Subspace Fitting Approaches for Frequency Estimation Using Real-Valued Data

Kaushik Mahata

Abstract-A novel data covariance model has recently been proposed for the subspace-based estimation of multiple real-valued sine wave frequencies. In this paper, we develop weighted subspace fitting approaches using this new data model. A new parameterization of the noise subspace is proposed. This parameterization is used to solve the subspace fitting problem analytically. An expression for the residual covariance matrix is derived. This covariance matrix is further used to obtain an optimally weighted Gauss-Markov estimator. A computationally efficient suboptimal weighting is also proposed, and the associated estimator is close to the Gauss-Markov estimator in performance. The suboptimal weighting strategy is quite general and can be used in other related applications. The performance of the algorithms are illustrated using numerical simulations. The proposed subspace fitting approach shows improved resolution performance. It is also robust to additive noise.

Index Terms—Frequency estimation, real-valued data, spectral analysis, subspace methods, weighted subspace fitting.

I. INTRODUCTION

T HE problem of estimating the spectral lines from a sequence of real-valued data samples arises naturally in many practical applications like nuclear magnetic resonance spectroscopy, passive sonar, ultrasound imaging, and so forth. When subspace methods like MUSIC [1] or ESPRIT [2] are employed to estimate real-valued sinusoidal frequencies, it is generally pretended that the data consist of a double number of complex-valued sinusoidal signals. However, the resolution performance of the conventional subspace algorithms (designed primarily for complex-valued sine waves) when employed to real-valued data are not as good as the case when they are employed to complex-valued data. Recently, an alternative data model (RDM) has been proposed for subspace-based estimation of real-valued sine wave frequencies [3]. The primary advantage of using RDM over the conventional complex-valued covariance matrix model is that the dimension of the signal subspace in RDM equals the number of sinusoids present in the data. This is half of the signal subspace dimension for the conventional complex-valued data model. As a result, the frequency estimates derived using RDM perform better for small dimensions of the covariance matrix model. This advantage has

The author is with the Center for Complex Dynamic Systems and Control, School of Electrical Engineering, University of Newcastle, Callaghan, NSW-2308, Australia (e-mail: kush@ee.newcastle.edu.au).

Digital Object Identifier 10.1109/TSP.2005.851129

been exploited in [3], where an ESPRIT like algorithm, named R-Esprit, was proposed. R-Esprit is equivalent to forward-back-ward ESPRIT [4] in performance but is computationally more efficient.

All subspace-based frequency estimation methods (including R-Esprit) exhibit excellent resolution performance when the signal-to-noise ratio (SNR) is high. In fact, the asymptotic (when the number of data samples is large) covariance matrix of the frequency estimates can be shown to be inversely proportional to the square of the SNR [4], [5]. However, the performance degrades significantly with decrease in the SNR. This phenomenon is even more prominent when the data are real-valued. We attempt to remedy this problem in this paper, where our prime objective is to combine the nice properties of the well-known subspace fitting approaches [4], [6], [7] with the advantages associated with RDM to develop noise tolerant estimators with improved resolution capability at low SNR. For the algorithms where the data are modeled as the superposition of complex-valued sine waves, a computationally efficient weighted subspace fitting (WSF) approach has been designed in [4]. An expression for the optimal weighting matrix was also derived in [4]. The same WSF problem was again considered in [6], where the results in [4] were refined, and further perspectives were given. Here, we follow the ideas presented in [6] to develop a computationally efficient WSF estimator for the real-valued data case. Consider

$$y_0(t) = \sum_{k=1}^d a_k \cos\{\omega_k t + \phi_k\}, \quad y(t) = y_0(t) + \tilde{y}(t) \quad (1)$$

where $\tilde{y}(t)$ is a zero mean white measurement noise of variance σ^2 , and ϕ_k is uniformly distributed on $[0, 2\pi)$ for all $k, 1 \leq k \leq d$. Without any loss of generality, we assume $0 < \omega_1 < \cdots < \omega_d < \pi$. The amplitudes $\{a_k\}_{k=1}^d$ and the noise variance σ^2 are assumed to be *unknown*. It is further assumed that the random variables $\{\phi_k\}_{k=1}^d$ are pairwise independent, as well as independent of $\tilde{y}(t)$. The central problem under consideration of this paper is to estimate the parameter vector

$$\boldsymbol{\omega}_0 := \begin{bmatrix} \omega_1 & \cdots & \omega_d \end{bmatrix} \tag{2}$$

from the observed data $\{y(t)\}_{t=1}^N$.

In the next section, we give a brief description of RDM. In Section III, we develop the subspace fitting algorithm and present its analysis. A comparative performance study of the different algorithms will be presented in Section IV, followed by conclusions in Section V.

Manuscript received March 5, 2004; revised October 12, 2004. This work was supported in part by Swedish Research Council for Engineering Sciences under Contract 2000-587 and by the Australian Research Council. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Petar M. Djuric.

II. COVARIANCE MATRIX MODEL

In this section, we revisit RDM: the compact covariance matrix model introduced in [3]. By the word *compact*, we mean that the noise-free part of the snapshot vector lies in a subspace of dimension d. Recall that the dimension of the signal subspace using the conventional complex-valued data model is 2d [8], [9]. The reason why it is beneficial to consider such an alternative model is the associated improvement in the estimation accuracy for small values of the dimension m of the snapshot vector. In order to spell out the details, let us introduce

$$\mathbf{y}_c(t) := \begin{bmatrix} y(t) & \cdots & y(t+m-1) \end{bmatrix}_{-}^{\top}$$
(3)

$$\mathbf{y}_b(t) := \begin{bmatrix} y(t-1) & \cdots & y(t-m) \end{bmatrix}^\top \tag{4}$$

$$\mathbf{y}(t) := \frac{1}{2} \{ \mathbf{y}_c(t) + \mathbf{y}_b(t) \}$$
(5)

where \top denotes the transpose and m > d. From the above definitions, it is easy to derive using (1) that the *j*th element of $\mathbf{y}(t)$ is given by

$$[\mathbf{y}(t)]_{j} = \frac{1}{2} \{ \tilde{y}(t+j-1) + \tilde{y}(t-j) \} + \sum_{k=1}^{d} a_{k} \cos\{\omega_{k}(t-1/2) + \phi_{k}\} \cos\{\omega_{k}(j-1/2)\}.$$
 (6)

Thus, from (6), we have

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\omega}_0)\mathbf{s}(t) + \tilde{\mathbf{y}}(t) \tag{7}$$

where $A(\omega_0)$ is an $m \times d$ full column rank [3] matrix given by

$$\mathbf{A}(\boldsymbol{\omega}_0) = \begin{bmatrix} \cos(\frac{\omega_1}{2}) & \cdots & \cos(\frac{\omega_d}{2}) \\ \cos(\frac{3\omega_1}{2}) & \cdots & \cos(\frac{3\omega_d}{2}) \\ \vdots & \ddots & \vdots \\ \cos\{(m-\frac{1}{2})\omega_1\} & \cdots & \cos\{(m-\frac{1}{2})\omega_d\} \end{bmatrix}.$$
(8)

 $\mathbf{s}(t)$ is a $d \times 1$ vector given by

$$\mathbf{s}(t) = \begin{bmatrix} a_1 \cos\{\omega_1 t + \phi_1^+\} \\ \vdots \\ a_d \cos\{\omega_d t + \phi_d^+\} \end{bmatrix}$$
(9)

and $\phi_k^+ = \phi_k - (1/2)\omega_k$ for $1 \le k \le d$. The noise snapshot vector $\tilde{\mathbf{y}}(t)$ is given by

$$\tilde{\mathbf{y}}(t) = \frac{1}{2} \{ \tilde{\mathbf{y}}_c(t) + \tilde{\mathbf{y}}_b(t) \}$$
(10)

where $\tilde{\mathbf{y}}_c(t)$ and $\tilde{\mathbf{y}}_b(t)$ are noise contributions to $\mathbf{y}_c(t)$ and $\mathbf{y}_b(t)$, respectively:

$$\tilde{\mathbf{y}}_c(t) = [\tilde{y}(t) \quad \cdots \quad \tilde{y}(t+m-1)]^{\top}$$
 (11)

$$\tilde{\mathbf{y}}_b(t) = [\tilde{y}(t-1) \quad \cdots \quad \tilde{y}(t-m)]^{\top}.$$
 (12)

Therefore, in contrast to the complex-valued data model, the noise-free part of y(t) lies in a d dimensional subspace. Hence,

we have now a more compact model, which is real-valued. In-troduce

$$\mathbf{P} := \mathbf{E}\{\mathbf{s}(t)\mathbf{s}^{\top}(t)\}$$
(13)

and note that $\tilde{\mathbf{y}}_c(t)$ and $\tilde{\mathbf{y}}_b(t)$ are mutually independent random vectors for all t, which would mean $\mathbf{E}{\{\tilde{\mathbf{y}}(t)\tilde{\mathbf{y}}^{\top}(t)\}} = (\sigma^2/2)\mathbf{I}_m$. It follows readily that

$$\mathbf{R} := \mathbf{E}\{\mathbf{y}(t)\mathbf{y}^{\top}(t)\} = \mathbf{A}(\boldsymbol{\omega}_0)\mathbf{P}\mathbf{A}^{\top}(\boldsymbol{\omega}_0) + \frac{\sigma^2}{2}\mathbf{I}_m.$$
 (14)

One can estimate **R** from the data. This estimate can be processed in several different ways to arrive at the estimates of the frequencies. In contrast to the conventional complex-valued covariance matrix model for complex-valued sine waves, in this case, the matrices involved are all real-valued. On the other hand, $\mathbf{A}(\omega_0)$ is no longer a Vandermonde matrix, which is the case for the conventional model. However, the structures present in $\mathbf{A}(\omega_0)$ can also be exploited to derive algorithms of low computational complexity, although the associated calculations are bit more involved as compared to the complex-valued case. Consider the eigenvalue decomposition

$$\mathbf{R} = \mathbf{S} \mathbf{\Lambda} \mathbf{S}^{\top} + \mathbf{G} \mathbf{\Sigma} \mathbf{G}^{\top} \tag{15}$$

where Λ is a $d \times d$ diagonal matrix having the d dominating eigenvalues of \mathbf{R} as its diagonal entries and the $m \times d$ matrix \mathbf{S} is composed of the corresponding eigenvectors. Similarly, $\Sigma = \sigma^2 \mathbf{I}_{m-d}$ is a diagonal matrix having the remaining m - d eigenvalues of \mathbf{R} as its diagonal entries. Columns of the $m \times (m - d)$ matrix \mathbf{G} constitute the corresponding eigenvectors. Note that the columns of \mathbf{S} constitute an orthonormal basis of the subspace spanned by the columns of $\mathbf{A}(\boldsymbol{\omega}_0)$. After a few steps of calculations [5], we can show using (14) and (15) that

$$\mathbf{S} = \mathbf{A}(\boldsymbol{\omega}_0)\mathbf{C} \tag{16}$$

where

$$\mathbf{C} = \mathbf{P}\mathbf{A}^{\top}(\boldsymbol{\omega}_0)\mathbf{S}\mathbf{D}^{-1} \tag{17}$$

$$\mathbf{D} = \mathbf{\Lambda} - \frac{\sigma^2}{2} \mathbf{I}_d.$$
 (18)

In other words, S contains necessary information about the frequencies, which is the primary motivation behind all the subspace approaches.

From the construction of the snapshot vector $\mathbf{y}(t)$ in (5), we see that the number of snapshot vectors available is N - 2m +1. On the other hand, using the conventional complex-valued data model, the number of snapshot vectors is N - m + 1. Hence, the *effective* data length is reduced for RDM. However, from (14), we can see that using RDM, the noise contribution to the covariance matrix \mathbf{R} of the snapshot vector $\mathbf{y}(t)$ is also reduced by factor of two. Therefore, if m is small compared to N, the performance loss due to the reduced effective data length is insignificant compared with the gain due to enhanced snapshot SNR. This is the main reason behind the significant performance improvement of R-Esprit using smaller values of m. For larger m, degradation in performance is expected (which is not a problem from a practical point of view). If the number of data samples N is also small, the reduction in the effective data length, however, is an important issue. This problem is remedied by using an alternative covariance estimate [see Section III-C for details] in this work.

III. WEIGHTED SUBSPACE FITTING

A. Background

In the rest of the paper, we use $\boldsymbol{\omega}$ as the dummy parameter vector, which will be given element-wise as

$$\boldsymbol{\omega} = [\boldsymbol{\varpi}_1 \quad \cdots \quad \boldsymbol{\varpi}_d].$$

We assume that the parameter vector $\boldsymbol{\omega} \in \mathcal{D}$, where we define the set $\mathcal{D} \subset \mathbb{R}^d$ such that

$$\mathcal{D} := \{ \boldsymbol{\omega} \, | \, 0 < \varpi_1 < \varpi_2 < \dots < \varpi_d < \pi \}.$$
(19)

Clearly, the true parameter vector $\boldsymbol{\omega}_0 \in \mathcal{D}$. For any value of the parameter vector $\boldsymbol{\omega} \in \mathcal{D}$, one can construct a full column rank matrix $\mathcal{G}(\boldsymbol{\omega})$ of dimension $m \times (m - d)$ such that

$$\mathcal{G}^{\top}(\boldsymbol{\omega})\mathbf{A}(\boldsymbol{\omega}) = \mathbf{0}_{(m-d)\times d}$$
(20)

is satisfied, i.e., the nullspace of $\mathbf{A}^{\top}(\boldsymbol{\omega})$ is spanned by the columns of $\mathcal{G}(\boldsymbol{\omega})$. Consider a *strongly consistent* estimate $\hat{\mathbf{R}}_N$ of \mathbf{R} , and let $\hat{\mathbf{S}}_N$ denote the $m \times d$ matrix composed of the d dominating eigenvectors of $\hat{\mathbf{R}}_N$. Introduce

$$\boldsymbol{\epsilon}(\boldsymbol{\omega}) = \operatorname{vec}[\mathcal{G}^{\top}(\boldsymbol{\omega})\mathbf{S}], \quad \hat{\boldsymbol{\epsilon}}_{N}(\boldsymbol{\omega}) = \operatorname{vec}[\mathcal{G}^{\top}(\boldsymbol{\omega})\hat{\mathbf{S}}_{N}]$$
(21)

where we denote the vectorization operation by vec[·]. The *residual* vector $\hat{\boldsymbol{\epsilon}}_N(\boldsymbol{\omega})$ is an estimate of $\boldsymbol{\epsilon}(\boldsymbol{\omega})$ obtained by replacing **S** by $\hat{\mathbf{S}}_N$ in the first equality of (21). Since the columns of **S** form an orthonormal basis of the subspace spanned by the columns of $\mathbf{A}(\boldsymbol{\omega}_0)$ [see (15) and (16)], we have

$$\boldsymbol{\epsilon}(\boldsymbol{\omega}_0) = \boldsymbol{0}_{d(m-d) \times 1}.$$
(22)

Therefore, given $\mathcal{G}(\boldsymbol{\omega})$, an estimate $\hat{\boldsymbol{\omega}}_N(\mathbf{W})$ of the parameter vector $\boldsymbol{\omega}_0$ can be obtained by solving the optimization problem

$$\hat{\boldsymbol{\omega}}_N(\mathbf{W}) = \arg\min_{\boldsymbol{\omega}\in\mathcal{D}} \hat{\boldsymbol{\epsilon}}_N^{\mathsf{T}}(\boldsymbol{\omega}) \mathbf{W} \hat{\boldsymbol{\epsilon}}_N(\boldsymbol{\omega})$$
(23)

where **W** is a user-defined positive definite weighting matrix. In the signal processing literature, a similar approach is quite common and usually referred to as the weighted subspace fitting algorithm [7], [10]. In the following lemma, we study the consistency property of $\hat{\omega}_N(\mathbf{W})$.

Lemma 1: Assume that $\hat{\mathbf{R}}_N \to \mathbf{R}$ with probability one as $N \to \infty$ and that \mathbf{W} is a positive definite weighting matrix. Then,

$$\lim_{N\to\infty}\hat{\boldsymbol{\omega}}_N(\mathbf{W}) = \boldsymbol{\omega}_0$$

with probability one.

Proof: See Appendix A.

Although the specific choice of the weighting matrix \mathbf{W} does not influence the consistency of the frequency estimates, it has a significant influence on the asymptotic accuracy of the parameter estimates. Assume that the necessary conditions needed to ensure consistency are satisfied. Let the Jacobian matrix of first-order derivatives of $\epsilon(\omega)$ evaluated at $\omega = \omega_0$ be given by

$$\mathcal{J} := \begin{bmatrix} \frac{\partial \boldsymbol{\epsilon}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}_1} & \cdots & \frac{\partial \boldsymbol{\epsilon}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}_d} \end{bmatrix} \Big|_{\boldsymbol{\omega} = \boldsymbol{\omega}_0}.$$
 (24)

Then, the asymptotic (as $N \to \infty$) covariance matrix of the estimate $\hat{\boldsymbol{\omega}}_N(\mathbf{W})$ is given by [6], [10]

$$\operatorname{cov}\{\sqrt{N}\hat{\omega}_{N}(\mathbf{W})\} = \{\mathcal{J}^{\top}\mathbf{W}\mathcal{J}\}^{-1} \times \{\mathcal{J}^{\top}\mathbf{W}\Delta\mathbf{W}\mathcal{J}\}\{\mathcal{J}^{\top}\mathbf{W}\mathcal{J}\}^{-1} \quad (25)$$

where Δ is the asymptotic covariance matrix of the residual vector:

$$\boldsymbol{\Delta} = \operatorname{cov}\{\sqrt{N}\hat{\boldsymbol{\epsilon}}_N(\boldsymbol{\omega}_0)\}.$$
(26)

The Gauss–Markov estimate $\hat{\boldsymbol{\omega}}_N(\mathbf{W}^*)$ has the minimum covariance matrix among the class of WSF estimators in the sense that $\operatorname{cov}\{\hat{\boldsymbol{\omega}}_N(\mathbf{W})\} - \operatorname{cov}\{\hat{\boldsymbol{\omega}}_N(\mathbf{W}^*)\}$ is a positive semidefinite matrix for all positive semidefinite \mathbf{W} . The corresponding *optimal* choice \mathbf{W}^* of \mathbf{W} is given by

$$\mathbf{W}^* = \boldsymbol{\Delta}^{-1},\tag{27}$$

so that the optimal covariance matrix is given by

$$\operatorname{cov}\{\sqrt{N}\hat{\boldsymbol{\omega}}_{N}(\mathbf{W}^{*})\} = \{\mathcal{J}^{\top}\boldsymbol{\Delta}^{-1}\mathcal{J}\}^{-1}.$$
 (28)

Using \mathbf{W}^* as the weighting matrix can be interpreted as prewhitening of the residual vector; see (23), (26), and (27). In Section III-D, we derive an explicit expression for Δ . Another important issue in a WSF approach is the associated computational complexity. For an arbitrary choice of $\mathcal{G}(\boldsymbol{\omega})$, the approach in (23) is computationally demanding in general, since the loss function in (23) must be optimized numerically. In the next section, we present a specific choice of $\mathcal{G}(\boldsymbol{\omega})$ for which the optimization problem can be solved analytically.

B. Noise Subspace Parameterization

In this section, our objective is to derive a new noise subspace parameterization for RDM so that the optimization problem in (23) can be solved analytically in a noniterative manner. Using similar recipes given in [4] and [11], we show that the problem in (23) can be reduced to a WLS problem, which can be solved readily using standard least squares theory. We have the following lemma which gives a specific noise subspace parameterization.

Lemma 2: Let $\omega \in \mathcal{D}$. Define the symmetric polynomial

$$G(z, \boldsymbol{\omega}) := \sum_{k=-d}^{d} g_k(\boldsymbol{\omega}) z^k := \prod_{k=1}^{d} \{ z^{-1} - 2\cos(\varpi_k) + z \}.$$
(29)

In addition, let $g_k(\boldsymbol{\omega}) = 0$ for |k| > d. Introduce the matrix $\mathbf{G}_0(\boldsymbol{\omega})$ of dimension $m \times (m-d)$, which is given element-wise as

$$[\mathbf{G}_0(\boldsymbol{\omega})]_{jk} = g_{j-k}(\boldsymbol{\omega}) + g_{j+k-1}(\boldsymbol{\omega}). \tag{30}$$

Then, $G_0(\boldsymbol{\omega})$ has a full column rank, and furthermore

$$\mathbf{G}_0^{\top}(\boldsymbol{\omega})\mathbf{A}(\boldsymbol{\omega}) = \mathbf{0}_{(m-d)\times d}.$$
 (31)

Proof: See Appendix B.

Lemma 2 provides us a way to parameterize the nullspace of $\mathbf{A}^{\top}(\boldsymbol{\omega})$, which will help us to simplify the optimization problem in (23). First, note that

$$g_d(\boldsymbol{\omega}) = g_{-d}(\boldsymbol{\omega}) = 1, \quad g_\tau(\boldsymbol{\omega}) = g_{-\tau}(\boldsymbol{\omega})$$
 (32)

for all τ . Hence, the polynomial $G(z, \boldsymbol{\omega})$ is completely specified by the vector

$$\mathbf{g}(\boldsymbol{\omega}) := [g_0(\boldsymbol{\omega}) \quad \cdots \quad g_{d-1}(\boldsymbol{\omega})]^\top.$$
 (33)

Since the symmetric polynomial $G(z, \omega_0)$ can be factorized uniquely in d pairs of mutually reciprocal roots lying on the unit circle, the parameter vector ω_0 can be recovered from $g(\omega_0)$ through polynomial rooting. The idea here is to use $G_0(\omega)$ for $\mathcal{G}(\omega)$ in (21) and solve the optimization problem (23) in $g(\omega)$ instead of ω . Subsequently, the resulting estimate of $g(\omega_0)$ is used to construct the estimate of the polynomial $G(z, \omega_0)$. Then, an estimate of ω_0 is obtained via polynomial rooting. In what follows next, we show that the loss function in (23) is quadratic in $g(\omega)$. To that aim, let us define the sequence of $m \times (m - d)$ matrices $\{\mathbf{K}_j\}_{j=0}^d$, which are given element-wise as

$$[\mathbf{K}_j]_{lk} = \delta_{|l-k|,j} + \delta_{l+k-1,j}$$
(34)

where $\delta_{k,l}$ is the Kronecker delta function. Then, it is readily verified from (30) and (34) that

$$\mathbf{G}_0(\boldsymbol{\omega}) = \sum_{j=0}^d \mathbf{K}_j g_j(\boldsymbol{\omega}). \tag{35}$$

Now, let us define the $d(m - d) \times 1$ vector \mathbf{u}_N and the $d(m - d) \times d$ matrix \mathbf{U}_N such that

$$\mathbf{u}_{N} = \operatorname{vec}\{\mathbf{K}_{d}^{\top}\hat{\mathbf{S}}_{N}\}$$
$$\mathbf{U}_{N} = [\operatorname{vec}\{\mathbf{K}_{0}^{\top}\hat{\mathbf{S}}_{N}\} \quad \cdots \quad \operatorname{vec}\{\mathbf{K}_{d-1}^{\top}\hat{\mathbf{S}}_{N}\}]. \quad (36)$$

Recall that $g_d(\boldsymbol{\omega}) = 1$ for all $\boldsymbol{\omega}$. Substituting $\mathcal{G}(\boldsymbol{\omega}) = \mathbf{G}_0(\boldsymbol{\omega})$ in (21) and using (35) and (36), we get

$$\hat{\boldsymbol{\epsilon}}_N(\boldsymbol{\omega}) = \mathbf{u}_N + \mathbf{U}_N \mathbf{g}(\boldsymbol{\omega}).$$
 (37)

Note that the vector \mathbf{u}_N and the matrix \mathbf{U}_N are functions of data only. Hence, the $\boldsymbol{\omega}$ -dependence of $\hat{\boldsymbol{\epsilon}}_N(\boldsymbol{\omega})$ is linear in the sense that $\hat{\boldsymbol{\epsilon}}_N(\boldsymbol{\omega})$ is a linear function of $\mathbf{g}(\boldsymbol{\omega})$. This observation leads us to a simple way to solve the optimization problem in (23), where the minimization is first carried out with respect to $\mathbf{g}(\boldsymbol{\omega})$ to obtain an estimate $\hat{\mathbf{g}}_N(\mathbf{W})$ of $\mathbf{g}(\boldsymbol{\omega}_0)$:

$$\hat{\mathbf{g}}_N(\mathbf{W}) = -[\mathbf{U}_N^\top \mathbf{W} \mathbf{U}_N]^{-1} \mathbf{U}_N^\top \mathbf{W} \mathbf{u}_N.$$
(38)

Using $\hat{\mathbf{g}}_N(\mathbf{W})$, we can form as estimate $\hat{G}(z)$ of $G(z, \omega_0)$; see (29) and (33). Subsequently, the estimate $\hat{\boldsymbol{\omega}}_N(\mathbf{W})$ is obtained by rooting $\hat{G}(z)$. Notice that the roots of the estimated polynomial

 $\hat{G}(z)$ will not lie on the unit circle in general, but they allways occur in complex conjugate pairs. Therefore, the arguments of complex-valued roots of $\hat{G}(z)$ give the frequency estimates.

C. Covariance Estimates

From (14), it is straightforward to see that \mathbf{R} can be estimated from the data as

$$\hat{\mathbf{R}}_{c} = \frac{1}{M} \sum_{t=m+1}^{N-m+1} \mathbf{y}(t) \mathbf{y}^{\top}(t)$$
(39)

where M = N - 2m + 1. It is well-known that $\hat{\mathbf{R}}_c$ is a strongly consistent estimator of R [12]. In that sense, the first requirement of Lemma 1 is satisfied if we use \mathbf{R}_c for \mathbf{R}_N . Note in (39) for relatively small values of N, the number of snapshots M available for the computation of \mathbf{R} can be quite small. For example, N = 25, and m = 10 implies M = 6 only. This can result in a poor estimate \mathbf{R}_c , which in fact is the principal disadvantage of R-Esprit [3]. However, it is well-known in the literature that the estimation accuracy can be improved significantly when the special structures present in the covariance matrix are considered in the estimation algorithm. In [4], forward-backward symmetry of the covariance matrix was exploited to improve the estimation accuracy significantly. This idea was extended in [6], where the Toeplitz structure in the covariance matrix was considered. Our objective here is to use the same notion in the problem under consideration. However, the structure here is more complicated. To proceed further, we need to define the autocorrelation sequence of the observed data y(t):

$$r_{\tau} = \mathbf{E}\{y(t+\tau)y(t)\}.$$
(40)

Now, from the definition (14) and (3)–(5), it is easy to verify that a generic element of \mathbf{R} is given as

$$[\mathbf{R}]_{kl} = \frac{1}{4} \{ r_{k-l} + r_{l-k} + r_{l+k-1} + r_{1-k-l} \}.$$
 (41)

Following the treatment given in [6], we see that the complete structure in \mathbf{R} can be exploited if we use an alternative consistent estimator $\hat{\mathbf{R}}_{\star}$, which is composed of the estimated autocorrelation sequence \hat{r}_{τ} for $0 \le \tau \le 2m - 1$ as

$$\hat{r}_{\tau} = \frac{1}{N} \sum_{t=1}^{N} y(t+\tau) y(t), \quad \hat{r}_{-\tau} = \hat{r}_{\tau}$$
 (42)

and obtain $[\hat{\mathbf{R}}_{\star}]_{kl}$ as

$$[\hat{\mathbf{R}}_{\star}]_{kl} = \frac{1}{2} \{ \hat{r}_{k-l} + \hat{r}_{k+l-1} \}.$$
(43)

Note in (42) that the estimates of the autocorrelation \hat{r}_{τ} are biased in small samples. We also emphasize that the frequency estimates using $\hat{\mathbf{R}}_{\star}$ will not be large SNR efficient. As a result, one can notice finite sample bias effects in the resulting estimates. As far as the weighted subspace fitting approach is concerned we prefer $\hat{\mathbf{R}}_{\star}$ over $\hat{\mathbf{R}}_{c}$, particularly when the SNR is relatively low. That is because of improved estimation performance and the possibility to explore the inherent structure in

further analysis. We also emphasize that $\hat{\mathbf{R}}_{\star}$ is a strongly consistent estimator of \mathbf{R} . Therefore, by Lemma 1, the estimate obtained using $\hat{\mathbf{R}}_{\star}$ is strongly consistent. From here on, we replace $\hat{\mathbf{R}}_N$ by $\hat{\mathbf{R}}_{\star}$. The matrix of d dominating eigenvectors of $\hat{\mathbf{R}}_{\star}$ will be denoted by $\hat{\mathbf{S}}_N$ to maintain the same notation. We conclude this section with the following lemma, which is a consequence of the structure imposed on $\hat{\mathbf{R}}_{\star}$.

Lemma 3: Let the sequence $q_{\tau}(\boldsymbol{\omega})$ be defined as

$$q_{\tau}(\boldsymbol{\omega}) = \sum_{k=-d}^{d} g_k(\boldsymbol{\omega}) \hat{r}_{k+\tau}$$
(44)

where $|\tau| < 2m - d$. Then, the matrix $\mathbf{G}_0^{\top}(\boldsymbol{\omega}) \hat{\mathbf{R}}_{\star}$ of dimension $(m - d) \times m$ is given element-wise as

$$[\mathbf{G}_{0}^{\top}(\boldsymbol{\omega})\hat{\mathbf{R}}_{\star}]_{lk} = \frac{1}{2}\{q_{l-k}(\boldsymbol{\omega}) + q_{l+k-1}(\boldsymbol{\omega})\}$$
(45)
$$= \frac{1}{4}\{q_{l-k}(\boldsymbol{\omega}) + q_{l-k-1}(\boldsymbol{\omega})\}$$

$$+ q_{l+k-1}(\boldsymbol{\omega}) + q_{l-k}(\boldsymbol{\omega}) + q_{l-k-l}(\boldsymbol{\omega}) + q_{l+k-1}(\boldsymbol{\omega}) + q_{1-k-l}(\boldsymbol{\omega}) \}.$$
(46)

Proof: See Appendix C.

D. Analysis of the Residual

In this section we present a large sample statistical analysis of the residual vector $\hat{\epsilon}_N(\omega_0)$. Such an analysis has a twofold motivation. Firstly, the covariance matrix Δ of the residual vector arises in the expression of the asymptotic covariance matrix of the parameter vector in (25). Secondly, the knowledge of Δ is required to find the optimal weighting \mathbf{W}^* in (27). Let $\hat{\mathbf{R}}_*$ be used to estimate \mathbf{R} . Then for large N, we have the following first order perturbation result [5]:

$$\hat{\boldsymbol{\epsilon}}_{N}(\boldsymbol{\omega}_{0}) = \operatorname{vec}\left\{\mathbf{G}_{0}^{\top}(\boldsymbol{\omega}_{0})\hat{\mathbf{S}}_{N}\right\}$$
$$= \operatorname{vec}\left\{\mathbf{G}_{0}^{\top}(\boldsymbol{\omega}_{0})\hat{\mathbf{R}}_{\star}\mathbf{S}\mathbf{D}^{-1}\right\}$$
$$= (\mathbf{D}^{-1}\mathbf{S}^{\top}\otimes\mathbf{I}_{m-d})\operatorname{vec}\left[\mathbf{G}_{0}^{\top}(\boldsymbol{\omega}_{0})\hat{\mathbf{R}}_{\star}\right]. \quad (47)$$

Next, we simplify (47). First, let us introduce the $(4m - 2d - 1) \times 1$ vector

$$\mathbf{q}(\boldsymbol{\omega}) := [q_{-2m+d+1}(\boldsymbol{\omega}) \quad \cdots \quad q_{2m-d-1}(\boldsymbol{\omega})]. \tag{48}$$

After a few steps of straightforward algebraic manipulations, it can be verified from (46) that

$$\mathbf{G}_{0}^{\top}(\boldsymbol{\omega}_{0})\hat{\mathbf{R}}_{\star} = [\mathbf{T}_{1}\mathbf{q}(\boldsymbol{\omega}_{0}) \quad \cdots \quad \mathbf{T}_{m}\mathbf{q}(\boldsymbol{\omega}_{0})]$$
(49)

where each of $\{\mathbf{T}_k\}_{k=1}^m$ is an $(m-d) \times (4m-2d-1)$ matrix given by

$$\mathbf{T}_{k} = \begin{bmatrix} \mathbf{0}_{(m-d)\times(m+k-1)} & \mathbf{J}_{m-d} & \mathbf{I}_{m-d} & \mathbf{0}_{(m-d)\times(m-k)} \end{bmatrix} \\ + \begin{bmatrix} \mathbf{0}_{(m-d)\times(m-k)} & \mathbf{J}_{m-d} & \mathbf{I}_{m-d} & \mathbf{0}_{(m-d)\times(m+k-1)} \end{bmatrix}$$
(50)

and \mathbf{J}_n is the $n \times n$ permutation matrix having ones along its antidiagonal and zeros elsewhere. Now, define

$$\mathbf{T} := [\mathbf{T}_1^\top \quad \cdots \quad \mathbf{T}_m^\top]^\top. \tag{51}$$

Immediately, using (47) and (49), we get

$$\hat{\boldsymbol{\epsilon}}_N(\boldsymbol{\omega}_0) = \frac{1}{4} (\mathbf{D}^{-1} \mathbf{S}^\top \otimes \mathbf{I}_{m-d}) \mathbf{T} \mathbf{q}(\boldsymbol{\omega}_0).$$
(52)

Thus, to find an expression for Δ , it only remains to find the covariance matrix of $\mathbf{q}(\omega_0)$. We have the following lemma in that direction.

Lemma 4: The asymptotic covariance matrix (as $N \to \infty$) Δ_0 of $\sqrt{N}\mathbf{q}(\boldsymbol{\omega}_0)$ is given by

$$\boldsymbol{\Delta}_0 = \sigma^4 \mathbf{G}_1(\boldsymbol{\omega}_0) \{ \mathbf{I}_{4m-1} + \mathbf{J}_{4m-1} \} \mathbf{G}_1^{\top}(\boldsymbol{\omega}_0)$$
 (53)

where $G_1(\boldsymbol{\omega})$ is a $(4m - 2d - 1) \times (4m - 1)$ matrix given by

$$\mathbf{G}_1(\boldsymbol{\omega})$$

$$= \begin{bmatrix} g_{-d}(\boldsymbol{\omega}) & \cdots & g_d(\boldsymbol{\omega}) & 0 & \cdots & 0\\ 0 & g_{-d}(\boldsymbol{\omega}) & \cdots & g_d(\boldsymbol{\omega}) & \ddots & 0\\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots\\ 0 & \cdots & 0 & g_{-d}(\boldsymbol{\omega}) & \cdots & g_d(\boldsymbol{\omega}) \end{bmatrix}.$$
 (54)

Proof: See Appendix D.

It can be easily verified that
$$\mathbf{I}_{4m-1} + \mathbf{J}_{4m-1} = \mathbf{L}\mathbf{L}^{\top}$$
, where

$$\mathbf{L} = \begin{bmatrix} \mathbf{I}_{2m-1} & \mathbf{0}_{(2m-1)\times 1} \\ \mathbf{0}_{1\times(2m-1)} & \sqrt{2} \\ \mathbf{J}_{2m-1} & \mathbf{0}_{(2m-1)\times 1} \end{bmatrix}.$$
 (55)

Then, it follows from (52) and (53) that

$$\boldsymbol{\Delta} = \boldsymbol{\Psi}(\boldsymbol{\omega}_0) \boldsymbol{\Psi}^\top(\boldsymbol{\omega}_0) \tag{56}$$

where $\Psi(\boldsymbol{\omega})$ is given by

$$\Psi(\boldsymbol{\omega}) = \frac{\sigma^2}{4} (\mathbf{D}^{-1} \mathbf{S}^\top \otimes \mathbf{I}_{m-d}) \mathbf{T} \mathbf{G}_1(\boldsymbol{\omega}) \mathbf{L}.$$
 (57)

From the above, it is clear that if d(m-d) > 2m, the residual covariance matrix Δ is singular. Note that if Δ is a singular matrix, one should replace Δ^{-1} in (27) and (28) by $\{\Delta + \mathcal{J}\mathcal{J}^{\top}\}^{\dagger}$, where \mathbf{X}^{\dagger} denotes the *pseudoinverse* of \mathbf{X} [12]. However, if

$$\operatorname{span}(\mathcal{J}) \subset \operatorname{span}(\Delta)$$
 (58)

it is sufficient to use Δ^{\dagger} as the weighting matrix [13]. In fact, (58) was shown to hold in [6] for the harmonic retrieval problem using complex-valued data. The proof in [6] can be extended readily in the current context. Therefore, for singular Δ , the associated optimal weighting matrix is given by Δ^{\dagger} . In addition, for a singular Δ , the inverse of the matrix $\mathcal{J}^{\top}\Delta^{\dagger}\mathcal{J}$ exists [see (28)] because (58) holds.¹

E. Implementation

The optimally weighted subspace fitting algorithm is a two-step procedure. The main reason for this is the fact that the residual covariance matrix depends on the true parameter vector $\boldsymbol{\omega}_0$. In order to estimate $\boldsymbol{\Psi}(\boldsymbol{\omega}_0)$, an estimate of $\mathbf{G}_1(\boldsymbol{\omega}_0)$ is required. Hence, it is required to replace $\boldsymbol{\omega}_0$ [or more precisely $\mathbf{g}(\boldsymbol{\omega}_0)$] by a consistent estimate in (54). An initial estimate

¹The calculations in [6] can be extended in a straightforward manner to show that \mathcal{J} has a rank d.

 $\hat{\mathbf{g}}(\mathbf{W})$ can be obtained using (38) for any positive definite \mathbf{W} [one can use $\mathbf{W} = \mathbf{I}_{m(m-d)}$, for example], and $\hat{\mathbf{G}}_1(\boldsymbol{\omega})$ is straightforward to obtain from $\hat{\mathbf{g}}(\mathbf{W})$. Note that in (57), we can drop the factor $\sigma^2/4$, since the subspace estimate of $\mathbf{g}(\boldsymbol{\omega}_0)$ in (38) is invariant up to a scaling of \mathbf{W} . Next, in (57), we replace \mathbf{S} by $\hat{\mathbf{S}}_N$ and \mathbf{D} by

$$\hat{\mathbf{D}} = \hat{\mathbf{\Lambda}} - \frac{1}{m-d} \operatorname{tr}\{\hat{\mathbf{\Sigma}}\}$$
(59)

where $\hat{\Lambda}$ and $\hat{\Sigma}$ are obtained from the eigenvalue decomposition of $\hat{\mathbf{R}}_{\star}$; see (15).

From (56) and (57), it appears that the computation of Δ is a computationally expensive operation, particularly for large value of *m*. This is because we need to multiply several matrices of large dimensions. However, by exploiting the structures in **T**, **L**, and **G**₁($\hat{\omega}$) [$\hat{\omega}$ is a consistent estimate of ω_0], we can simplify the computations to a large extent. In fact, it turns out that the task can be achieved without any matrix multiplication. To spell out the details, let us introduce

$$\mathbf{B} = \mathbf{D}^{-1} \mathbf{S}^{\top}, \quad \mathbf{H}_k(\hat{\boldsymbol{\omega}}) = \mathbf{T}_k \mathbf{G}_1(\hat{\boldsymbol{\omega}}) \mathbf{L}$$
(60)

where **B** is a $d \times m$ matrix, and each of $\{\mathbf{H}_k(\hat{\boldsymbol{\omega}})\}_{k=1}^m$ is an $(m - d) \times 2m$ matrix. Now, from (51) and (57), it is readily verified that $\Psi(\hat{\boldsymbol{\omega}})$ can be partitioned as

$$\Psi(\hat{\boldsymbol{\omega}}) = \begin{vmatrix} \Psi_1(\hat{\boldsymbol{\omega}}) \\ \vdots \\ \Psi_d(\hat{\boldsymbol{\omega}}) \end{vmatrix}, \quad \Psi_k(\hat{\boldsymbol{\omega}}) = \sum_{j=1}^m [\mathbf{B}]_{kj} \mathbf{H}_j(\hat{\boldsymbol{\omega}}) \quad (61)$$

for $1 \le k \le m$. The structures in **T**, **L**, and $G_1(\hat{\omega})$ can be exploited to compute $\{H_k(\hat{\omega})\}_{k=1}^m$ in an efficient manner without any matrix multiplication. Again, **D** is a diagonal matrix. Therefore, we can compute **B** in a very efficient way. In what follows next, we show that it is sufficient to compute $\Psi(\hat{\omega})$ to compute the optimal WSF estimate [hence, we need not compute and invert Δ]. Three different cases can arise.

In case d(m − d) < 2m, the number of columns in Ψ(ŵ) is more than the number of rows. In this case, we carry out the QR factorization of Ψ^T(ŵ):

$$\Psi^{+}(\hat{\omega}) = Q\mathcal{R} \tag{62}$$

where Q is a $2m \times d(m-d)$ matrix with mutually orthogonal columns, and \mathcal{R} is a full-rank $d(m-d) \times d(m-d)$ matrix. Then,

$$\mathbf{W}^* = \boldsymbol{\Delta}^{-1} = [\mathcal{R}^\top \mathcal{R}]^{-1}.$$
 (63)

Hence, we can obtain the optimal WSF estimate of $\mathbf{g}(\boldsymbol{\omega}_0)$ by solving

$$\mathcal{R}^{-\top}\mathbf{U}_{N}\mathbf{g} = -\mathcal{R}^{-\top}\mathbf{u}_{N}$$
(64)

for g in the least squares sense.

TABLE I Outline of the Optimally WSF Algorithm

- 1. Compute the covariance sequence ${\{\hat{r}_{\tau}\}}_{-2m+1}^{2m-1}$ in (42) and form the matrix $\hat{\mathbf{R}}_{\star}$ in (43).
- 2. Compute the eigenvalue decomposition of $\widehat{\mathbf{R}}_{\star}$, and form $\widehat{\mathbf{S}}_{N}$ using d dominating eigenvectors of $\widehat{\mathbf{R}}_{\star}$.
- 3. Compute \widehat{D} from the eigenvalues of \widehat{R}_{\star} as in (59), and compute $\mathbf{B} = \widehat{D}^{-1}\widehat{\mathbf{S}}_{N}^{\top}$.
- 4. Form the matrix \mathbf{U}_N and the vector \mathbf{u}_N using (34) and (36).
- 5. Compute $\mathbf{g}(\hat{\omega}_1) = [\mathbf{U}_N^\top \mathbf{U}_N]^{-1} \mathbf{U}_N^\top \mathbf{u}_N$ and $\mathbf{G}_1(\hat{\omega}_1)$ using (54), see also (32) and (33).
- 6. Form the matrices $\{\mathbf{T}_k\}_{k=1}^m$ in (50), L in (55), and $\{\mathbf{H}_k\}_{k=1}^m$ in (60). Compute $\Psi(\hat{\omega})$ using (61).
- 7. If d(m-d) < 2m, use (62) and (64) compute $\hat{\mathbf{g}}_N(\mathbf{W}^*)$, else if d(m-d) = 2m use (65) to compute $\hat{\mathbf{g}}_N(\mathbf{W}^*)$, else use (66) and (68) compute $\hat{\mathbf{g}}_N(\mathbf{W}^*)$.
- 8. Using $\hat{\mathbf{g}}_N(\mathbf{W}^*)$ form $\hat{G}(z)$, see (29) and (33), and compute $\hat{\omega}_N(\mathbf{W})$ by rooting $\hat{G}(z)$.
 - If d(m − d) = 2m, we have Ψ(ŵ) as a full rank square matrix. Then, we can obtain the optimal WSF estimate of g(ω₀) by solving g in

$$\Psi^{-1}(\hat{\boldsymbol{\omega}})\mathbf{U}_N\mathbf{g} = -\Psi^{-1}(\hat{\boldsymbol{\omega}})\mathbf{u}_N.$$
 (65)

In case d(m − d) > 2m, the number of columns in Ψ(ŵ) is less than the number of rows. In this case, Δ is rank deficient. Introduce the QR factorization of Ψ(ŵ):

$$\Psi(\hat{\boldsymbol{\omega}}) = \mathcal{QR} \tag{66}$$

where Q is $d(m-d) \times 2m$ matrix, and \mathcal{R} is a full-rank $2m \times 2m$ matrix. Then

$$\mathbf{W}^* = \mathbf{\Delta}^{\dagger} = \mathcal{Q}[\mathcal{R}\mathcal{R}^{\top}]^{-1}\mathcal{Q}^{\top}.$$
 (67)

Hence, we can obtain the optimal WSF estimate of $g(\omega_0)$ by solving

$$\mathcal{R}^{-1}\mathcal{Q}^{\top}\mathbf{U}_{N}\mathbf{g} = -\mathcal{R}^{-1}\mathcal{Q}^{\top}\mathbf{u}_{N}$$
(68)

for g in the least squares sense.

Note that if m = d + 1, then U_N is a $d \times d$ square matrix. Then, the system of equations in (37) is no longer overdetermined. In this case, the weight W has no impact on the estimates. However, this situation is not practical in the sense that using m = d + 1 generally leads to inaccurate estimates.

The summary of the optimally weighted subspace fitting algorithm is given in Table I. Although the optimal weight computation can be carried out in a computationally efficient way, the optimally weighted WSF algorithm is a two-step procedure anyway. To remedy this difficulty, we introduce an *ad hoc* suboptimal weighting. This is motivated by the second equality in (47). We note for large N that

$$\mathbf{G}_0^{\top}(\boldsymbol{\omega}_0)\hat{\mathbf{S}}_N = \mathbf{G}_0^{\top}(\boldsymbol{\omega}_0)\hat{\mathbf{R}}_{\star}\mathbf{S}\mathbf{D}^{-1}.$$
 (69)

Clearly, the perturbation due to additive noise in the *k*th column of $\mathbf{G}_0^{\top}(\boldsymbol{\omega}_0)\hat{\mathbf{S}}_N$ is inversely proportional to $[\mathbf{D}]_{kk}$. Since the condition number of **D** is often quite large, the estimate of the first column of $\mathbf{G}_0^{\top}(\boldsymbol{\omega}_0)\mathbf{S}$ is much more reliable compared with that of its *d*th column.² Hence, it is natural to weight the *k*th column of $\mathbf{G}_0^{\top}(\boldsymbol{\omega}_0)\hat{\mathbf{S}}_N$ by a factor $[\mathbf{D}]_{kk}$. This would mean that

$$\mathbf{W}_1 = [\mathbf{D} \otimes \mathbf{I}_{m-d}]^\top [\mathbf{D} \otimes \mathbf{I}_{m-d}]$$
(70)

is used as the weighting matrix. As before, for practical implementation, one needs to replace \mathbf{D} by $\hat{\mathbf{D}}$. However, this suboptimal weighting can be implemented without any knowledge of $\boldsymbol{\omega}_0$, and hence, it is a one-step procedure. In addition, the implementation of the weighting can be done in an efficient manner, where one multiplies each individual column of $\hat{\mathbf{S}}_N$ by the corresponding diagonal entry of $\hat{\mathbf{D}}$, which can be accomplished by md real multiplications. As a result, the suboptimal weighting has much smaller computational complexity. We show in the next section that the estimate corresponding to this *ad hoc* weighting yields estimates having accuracy close to that of the optimal estimate $\hat{\boldsymbol{\omega}}_N(\mathbf{W}^*)$.

IV. NUMERICAL ILLUSTRATIONS

In this section, we illustrate the performance of the algorithms discussed in the previous sections using numerical simulations. We consider the optimally WSF approach using both the covariance estimates $\hat{\mathbf{R}}_{\star}$ and $\hat{\mathbf{R}}_{c}$ and the subspace fitting with *ad hoc* weighting (denoted AD-HOC) in (70). The optimally weighted subspace fitting approach using $\hat{\mathbf{R}}_{\star}$ will be denoted WSF-1, whereas that using $\hat{\mathbf{R}}_c$ will be denoted WSF-2. We compare the performance of the proposed algorithms with forward-backward ESPRIT (FB-ESPRIT) (as a representative of the traditional subspace approaches), R-Esprit developed in [3], and the nonlinear least squares (NLLS) [14]. We choose FB-ESPRIT as a benchmark since it is known to yield accurate estimates (as good as WSF estimates for complex data model [4], [6], and other subspace algorithms) at comparatively less computational burden. The nonlinear least squares algorithm is chosen as another benchmark because it is the maximum likelihood algorithm and is asymptotically efficient (in a statistical sense). The root mean square estimation error will be used as the quantitative performance measure of different algorithms. Let $\{\hat{\theta}_k\}_{k=1}^{N_r}$ be the estimates of θ in N_r independent Monte Carlo simulations. Then, we express the root mean square error (RMSE) is computed as

$$\text{RMSE}(\hat{\theta}) = \sqrt{\frac{1}{N_r}\sum_{k=1}^{N_r} (\hat{\theta}_k - \theta)^2}.$$

²Note that we assume

$$[D]_{11} > \cdots > [D]_{dd}$$



Fig. 1. Comparison of the RMSE as a function of m obtained from different approaches ($N = 30, \omega_1 = 1.4, \omega_2 = 1.5$) at 7 dB SNR. (a) ω_1 . (b) ω_2 .

In Fig. 1, we have plotted the RMSE obtained using different algorithms as functions of m: the dimension of the snapshot vector. The frequencies are $\omega_1 = 1.4$ and $\omega_2 = 1.5$, whereas the amplitudes of the sinusoids are $a_1 = 1$ and $a_2 = 1$, respectively. The SNR $(a_k^2/2\sigma^2)$ is 7 dB [corresponding to $\sigma = 0.32$]. The number of data samples N is 30. For each value of m, the individual RMSE of the estimates were computed from 100 independent Monte Carlo simulations. Note that R-Esprit and WSF-2 use $\hat{\mathbf{R}}_c$ as the covariance estimate. Therefore, for these two algorithms, we need to satisfy $N - 2m + 1 \ge d$. This implies that the maximum permissible value of m is 14 in this example. Hence, we have plotted the RMSE of WSF-2 and R-Esprit for m < 14. Since we use $\hat{\mathbf{R}}_{\star}$ to compute the WSF-1 and AD-HOC estimates, there is no such limitation on m. Note that in the NLLS method, we need to solve an optimization problem numerically. In fact, it is not always possible to find the global minimum point. It is therefore suggested that we use a suboptimal method like ESPRIT or the high-order Yule-Walker approach to obtain the initial point [14]. However, in this example, the SNR is *low*, and these suboptimal initializing methods can lead to misleading results (see below). Hence, for performance comparison, the NLLS optimization was initialized at the *true value* of the parameter vector.

Notice that in Fig. 1, the RMSEs of FB-ESPRIT, R-Esprit, and WSF-2 are greater than the difference between the two frequencies [0.1 rad]. This is a clear indication that at this noise level none of FB-ESPRIT, R-Esprit, or WSF-2 is able to resolve the frequencies. On the other hand, the subspace fitting based approaches based on $\hat{\mathbf{R}}_{\star}$ [WSF-1 and AD-HOC] and NLLS are able to resolve the frequencies. We emphasize that the NLLS estimate does not depend on m. Hence, the RMSE of the NLLS is independent of m. It is interesting to notice that WSF-1 and AD-HOC outperform the NLLS method (although NLLS is actually the maximum likelihood estimate). The reason behind this phenomenon is the fact that the subspace fitting-based approaches are more robust to outliers.³

In Fig. 2, we have depicted the variations of the RMSE obtained using different methods as functions of the SNR. The values of m for subspace-based methods were kept fixed. For R-Esprit and WSF-2, we have chosen m = 10. For WSF-2 and AD-HOC, we have m = 12. For FB-ESPRIT, m = 15. These values were chosen in order to achieve the best performances (approximately) for N = 30. In Fig. 2, we note for SNR below 15 dB, FB-ESPRIT and WSF-2 are unable to resolve the frequencies. R-Esprit can resolve the frequencies at 15 dB (although the performance is poor). On average, WSF-1 and AD-HOC performs as well as NLLS in the range 5–12 dB. For SNR below 5 dB, NLLS outperforms WSF-1 and AD-HOC. However, we emphasize that this performance of NLLS is obtained using initialization at the true parameter values, which is not practical. It is also interesting to note that the improved SNR beyond 7 dB does not bring significant improvement in the performance of WSF-1 and AD-HOC estimates. This is because WSF-1 and AD-HOC use $\hat{\mathbf{R}}_{\star}$ as the estimate of \mathbf{R} and, hence, are not SNR efficient. Note that even at 15 dB SNR, FB-ESPRIT is unable to resolve the frequencies. We emphasize in this context that FB-ESPRIT is generally capable of resolving frequencies at similar spacing using the same number of data samples when the data are composed of complex-valued sinusoids [4], [6]. It is also interesting to note both in Figs. 1 and 2 that the loss in the estimation performance using the ad hoc weighting is not significant (about 0.03 rad in RMSE) when compared with the optimally WSF approach.

The computational complexity of different algorithms are compared in Fig. 3, where we have shown the Matlab floating-point operations required to compute different estimates. The implementation was done using Matlab 5.3. In the implementation, all the linear systems of equations were (arising at different steps of ESPRIT and subspace fitting algorithms) solved using a QR factorization-based algorithm for solving linear equations. In the optimal weight computation for the WSF approach, the structures in the matrices in (57) were exploited in order to minimize the number of matrix



Fig. 2. Comparison of the RMSE of different methods as functions of the SNR for fixed $m(N = 30, \omega_1 = 1.4, \omega_2 = 1.5)$. (a) ω_1 . (b) ω_2 .

multiplications. The numerical optimization involved in the nonlinear least squares algorithm were carried out using the separable least squares method (known as the variable projection algorithm in the literature) [15] using the fminsearch function in Matlab. The NLLS optimization was initialized using the R-Esprit algorithm.⁴ In Fig. 3(a), we have plotted the Matlab floating-point operations required by different algorithms as functions of m, whereas the number of data points N is kept fixed at 30. It is interesting to note in Fig. 3(a) that the complexity of the WSF-based approaches are even smaller than that of the FB-ESPRIT algorithm for smaller values of m. In fact, for the values of m of interest, the computational complexity of the AD-HOC approach is comparable to that of R-Esprit, whereas the complexity of WSF-1 lies between that of FB-ESPRIT and R-Esprit. For a larger m (which is not a

³The probability of getting an outlier at 7 dB SNR is reasonably high. From the simulation, this probability was estimated to be 9%.

⁴We have used noise-free data for generating plots in Fig. 3. R-Esprit is the computationally most economical alternative to compute the initial values. It is also reliable at high SNR. However, as we have seen before, this is not a good choice at low SNR.



Fig. 3. Comparison of the Matlab floating-point operations as a function of m obtained from different approaches (N = 30). (b) Matlab floating-point operations as a function of N for a fixed m.

practical choice), however, the computational complexities of WSF-1 and AD-HOC increase significantly. This is because we need floating-point operations of the order m^3 to compute the eigen-decomposition of $\hat{\mathbf{R}}_{\star}$. Solving the normal equations (38) is also computationally demanding for larger values of m. In Fig. 3(b), the Matlab floating-point operations associated with different approaches are plotted as functions of N, whereas m is kept fixed. As before, we have m = 10 for R-Esprit and WSF-2; m = 12 for WSF-1 and AD-HOC; and m = 15 for FB-ESPRIT. Note that the complexity of WSF-1 and AD-HOC remains relatively constant with the increase in the data size. This is because the computational cost of computing the correlation estimates \hat{r}_{τ} increase at a very slow rate with the data size, whereas the computation involved in the subsequent steps remain unchanged. This makes WSF-1 and AD-HOC methods to be computationally economical alternatives for large N. Another interesting point here is that the complexity of AD-HOC weighting method is about 30% of that of WSF-1 for the values

of m of interest. However, associated loss in accuracy is not that significant.

V. CONCLUSION

In this paper, we have proposed a subspace-fitting approach for estimation of the real-valued sine wave frequencies. We have presented a novel noise subspace parameterization for the realvalued data model. This new parameterization helps to simplify the associated optimization problem and solve it in an analytical and noniterative manner. We have also presented a statistical analysis of the residual vector, which leads to the optimally WSF estimate. Subsequently, we have proposed a suboptimal weighting for which the corresponding estimation accuracy is close to that of the optimal one, whereas the savings in the computational load is significant. The same suboptimal weighting can be used in any subspace fitting problem in a broader perspective, sensor array processing being an example. The main advantage of the proposed algorithm is its improved resolution capability in the presence of noise level. Using simulation study, we have shown that the proposed algorithms can resolve frequencies when the conventional subspace approaches fail. At higher noise levels, the proposed approach can outperform the maximum likelihood method. The optimally WSF approach gives best results with N/3 < m < N/2 (N is the number of data samples). At this range, the complexity of optimally weighted WSF approach is less than that of the FB-ES-PRIT method. The ad hoc weighting strategy can be seen as a very low complexity alternative, where we need to compromise a little in the accuracy. However, the resulting estimates are biased. This can be seen as a bias-variance tradeoff that is done in the proposed WSF approach, where the effect of noise on the covariance matrix of the estimates is reduced significantly for the expense of finite sample bias. This bias effects can be eliminated if the covariance estimate $\hat{\mathbf{R}}_c$ is used. Hence, for data with large SNR (or large N), \mathbf{R}_c should be used, whereas for low SNR data and small $N, \hat{\mathbf{R}}_{\star}$ is the user's choice. It is also interesting to study the effect of using a convex combination

$$\hat{\mathbf{R}}_{\alpha} = \alpha \hat{\mathbf{R}}_{\star} + (1 - \alpha) \hat{\mathbf{R}}_{c}, \quad 0 \le \alpha \le 1$$

as the estimator of **R**, where the parameter α depends on the SNR and N.

APPENDIX A PROOF OF LEMMA 1

In this Appendix, we denote the loss function in (23) by $\ell_N(\boldsymbol{\omega}, \mathbf{W})$ and introduce the *limiting* loss function $\ell_{\star}(\boldsymbol{\omega}, \mathbf{W})$:

$$\ell_N(\boldsymbol{\omega}, \mathbf{W}) = \hat{\boldsymbol{\epsilon}}_N^{\top}(\boldsymbol{\omega}) \mathbf{W} \hat{\boldsymbol{\epsilon}}_N(\boldsymbol{\omega})$$

$$\ell_{\star}(\boldsymbol{\omega}, \mathbf{W}) = \boldsymbol{\epsilon}^{\top}(\boldsymbol{\omega}) \mathbf{W} \boldsymbol{\epsilon}(\boldsymbol{\omega}).$$
(71)

The proof of the lemma will be given in three main steps. First, we establish that $\ell_N(\boldsymbol{\omega}, \mathbf{W})$ converges uniformly with probability one to $\ell_{\star}(\boldsymbol{\omega}, \mathbf{W})$ as $N \to \infty$. Then, we show that $\boldsymbol{\omega}_0$ is the unique global minimum point of the limiting cost function $\ell_{\star}(\boldsymbol{\omega}, \mathbf{W})$ in \mathcal{D} . Finally, these two facts will be used to show that $\hat{\boldsymbol{\omega}}(\mathbf{W}) \to \boldsymbol{\omega}_0$ as $N \to \infty$ with probability one.

The covariance estimate $\hat{\mathbf{R}}_N$ converges to \mathbf{R} as $N \to \infty$ with probability one.⁵ Hence, $\hat{\mathbf{S}}_N$ converges to \mathbf{S} as $N \to \infty$ with probability one [16]. Then, from (21), we can see that $\hat{\boldsymbol{\epsilon}}_N(\boldsymbol{\omega}) \to$ $\boldsymbol{\epsilon}(\boldsymbol{\omega})$ uniformly as $N \to \infty$ with probability one. Note that for uniform convergence, we need to have elements of $\mathcal{G}(\boldsymbol{\omega})$ to be uniformly bounded, which is not a restriction. Then, from the definition in (71), we get $\ell_N(\boldsymbol{\omega}, \mathbf{W}) \to \ell_{\star}(\boldsymbol{\omega}, \mathbf{W})$ uniformly with probability one as $N \to \infty$.

To show that $\boldsymbol{\omega}_0$ is the unique global minimum point of $\ell_{\star}(\boldsymbol{\omega}, \mathbf{W})$ in \mathcal{D} , first, note from (5) and (71) that $\ell_{\star}(\boldsymbol{\omega}_0, \mathbf{W}) = 0$. Since \mathbf{W} is a positive definite matrix, we note that $\boldsymbol{\omega}_0$ is a global minimum of $\ell_{\star}(\boldsymbol{\omega}, \mathbf{W})$ in \mathcal{D} . To prove the uniqueness, let us assume that there exists another point $\boldsymbol{\omega}_1 \in \mathcal{D}$ such that $\ell_{\star}(\boldsymbol{\omega}_1, \mathbf{W}) = 0$ and $\boldsymbol{\omega}_0 \neq \boldsymbol{\omega}_1$. From positive definiteness of \mathbf{W} , this implies that

$$\boldsymbol{\epsilon}(\boldsymbol{\omega}_1) = \mathbf{0}_{d(m-d) \times 1} \quad \Rightarrow \mathcal{G}^{\top}(\boldsymbol{\omega}_1) \mathbf{S} = \mathbf{0}_{(m-d) \times d}.$$

Since both $\mathcal{G}(\boldsymbol{\omega}_1)$ and **S** have full column rank, the last equation implies that **S**, $\mathbf{A}(\boldsymbol{\omega}_1)$, and $\mathbf{A}(\boldsymbol{\omega}_0)$ have the identical column space; see also (16). Thus, the rank of the matrix

$$\mathbf{A}_{\star} = \begin{bmatrix} \mathbf{A}(\boldsymbol{\omega}_0) & \mathbf{A}(\boldsymbol{\omega}_1) \end{bmatrix}$$

is d. However, this is contradiction since $\omega_0 \neq \omega_1$ implies that the rank of \mathbf{A}_{\star} is at least d + 1 [3]. Hence, ω_0 is the unique global minimum point of $\ell_{\star}(\omega, \mathbf{W})$ in \mathcal{D} .

Now, $\boldsymbol{\omega}_0$ is an interior point of the open set \mathcal{D} . Using a wellknown result in [17] (see also [18]), we conclude that the global minimum point $\hat{\boldsymbol{\omega}}_N(\mathbf{W})$ of the loss function $\ell_N(\boldsymbol{\omega}, \mathbf{W})$ converges to the unique global minimum point $\boldsymbol{\omega}_0$ of the limiting loss function $\ell_{\star}(\boldsymbol{\omega}, \mathbf{W})$ with probability one as $N \to \infty$.

APPENDIX B PROOF OF LEMMA 2

From the construction of the polynomial $G(z, \boldsymbol{\omega})$ in (29), it follows that

$$G(e^{\mathbf{i}\varpi_k}, \boldsymbol{\omega}) = 0, \quad 1 \le k \le d.$$
(72)

Multiplying both sides of (72) by $e^{i\phi}$ (for any arbitrary ϕ) and taking the real part, we have

$$\sum_{\tau=-d}^{d} g_{\tau}(\boldsymbol{\omega}) \cos(\varpi_k \tau + \phi) = 0.$$
 (73)

Next, using (73), we show that every element of the matrix

$$\mathbf{X} := \mathbf{G}_0^\top(\boldsymbol{\omega}) \mathbf{A}(\boldsymbol{\omega}) \tag{74}$$

equals zero. To that aim, we first note using (8) and (30) that a generic element of **X** is given by

$$[\mathbf{X}]_{kl} = \sum_{j=1}^{m} \{g_{j-k}(\boldsymbol{\omega}) \cos[(j-1/2)\varpi_l] + g_{1-j-k}(\boldsymbol{\omega}) \cos[(1/2-j)\varpi_l]\}$$
(75)

⁵The same is true for $\hat{\mathbf{R}}_c$ and $\hat{\mathbf{R}}_{\star}$. Therefore, the proof also holds for weighted subspace fitting estimate using $\hat{\mathbf{R}}_c$ and $\hat{\mathbf{R}}_{\star}$.

where we have used $\cos(-\theta) = \cos(\theta)$ and $g_k(\omega) = g_{-k}(\omega)$. Now, with a straightforward change in variables in (75), it is easy to derive that

$$[\mathbf{X}]_{kl} = \sum_{\tau = -m-k+1}^{m-k} g_{\tau}(\boldsymbol{\omega}) \cos[(\tau + k - 1/2)\varpi_l].$$
(76)

Now, recall that $1 \le k \le m - d$. Hence, we see that $m - k \ge d$ and $-m + 1 - k \le -m < -d$. Since $g_{\tau}(\omega) = 0$ if $|\tau| > d$, we have from (76)

$$[\mathbf{X}]_{kl} = \sum_{\tau = -m-k+1}^{m-k} g_{\tau}(\boldsymbol{\omega}) \cos[\varpi_l \tau + \varpi_l(k-1/2)] = 0 \quad (77)$$

using (73). It remains to show that $\mathbf{G}_0(\boldsymbol{\omega})$ has a full column rank. Partition $\mathbf{G}_0(\boldsymbol{\omega})$ into a $d \times (m - d)$ matrix $\mathbf{G}_{11}(\boldsymbol{\omega})$ and a $(m - d) \times (m - d)$ square matrix $\mathbf{G}_{21}(\boldsymbol{\omega})$ such that

$$\mathbf{G}_{0}(\boldsymbol{\omega}) = \begin{bmatrix} \mathbf{G}_{11}(\boldsymbol{\omega}) \\ \mathbf{G}_{21}(\boldsymbol{\omega}) \end{bmatrix}.$$
 (78)

We show that $\mathbf{G}_{21}(\boldsymbol{\omega})$ is an upper triangular matrix. Consider the element at the (j + k)th row and jth column of $\mathbf{G}_{21}(\boldsymbol{\omega})$, where k > 0. We have

$$[\mathbf{G}_{21}(\boldsymbol{\omega})]_{(j+k),j} = [\mathbf{G}_0(\boldsymbol{\omega})]_{(d+j+k),j}$$
$$= g_{d+k}(\boldsymbol{\omega}) + g_{d+2j+k-1}(\boldsymbol{\omega}) = 0 \quad (79)$$

since j > 0, and k > 0, and by definition, $g_{\tau}(\boldsymbol{\omega}) = 0$ for all $|\tau| > d$. Similarly, we can verify that the elements in the upper triangle of $\mathbf{G}_{21}(\boldsymbol{\omega})$ are nonzero. In fact, using k = 0 in (79), we see that $[\mathbf{G}_{21}(\boldsymbol{\omega})]_{jj} = g_d(\boldsymbol{\omega})$ for all j. With this observation, the lemma is proved.

APPENDIX C PROOF OF LEMMA 3

Let

$$\mathbf{P}_1 = \mathbf{G}_0^\top(\boldsymbol{\omega}) \hat{\mathbf{R}}_\star. \tag{80}$$

Using (30) and (43), we get

$$[\mathbf{P}_1]_{ij} = \frac{1}{2} \sum_{k=1}^{m} \{g_{i-k} + g_{i+k-1}\} \{\hat{r}_{k-j} + \hat{r}_{k+j-1}\}$$
(81)

where we have omitted the argument $\boldsymbol{\omega}$ for simplicity. Using $g_{i-k} = g_{k-i}$, we have after a straightforward change of sum indices

$$2[\mathbf{P}_{1}]_{ij} = \sum_{\tau=1-i}^{m-i} g_{\tau} \hat{r}_{\tau+i-j} + \sum_{\tau=i}^{m+i-1} g_{\tau} \hat{r}_{\tau+j-i} + \sum_{\tau=1-i}^{m-i} g_{\tau} \hat{r}_{\tau+i+j-1} + \sum_{\tau=i}^{m+i-1} g_{\tau} \hat{r}_{\tau-i-j+1} = \sum_{\tau=-m-i+1}^{m-i} g_{\tau} \hat{r}_{\tau+i-j} + \sum_{\tau=-m-i+1}^{m-i} g_{\tau} \hat{r}_{\tau+i+j-1}$$
(82)

where in the last equation, we have used that $g_t = g_{-t}$, and $\hat{r}_t = \hat{r}_{-t}$. Recall that $1 \le i \le m - d$. Therefore, the maximum value of the lower sum index is -m, whereas the minimum value of

the upper sum index is d. In addition, $g_{\tau} = 0$ for $|\tau| > d$. Hence, The newly introduced matrix Q is of the form from (44), we get

$$2[\mathbf{P}_1]_{ij} = q_{i-j} + q_{i+j-1} \tag{83}$$

which proves (45). Note that by definition, $q_{\tau}(\boldsymbol{\omega})$ is a symmetric sequence. This fact leads us to (46).

APPENDIX D **PROOF OF LEMMA 4**

Let us define the correlation vector

$$\mathbf{r} = \begin{bmatrix} r_{-2m+1} & \cdots & r_0 & \cdots & r_{2m-1} \end{bmatrix}^\top.$$
(84)

Then, from the definition of G_1 in (54) and the definition of qin (48), we have, using (44)

$$\mathbf{q}(\boldsymbol{\omega}) = \mathbf{G}_1(\boldsymbol{\omega})\hat{\mathbf{r}} \tag{85}$$

where $\hat{\mathbf{r}}$ denotes the estimate of \mathbf{r} . Introduce

$$\boldsymbol{y}_t = \begin{bmatrix} y(t-2m+1) & \cdots & y(t) & \cdots & y(t+2m-1) \end{bmatrix}^\top \\ = \boldsymbol{y}_{0t} + \tilde{\boldsymbol{y}}_t \tag{86}$$

where y_{0t} is the noise-free part of y(t), whereas \tilde{y}_t is the noise contribution to y(t). In this Appendix, we use

$$\hat{\mathbf{r}} = \frac{1}{N} \sum_{t=2m}^{N-2m+1} \boldsymbol{y}_t y(t).$$
(87)

Note that the estimate of the autocorrelation sequence according to (87) is different from that obtained by (42). However, this difference is only due to the *edge effects* and vanishes as N increases. Since we are interested only in the large sample properties $(N \to \infty)$, in this section, we are permitted to use the estimate (87) instead of (42). It is easy to show⁶ from (1), (54), and (73) that

$$\mathbf{G}_1(\boldsymbol{\omega}_0)\boldsymbol{y}_{0t} = \boldsymbol{0}_{(2m-1)\times 1}.$$
(88)

Hence, we have

$$\mathbf{q}(\boldsymbol{\omega}_0) = \frac{1}{N} \mathbf{G}_1(\boldsymbol{\omega}_0) \sum_{t=2m}^{N-2m+1} \tilde{\boldsymbol{y}}_t \{ y_0(t) + \tilde{y}(t) \}.$$
(89)

Since $y_0(t)$ and $\tilde{y}(t)$ are independent of each other, it follows from (89) that

$$\mathbf{Eq}(\boldsymbol{\omega}_0) = \frac{\sigma^2 M_1}{N} \mathbf{G}_1(\boldsymbol{\omega}_0) \boldsymbol{e}$$
(90)

where we denote $M_1 = N - 4m + 2$ and

$$\boldsymbol{e} = \begin{bmatrix} \boldsymbol{0}_{1 \times (2m-1)} & 1 & \boldsymbol{0}_{1 \times (2m-1)} \end{bmatrix}^\top.$$
(91)

Next, in the aim of computing the second-order moments, we note that the third-order moments of a jointly Gaussian distributed random variable are zero. Using the well-known formula for the fourth-order moments of jointly Gaussian random variables, we have

$$\mathbf{E}\mathbf{q}(\boldsymbol{\omega}_0)\mathbf{q}^{\top}(\boldsymbol{\omega}_0) = \frac{1}{N^2}\mathbf{G}_1(\boldsymbol{\omega}_0)\mathbf{Q}\mathbf{G}_1^{\top}(\boldsymbol{\omega}_0). \tag{92}$$

⁶Arguments are similar to that of the Proof of Lemma 1.

$$\mathbf{Q} = \mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{Q}_3 + \mathbf{Q}_4 \tag{93}$$

where

$$\mathbf{Q}_{1} = \sum_{t_{1}=2m}^{N-2m+1} \sum_{t_{2}=2m}^{N-2m+1} \mathbf{E} \left\{ \tilde{\boldsymbol{y}}_{t_{1}} \tilde{\boldsymbol{y}}_{t_{2}}^{\top} \right\} \mathbf{E} \{ y_{0}(t_{1}) y_{0}(t_{2}) \}$$
(94)
$$\mathbf{Q}_{2} = \sum_{t_{1}=2m}^{N-2m+1} \sum_{t_{2}=2m}^{N-2m+1} \mathbf{E} \left\{ \tilde{\boldsymbol{y}}_{t_{1}} \tilde{\boldsymbol{y}}(t_{1}) \right\} \mathbf{E} \left\{ \tilde{\boldsymbol{y}}(t_{2}) \tilde{\boldsymbol{y}}_{t_{2}}^{\top} \right\}$$
$$= \sigma^{4} M_{1}^{2} \boldsymbol{e} \boldsymbol{e}^{\top}$$
(95)

$$\mathbf{Q}_{3} = \sum_{t_{1}=2m}^{N-2m+1} \sum_{t_{2}=2m}^{N-2m+1} \mathbf{E}\left\{\tilde{\boldsymbol{y}}_{t_{1}}\tilde{\boldsymbol{y}}_{t_{2}}^{\top}\right\} \mathbf{E}\left\{\tilde{y}(t_{1})\tilde{y}(t_{2})\right\} \quad (96)$$

$$\mathbf{Q}_{4} = \sum_{t_{1}=2m}^{N-2m+1} \sum_{t_{2}=2m}^{N-2m+1} \mathbf{E} \left\{ \tilde{\boldsymbol{y}}_{t_{1}} \tilde{\boldsymbol{y}}(t_{2}) \right\} \mathbf{E} \left\{ \tilde{\boldsymbol{y}}(t_{1}) \tilde{\boldsymbol{y}}_{t_{2}}^{\top} \right\}. (97)$$

In order to evaluate the right-hand side of (94), we get by straightforward calculations that Q_1 is a symmetric Toeplitz matrix, which is given element-wise as

$$[\mathbf{Q}_1]_{ij} = \sigma^2 (M_1 - |i - j|) \sum_{k=1}^n \frac{a_k^2}{2} \cos\{(i - j)\omega_k\}.$$
 (98)

After a few more steps of straightforward calculations, one can show using (73) and (54) that as $N \to \infty$

$$\mathbf{G}_{1}(\boldsymbol{\omega}_{0})\mathbf{Q}_{1}\mathbf{G}_{1}^{\top}(\boldsymbol{\omega}_{0}) = \mathbf{0}_{(4m-2d-1)\times(4m-2d-1)}.$$
 (99)

Similarly, it is easy to verify that Q_3 is a diagonal matrix and that \mathbf{Q}_4 is an antidiagonal matrix, which are given element-wise as

$$[\mathbf{Q}_3]_{ij} = \sigma^4 (M - |2m - i|) \delta_{i,j}$$
(100)

$$[\mathbf{Q}_4]_{ij} = \sigma^4 (M - |2m - i|) \delta_{i+j,4m}.$$
 (101)

Taking $N \to \infty$ and combining (90)–(101), the lemma follows.

ACKNOWLEDGMENT

The author would like to thank Prof. T. Söderström for his valuable comments on this manuscript. He is also indebted to the anonymous reviewers for their constructive suggestions.

REFERENCES

- [1] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in Proc. RADC Spectral Estimation Workshop, Rome, NY, 1979, pp. 243-258
- [2] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," IEEE Trans. Acoust., Speech, Signal Process., vol. 37, no. 7, pp. 984-995, Jul. 1989.
- [3] K. Mahata and T. Söderström, "Esprit like estimation of real-valued sinusoidal frequencies," IEEE Trans. Signal Process., vol. 52, no. 5, pp. 1161-1170, May 2004.
- [4] A. Eriksson, P. Stoica, and T. Söderström, "Markov-based eigenanalysis method for frequency estimation," IEEE Trans. Signal Process., vol. 42, no. 3, pp. 586-594, Mar. 1994.
- P. Stoica and T. Söderström, "Statistical analysis of MUSIC and sub-[5] space rotation estimates of sinusoidal frequencies," IEEE Trans. Signal Process., vol. 39, no. 8, pp. 1836–1847, Aug. 1991.

- [6] M. Kristensson, M. Jansson, and B. Ottersten, "Further results and insights on subspace based sinusoidal frequency estimation," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 2962–2974, Dec. 2001.
 [7] M. Viberg and B. Ottersten, "Sensor array processing based on subspace
- [7] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1110–1121, May 1991.
- [8] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1997.
- [9] P. Stoica and A. Eriksson, "MUSIC estimation of real-valued sine wave frequencies," *Signal Process.*, vol. 42, no. 4, pp. 139–146, Apr. 1995.
- [10] B. Ottersten, M. Viberg, and T. Kailath, "Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data," *IEEE Trans. Signal Process.*, vol. 40, no. 3, pp. 590–600, Mar. 1992.
- [11] P. Stoica and K. C. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 7, pp. 1132–1143, Jul. 1990.
- [12] T. Söderström and P. Stoica, *System Identification*. Hemel Hempstead, U.K.: Prentice-Hall, 1989.
- [13] C. R. Rao and S. K. Mitra, Generalized Inverse of Matrices and Its Applications, First ed. New York: Wiley, 1971.
- [14] P. Stoica, R. Moses, B. Friedlander, and T. Söderström, "Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 3, pp. 378–392, Mar. 1989.
- [15] G. H. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM J. Numerical Anal.*, vol. 10, no. 2, pp. 413–432, 1973.

- [16] P. Stoica and T. Söderström, "Eigenelement statistics of sample covariance matrix in the correlated data case," *Digital Signal Process.*, vol. 7, pp. 136–143, 1997.
- [17] T. Söderström, "Convergence properties of the generalized least squares identification algorithm," *Automatica*, vol. 10, pp. 617–626, 1974.
- [18] J. Schoukens, R. Pintelon, G. Vandersteen, and P. Guillaume, "Frequency-domain system identification using nonparametric noise models estimated from a small number of data sets," *Automatica*, vol. 33, no. 6, pp. 1073–1086, June 1997.



Kaushik Mahata received the M.E. degree in signal processing from Indian Institute of Science, Bangalore, India, in 2000 and the Ph.D. degree in signal processing from Uppsala University, Uppsala, Sweden in 2003.

Currently, he is a research academic with the university of Newcastle, Callaghan, Australia. His research interests include estimation identification and spectrum analysis.