilitates tracking of single elements. This turns out to be most useful in the statistical analysis in the sequel. It follows from (23), (12), and the complex conjugate symmetry of g_k that a short hand formula for $\boldsymbol{\delta}$ is given by

$$\boldsymbol{\delta} = \mathbf{G}_{2m-1}^* \mathbf{r}. \tag{26}$$

Now, define the $m \times (2m-d-1)$ matrices

$$\mathbf{Q}_k = \begin{bmatrix} \mathbf{0} & \mathbf{I}_m & \mathbf{0}_{m\times k} \end{bmatrix}, \quad k = 0, 1, \dots, m-d-1. \tag{27}$$

Placed on top of each other, $\mathbf{Q}_k$ define the new matrix

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_0^T & \mathbf{Q}_1^T & \dots & \mathbf{Q}_{m-d-1}^T \end{bmatrix}^T. \tag{28}$$

It is now easily verified that

$$\mathbf{R}\mathbf{G}_m = \begin{bmatrix} \mathbf{Q}_0\boldsymbol{\delta} & \mathbf{Q}_1\boldsymbol{\delta} & \dots & \mathbf{Q}_{m-d-1}\boldsymbol{\delta} \end{bmatrix} \tag{29}$$

and, hence,

$$\mathrm{vec}[\mathbf{R}\mathbf{G}_m] = \mathbf{Q}\mathbf{G}_{2m-1}^* \mathbf{r}. \tag{30}$$

Equation (30) is used extensively in the statistical analysis of the estimator. Given a Toeplitz estimate of $\mathbf{R}$, e.g., $\hat{\mathbf{R}}_{\mathrm{T}}$, we can in an obvious fashion define sample versions of the above quantities.

5. STATISTICAL PROPERTIES OF THE RESIDUAL

To analyze the performance of the frequency estimators as in (16), it is necessary to determine the second order moments of the residual in the cost function. This has been accomplished previously in [1]. However, the complicated expression for the residual covariance matrix in that contribution obstructs the analysis. In this section, compact matrix expressions are derived for the covariance matrix of the residual and the rank properties of this matrix are established. The explicit determination of the rank and the null space of the residual covariance matrix were left as open questions in [1].

We obtain the large sample covariance of the residual by relating the statistical properties of the residual vector in (14) to the properties of the sample covariance matrix via

$$\begin{aligned} \boldsymbol{\epsilon} &= \mathrm{vec}[\hat{\mathbf{U}}_s^* \mathbf{G}_m] = \mathrm{vec}[\tilde{\boldsymbol{\Lambda}}_s^{-1} \mathbf{U}_s^* \hat{\mathbf{R}} \mathbf{G}_m] + O(1/N) \\ &= \mathrm{vec}[\tilde{\boldsymbol{\Lambda}}_s^{-1} \mathbf{U}_s^* \hat{\mathbf{R}}_{\mathrm{T}} \mathbf{G}_m] + O(1/N) \end{aligned} \tag{31}$$

where we have defined $\tilde{\boldsymbol{\Lambda}}_s = \boldsymbol{\Lambda}_s - \sigma\mathbf{I}$. For a proof of the second equality in (31) see, e.g., [1]. The third equality follows from the large sample equivalence of $\hat{\mathbf{R}}$ and $\hat{\mathbf{R}}_{\mathrm{T}}$. By making use of the formula $\mathrm{vec}[\mathbf{ABC}] = (\mathbf{C}^T \otimes \mathbf{A})\,\mathrm{vec}[\mathbf{B}]$ for any matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ of compatible dimensions we can rewrite (31) as

$$\begin{aligned} \boldsymbol{\epsilon} &= (\mathbf{I}_{m-d} \otimes \tilde{\boldsymbol{\Lambda}}_s^{-1} \mathbf{U}_s^*)\,\mathrm{vec}[\hat{\mathbf{R}}_{\mathrm{T}} \mathbf{G}_m] + O(1/N) \\ &= (\mathbf{I}_{m-d} \otimes \tilde{\boldsymbol{\Lambda}}_s^{-1} \mathbf{U}_s^*) \mathbf{Q} \mathbf{G}_{2m-1}^* \hat{\mathbf{r}} + O(1/N), \end{aligned} \tag{32}$$

where the sample counterpart to (30) has been used in the second step. This relation shows that it is only necessary to study the statistical properties of $\mathbf{G}_{2m-1}^* \hat{\mathbf{r}}$ in the sequel. Before studying these statistical properties we introduce the notation

$$\boldsymbol{\Psi} \triangleq (\mathbf{I}_{m-d} \otimes \tilde{\boldsymbol{\Lambda}}_s^{-1} \mathbf{U}_s^*) \mathbf{Q}. \tag{33}$$

With this definition the residual vector is compactly written

$$\boldsymbol{\epsilon} = \boldsymbol{\Psi} \mathbf{G}_{2m-1}^* \hat{\mathbf{r}} + O(1/N). \tag{34}$$

This formula separates the residual in a product with two factors. The first factor, $\boldsymbol{\Psi}$, describes the structure in the residual originating from the Toeplitz structured covariance estimate. The second factor, $\mathbf{G}_{2m-1}^* \hat{\mathbf{r}}$, contains the "statistical kernel" of the residual.

Theorem 1. *Let $\boldsymbol{\omega}_0$ denote the true frequencies and $\boldsymbol{\epsilon} = \boldsymbol{\epsilon}(\boldsymbol{\omega}_0)$ the corresponding residuals defined in (14). Then*

$$\begin{aligned} \boldsymbol{\Sigma} &= \lim_{N\to\infty} N\,\mathrm{E}\{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^*\} = \sigma^2 \boldsymbol{\Psi} \mathbf{G}_{2m-1}^* \mathbf{G}_{2m-1} \boldsymbol{\Psi}^* \\ \bar{\boldsymbol{\Sigma}} &= \lim_{N\to\infty} N\,\mathrm{E}\{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\} = \boldsymbol{\Sigma}(\mathbf{J} \otimes \mathbf{D}^*) \\ \boldsymbol{\Sigma}_r &= \lim_{N\to\infty} N\,\mathrm{E}\{\boldsymbol{\epsilon}_r \boldsymbol{\epsilon}_r^T\} \\ &= \mathbf{L} \begin{bmatrix} \mathbf{I} \\ (\mathbf{J} \otimes \mathbf{D}) \end{bmatrix} \boldsymbol{\Sigma} \begin{bmatrix} \mathbf{I} & (\mathbf{J} \otimes \mathbf{D}^*) \end{bmatrix} \mathbf{L}^* \\ &\triangleq \bar{\mathbf{L}} \boldsymbol{\Sigma} \bar{\mathbf{L}}^*. \end{aligned}$$

Here, $\mathbf{D} = \mathbf{U}_s^T \mathbf{J} \mathbf{U}_s$ is a unitary diagonal matrix and the $(m-d) \times (m-d)$ matrix $\mathbf{J}$ is the exchange matrix with ones along the anti-diagonal and zeros elsewhere.

Proof: See [3].

When using the results in Theorem 1 to determine optimal weighting matrices for the class of subspace estimators described in Section 3, the rank and range properties of the residual covariances are important. From Theorem 1 it is clear that $\boldsymbol{\Sigma}$ is rank deficient if and only if $\boldsymbol{\Psi}$ is not full row rank. This observation is the reason for writing the residual on the special form in (34). The following theorem, proved in [3], specifies the rank of $\boldsymbol{\Psi}$:

Theorem 2. *Consider the $d(m-d) \times (2m-d-1)$ matrix $\boldsymbol{\Psi}$ defined in (33). If the number of sinusoids is equal to one ($d = 1$), or if the columns in $\mathbf{A}_m(\boldsymbol{\omega})$ are orthogonal, then $\mathrm{rank}\{\boldsymbol{\Psi}\} = m-1$. Otherwise, when the number of sinusoids is strictly greater than one, ($d > 1$), and the columns in $\mathbf{A}_m(\boldsymbol{\omega})$ are not orthogonal, then $\mathrm{rank}\{\boldsymbol{\Psi}\} = 2m-d-2$.*

For the case of only one sinusoid or orthogonal columns in $\mathbf{A}_m(\boldsymbol{\omega})$, the result of Theorem 2 together with the observation that $\mathbf{G}_{2m-1}^* \mathbf{G}_{2m-1}$ is a full rank matrix yield that the dimension of the null space of $\boldsymbol{\Sigma}$ is: $\dim \mathcal{N}(\boldsymbol{\Sigma}) = (d-1)(m-d-1)$. When the sinusoids are more than one and the columns in $\mathbf{A}_m$ are not orthogonal we get $\dim \mathcal{N}(\boldsymbol{\Sigma}) = (d-2)(m-d-1)$. The conditions for when the residual covariance $\boldsymbol{\Sigma}$ is positive definite ($\dim \mathcal{N}(\boldsymbol{\Sigma}) = 0$) are summarized in Table 1.

d	Comment on rank
1	$\mathbf{\Sigma} > \mathbf{0}$ holds for all $\boldsymbol{\omega}$
2	$\mathbf{\Sigma} > \mathbf{0}$ if either the sinusoids are "non-orthogonal" or $m = d + 1$.
> 2	$\mathbf{\Sigma} > \mathbf{0}$ only if $m = d + 1$.

Table 1: Conditions for when $\mathbf{\Sigma}$ is positive definite.

6. STATISTICAL ANALYSIS

In general, we assume that the weighting matrix $\mathbf{W}_r$ may depend on the parameters as well as on data. However, in the analysis presented below, we consider $\mathbf{W}_r$ to be a constant parameter independent matrix. Since $\boldsymbol{\epsilon}_r = O(1/\sqrt{N})$, this is valid asymptotically. The statistical analysis is separated in two parts. The first part solely treats the subspace estimator whereas the second part establishes the large sample equivalence to AML in [6]. From the theory presented in, e.g., Appendix C4.4 [5], it follows that the large sample covariance for the subspace based estimate is

$$\lim_{N\to\infty} N\,\mathrm{E}\{(\boldsymbol{\omega}-\boldsymbol{\omega}_0)(\boldsymbol{\omega}-\boldsymbol{\omega}_0)^T\} = (\mathbf{\Phi}_r^T\mathbf{W}_r\mathbf{\Phi}_r)^{-1}\mathbf{\Phi}_r^T\mathbf{W}_r\mathbf{\Sigma}_r\mathbf{W}_r^T\mathbf{\Phi}_r(\mathbf{\Phi}_r^T\mathbf{W}_r\mathbf{\Phi}_r)^{-1}. \quad (35)$$

Here, $\mathbf{\Phi}_r$ is the limiting Jacobian of the real-valued residual vector evaluated at the true frequencies $\boldsymbol{\omega}_0$,

$$\mathbf{\Phi}_r = \lim_{N\to\infty} \left.\frac{\partial \boldsymbol{\epsilon}_r}{\partial \boldsymbol{\omega}}\right|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0}. \quad (36)$$

Next, we show how to choose the weighting matrix $\mathbf{W}_r$ so that the estimation error covariance in (35) is minimized. It is well known that if $\mathbf{\Sigma}_r$ is non-singular then an optimal weighting matrix is given by $\mathbf{W}_r = \mathbf{\Sigma}_r^{-1}$. However, here $\mathbf{\Sigma}_r$ is singular and another approach is necessary. It is shown in [3] that $\mathrm{span}\{\mathbf{\Phi}_r\} \subset \mathrm{span}\{\mathbf{\Sigma}_r\}$ holds. Here, $\mathrm{span}(\cdot)$ denotes the range space of the corresponding matrix. When this relation holds it follows from the theory for weighting with pseudo-inverses [4] that $\mathbf{W}_r = \mathbf{\Sigma}_r^\dagger$ minimizes the estimation error covariance. Here, $(\cdot)^\dagger$ denotes the Moore-Penrose pseudo inverse; see, e.g., [4].

Theorem 3. *Assume that the frequency estimate $\hat{\boldsymbol{\omega}}$ is given by (16) with the weighting matrix given by $\mathbf{W}_r = \mathbf{\Sigma}_r^\dagger = \mathbf{L}\mathbf{\Sigma}^\dagger\mathbf{L}^*$. Then $\hat{\boldsymbol{\omega}}$ is the asymptotically best consistent estimate within the class of subspace based estimators and in large samples it converges to a distribution with the covariance*

$$\lim_{N\to\infty} N\,\mathrm{E}\{(\hat{\boldsymbol{\omega}}-\boldsymbol{\omega}_0)(\hat{\boldsymbol{\omega}}-\boldsymbol{\omega}_0)^T\} = (\mathbf{\Phi}_r^T\mathbf{\Sigma}_r^\dagger\mathbf{\Phi}_r)^{-1}$$

In [3] it is shown that the performance of the optimally weighted real-valued estimator can also be obtained without separating the residual in its real and imaginary parts.

The optimal subspace based estimator is now shown to be equivalent to the asymptotic maximum likelihood frequency estimator (AML) proposed in [6]. The AML algorithm is in principle a weighted least squares fit of the estimated covariance and the parameterized version thereof in (22). In [3] it is proved that:

Theorem 4. *Assume that the sinusoidal frequencies are such that not all the columns in $\mathbf{A}_m$ are orthogonal, then the large sample covariances of the optimally weighted subspace estimate and the AML estimate in [6] are equal. This establishes that the subspace approach provides the minimum asymptotic error covariance in the class of all estimators based on a given set of covariance estimates.*

Observe that the theorem is derived under the assumption that the sinusoids are non-orthogonal. Surprisingly, the result in the theorem is not valid when the sinusoids are orthogonal. Numerical investigations indicate that the subspace based estimator is in the orthogonal case equivalent to ESPRIT and thus suboptimal! For non-orthogonal sinusoids the equivalence of AML and the subspace based estimator is perhaps somehow surprising since AML explicitly exploits the diagonal structure of $\mathbf{S}$. Since $\hat{\mathbf{U}}_s$ is computed without the constraint that $\mathbf{S}$ is diagonal ($\mathbf{R}$ is Toeplitz), this is not obviously the case for the subspace method discussed in this paper. In comparison, neglecting the exploitation of a diagonal signal covariance matrix (uncorrelated sources) in direction estimation results in suboptimal performance [2].

7. CONCLUSIONS

Herein, a Markov-like subspace based procedure for sinusoidal frequency estimation has been re-investigated. Compact formulas for the covariance matrix of the residual in the criterion function have been derived. These expressions facilitated an analysis of the estimator and with certain rank considerations optimality claims were established. In addition, the large sample equivalence to AML in [6] was established using these expressions.

The rank investigations show that when the number of sinusoids is one or two, then the residual covariance matrices are in most cases full rank and the optimal weighting matrices can be computed using standard matrix inversion. However, when the number of sinusoids is strictly greater than two, then these matrices are always rank deficient. In addition, the dimension of the nullspace of the residual covariance matrix depends on the sinusoidal frequencies to be estimated. This complicates the computation of the optimal weighting matrices. Some comments on implementational aspects can be found in [3].

8. REFERENCES

[1] A. Eriksson, P. Stoica, and T. Söderström. Markov-based eigenanalysis method for frequency estimation. *IEEE Transactions on Signal Processing*, 42(3):586–594, March 1994.

[2] M. Jansson, B. Göransson, and B. Ottersten. A subspace method for direction of arrival estimation of uncorrelated emitter signals. To appear in IEEE Transactions on Signal Processing.

[3] M. Kristensson, M. Jansson, and B. Ottersten. Further results and insights on subspace based sinusoidal frequency estimation. Submitted to IEEE Transactions on Signal Processing.

[4] C. R. Rao and S. K. Mitra. *Generalized Inverse of Matrices and its Applications*. John Wiley & Sons, first edition, 1971.

[5] T. Söderström and P. Stoica. *System Identification*. Prentice Hall International, 1989.

[6] P. Stoica, P. Händel, and T. Söderström. Approximate maximum-likelihood frequency estimation. *Automatica*, 30(1):131–145, January 1994.