Subspace estimation of real-valued sinusoidal frequencies *

Kaushik Mahata and Torsten Söderström

Systems and Control, Department of Information Technology,
Uppsala University, P O Box 337, SE-75105 Uppsala, Sweden
email: Kaushik.Mahata@it.uu.se, Torsten.Soderstrom@it.uu.se

version January 30, 2003

Abstract

Subspace based estimation of multiple real-valued sine wave frequencies is considered in this paper. A novel data covariance model is proposed. In the proposed model the dimension of the signal subspace equals the number of frequencies present in the data, which is half of the signal subspace dimension for the conventional model. Consequently, an ESPRIT like algorithm using the proposed data model is presented. The proposed algorithm is then extended for the case of complex-valued sine waves. Performance analysis of the proposed algorithms are also carried out. The algorithms are tested in numerical simulations. The newly proposed algorithm outperforms ESPRIT by a significant margin and is computationally efficient as well.

*This work was supported by Swedish Research Council for Engineering Sciences under contract 2000-587.

1 Introduction

Super-resolution subspace based estimation of the frequencies of complex-valued sinusoids is well documented in the literature, see [23] and references therein. However, the literature is scarce when it comes to the application of the same methodology to the frequency estimation of real-valued sinusoidal signals. According to the best of our knowledge, such an attempt was made only in [22], where it was shown that the application of MUSIC algorithm to real-valued sinusoidal signals leads to a frequency estimation problem, which has a whole class of solutions and the conventional MUSIC estimator is an *ad hoc* choice in that class. Although it is generally argued that all the algorithms for complex-valued sinusoids are applicable to real-valued data pretending that the data consists of a double number of complex-valued sinusoidal signals, it is reasonable to expect better results if the information that the data is real-valued can somehow be incorporated in the algorithm. However, incorporation of such a constraint may lead to additional computational complexity. In fact, in [22] it was reported that the accuracy degradation using the conventional MUSIC approach is not significant enough as compared to the corresponding savings in the computational cost. Thus, from a computational standpoint, the conventional approach using complex-valued signal model is the method of choice for most applications. Therefore it is still an open question whether it is possible to incorporate the information that the data is real-valued in such a fashion that the resulting algorithm is computationally efficient and also more accurate. In this paper we propose a modified ESPRIT algorithm for real-valued sine waves. We shall show that it is possible to achieve improved accuracy of the estimates without increasing the computational cost when the modified ESPRIT method is employed. In that aim let us consider

$$y(t) = \sum_{k=1}^{n} a_k \cos\{\omega_k t + \phi_k\} + \epsilon(t) \tag{1.1}$$

where $\epsilon(t)$ is a zero mean white noise of variance σ_1^2, and ϕ_k is uniformly distributed in $[0, 2\pi)$ for all $1 \leq k \leq n$. It is further assumed that the random variables $\{\phi_k\}_{k=1}^{n}$ are pairwise independent, and also independent of $\epsilon(t)$. The central problem under consideration of this paper is to estimate the frequencies $\{\omega_k\}_{k=1}^{n}$ from the observed data $\{y(t)\}_{t=1}^{N}$, where $\omega_k \neq 0$ for all k.

The first step to all subspace based methods, is the formation of a finite sample estimate of the data covariance matrix. To form the data covariance matrix, the data set is broken into small overlapping fragments or *snapshots* of length m using a sliding window technique known as spatial smoothing[1]. As a next step it is shown that the noise free component of each of the snapshot vectors lie in a smaller subspace of an m dimensional space, commonly known as the *signal subspace*. The orthogonal complement of the signal subspace is referred to as the *noise subspace*. The singular value decomposition of the data covariance matrix then leads us to

[1]Strictly speaking, the terminology spatial smoothing is used more in Array Signal Processing literature, see [17]. However, the basic data processing steps being same in DOA estimation in spatial spectrum analysis and frequency estimation in temporal spectrum analysis problem, use of the same terminology has been extended to frequency estimation problem also, see for example [5].

an estimated orthogonal basis of the signal subspace, which is then used to extract the frequency estimates. The estimation accuracy of the subspace based technique is closely related to the relative dimensions of the signal and noise subspaces. In general the estimation accuracy increases significantly with the increase in the dimension of the noise subspace [2, 25]. Using the usual approach of modeling a real-valued signal gives a signal subspace of dimension $2n$, see [23] or the next section of this paper. One of the major contributions in this paper is to construct a data covariance model where the noise free part of the snapshot vector will be shown to lie in a subspace of dimension n. Such a reduction of dimension of the signal subspace (*i.e.* the increase in the dimension of the noise subspace for a given m) will lead us to estimates of improved accuracy, particularly for small m. As a next step we propose an ESPRIT like algorithm, called R-Esprit, to estimate the frequencies in a non-iterative and computationally efficient way. Further, we carry out an asymptotic analysis of both the ESPRIT and R-Esprit algorithms for real-valued sine waves. R-Esprit has a natural extension to complex-valued sinusoids too, since any complex-valued signal can be decomposed into two real-valued signals corresponding to its real and imaginary parts. After extending R-Esprit to complex-valued sine waves, we present an asymptotic analysis of the extended R-Esprit algorithm for the complex-valued case. From a simulation study it is shown that the R-Esprit algorithm yields accurate estimates at smaller values of m, which gives significant computational advantage. R-Esprit performs significantly better than ESPRIT for closely spaced sine waves. The asymptotic performance of R-Esprit outperforms the ESPRIT algorithm.

The paper is structured as follows. In the next section, we review the conventional complex-valued data model and propose the new compact data model for real-valued sine waves. In Section 3, we propose the R-Esprit algorithm and carry out an analysis of R-Esprit and ESPRIT for real-valued sinusoids. We extend R-Esprit to complex-valued sinusoids in Section 4. In Section 5, we illustrate the performance of the algorithms using numerical simulations. The results are analyzed in Section 6 and we provide conclusions in Section 7.

2 Data Models

2.1 The complex-valued signal model

In this section, we review the conventional complex-valued low rank data model (CDM) [9, 23]. This model was primarily introduced in the context of subspace based estimation of the parameters of complex-valued sinusoids observed in noise, and hence CDM constitutes the basic building block of all such algorithms (see also [5] for a variant of basic CDM). Since any real-valued sinusoid can be expressed as a superposition of two complex-valued sinusoids, CDM can readily be extended to real valued sinusoids. To spell out the details, consider (1.1), we see that

$$y(t) = \sum_{k=1}^{n} \frac{a_k}{2} \left\{ \mathrm{e}^{\mathrm{i}\,(\omega_k t + \phi_k)} + \mathrm{e}^{-\mathrm{i}\,(\omega_k t + \phi_k)} \right\} + \epsilon(t) \tag{2.1}$$

where we have expressed each of the n real-valued sinusoids as a sum of two complex-valued sinusoids. Introduce the $m|1$ ($m > 2n$) dimensional snapshot vector

$$\mathbf{y}_c(t) := \left[\begin{array}{ccc} y(t) & \dots & y(t+m-1) \end{array}\right]^\top, \tag{2.2}$$

where $^\top$ is used to denote the transpose. It is readily verified that

$$\mathbf{y}_c(t) = \mathbf{A}_c\mathbf{s}_c(t) + \boldsymbol{\epsilon}_c(t), \tag{2.3}$$

where $\mathbf{A}_c$ is a $m|2n$ is matrix and $\mathbf{s}_c(t)$ is a $2n|1$ vector given by

$$\begin{aligned}
\mathbf{A}_c &= \left[\begin{array}{cccccc} 1 & \dots & 1 & 1 & \dots & 1 \\ \mathrm{e}^{\mathrm{i}\,\omega_1} & \dots & \mathrm{e}^{\mathrm{i}\,\omega_n} & \mathrm{e}^{-\mathrm{i}\,\omega_1} & \dots & \mathrm{e}^{-\mathrm{i}\,\omega_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{e}^{\mathrm{i}\,(m-1)\omega_1} & \dots & \mathrm{e}^{\mathrm{i}\,(m-1)\omega_n} & \mathrm{e}^{-\mathrm{i}\,(m-1)\omega_1} & \dots & \mathrm{e}^{-\mathrm{i}\,(m-1)\omega_n} \end{array}\right], && (2.4) \\
\mathbf{s}_c(t) &= \left[\begin{array}{cc} \mathbf{s}_0^\top(t) & \mathbf{s}_0^*(t) \end{array}\right]^\top, && (2.5) \\
\mathbf{s}_0(t) &= \frac{1}{2}\left[\begin{array}{ccc} a_1\,\mathrm{e}^{\mathrm{i}\,(\omega_1 t+\phi_1)} & \dots & a_n\,\mathrm{e}^{\mathrm{i}\,(\omega_n t+\phi_n)} \end{array}\right]^\top. && (2.6)
\end{aligned}$$

Note that $\mathbf{s}_0^*(t)$ denotes the conjugate transpose of $\mathbf{s}_0(t)$. The noise snapshot vector $\boldsymbol{\epsilon}_c(t)$ of dimension $(m|1)$ is given by

$$\boldsymbol{\epsilon}_c(t) = \left[\begin{array}{ccc} \epsilon(t) & \dots & \epsilon(t+m-1) \end{array}\right]^\top. \tag{2.7}$$

It is now clear from (2.3) that the noise free part of the snapshot vector $\mathbf{y}_c(t)$ lies in the column space of $\mathbf{A}_c$ for all t. Thus (2.3) is a low rank signal model, which is exploited in all super-resolution subspace based approaches for frequency estimation. The dimension of the signal subspace in CDM equals the number of complex-valued sinusoids. Hence for the real-valued sinusoids the dimension of the signal subspace is twice the number of frequencies. It can also be observed that $\mathbf{A}_c$ and $\mathbf{s}_c(t)$ are complex-valued but the product $\mathbf{A}_c\,\mathbf{s}_c(t)$ is real-valued for all t. Also $\boldsymbol{\epsilon}_c(t)$ is zero mean real-valued random variable with covariance matrix $\sigma_1^2\mathbf{I}_m$, where $\mathbf{I}_m$ denotes the identity matrix of dimension m. Denote the correlation matrix

$$\mathbf{P}_c(\tau) := \mathbf{E}\mathbf{s}_c(t+\tau)\,\mathbf{s}_c^*(t). \tag{2.8}$$

With a slight abuse of notation, we use $\mathbf{P}_c := \mathbf{P}_c(0)$. It is easy to see that

$$\mathbf{R}_c := \mathbf{E}\,\mathbf{y}_c(t)\mathbf{y}_c^\top(t) = \mathbf{A}_c\mathbf{P}_c\mathbf{A}_c^* + \sigma_1^2\mathbf{I}_m. \tag{2.9}$$

It is straightforward to estimate $\mathbf{R}_c$ consistently from the data. Once the correlation matrix $\mathbf{R}_c$ is obtained one has several different efficient possibilities to process $\mathbf{R}_c$ further to obtain the consistent estimates of the frequencies $\{\omega_i\}_{i=1}^n$. We shall come back to this point later.

2.2 A compact subspace representation

In this section, our goal is to derive a more compact subspace representation of the data. By the word 'compact', we mean that we seek for a signal model (or an alternative snapshot vector) where we can assure that the noise-free part of the snapshot vector lies in a subspace of smaller dimension than $2n$. The primary motivation comes from the CDM for complex-valued signal. We see that the dimension of the low rank signal subspace for n complex-valued sine waves is n. Similarly, it is reasonable to expect that it is possible to construct a snapshot vector also for real-valued sine waves, such that the dimension of the signal subspace equals the number of frequencies. The reason why it is beneficial to seek for such a representation is the associated improvement in the estimation accuracy for the same dimension m of the snapshot vector. It is known that the estimation accuracy of a subspace based estimation method can be improved significantly if the dimension of the subspace representation can be reduced, by proper incorporation of the available information. First, let us examine the possibility of existence of such a compact representation. In (2.4) we can see that the first n columns of $\mathbf{A}_c$ are precisely the complex conjugate of the next n columns. This means that the first n columns of $\mathbf{A}_c$ are dependent on the next n columns. However, this dependence is nonlinear, and cannot be exploited by any simple method. One can of course use nonlinear techniques but that would lead to a heavier computational load. In fact, this was the major disadvantage encountered by the authors of [22]. Here we attempt another approach. The idea is to modify the snapshot vector $\mathbf{y}_c(t)$ so that the nonlinear dependence between the individual columns of $\mathbf{A}_c$ is can be converted to a linear dependence in the new model. In that aim let us introduce

$$\begin{aligned} \mathbf{y}_b(t) &:= \left[\begin{array}{ccc} y(t-1) & \dots & y(t-m) \end{array} \right]^\top && (2.10) \\ \mathbf{y}_r(t) &:= \frac{1}{2}\left\{ \mathbf{y}_c(t) + \mathbf{y}_b(t) \right\}. && (2.11) \end{aligned}$$

From the above definitions it is easy to see that the jth element of $\mathbf{y}_r(t)$ is given by

$$\begin{aligned} \left[\mathbf{y}_r(t)\right]_j &= \frac{1}{2}\sum_{k=1}^{n} a_k \left[\cos\left\{\omega_k(t+j-1)+\phi_k\right\} + \cos\left\{\omega_k(t-j)+\phi_k\right\}\right] \\ &\quad + \frac{1}{2}\left\{\epsilon(t+j-1)+\epsilon(t-j)\right\} \\ &= \sum_{k=1}^{n} a_k \cos\left\{\omega_k\left(t-\frac{1}{2}\right)+\phi_k\right\}\cos\left\{\omega_k\left(j-\frac{1}{2}\right)\right\} \\ &\quad + \frac{1}{2}\left\{\epsilon(t+j-1)+\epsilon(t-j)\right\}. \qquad (2.12) \end{aligned}$$

Thus from (2.12) we have

$$\mathbf{y}_r(t) = \mathbf{A}_r \mathbf{s}_r(t) + \boldsymbol{\epsilon}_r(t), \qquad (2.13)$$

where $\mathbf{A}_r$ is an $(m|n)$ matrix given by

$$\mathbf{A}_r = \begin{bmatrix} \cos\left(\frac{\omega_1}{2}\right) & \dots & \cos\left(\frac{\omega_n}{2}\right) \\ \cos\left(\frac{3\omega_1}{2}\right) & \dots & \cos\left(\frac{3\omega_n}{2}\right) \\ \vdots & \ddots & \vdots \\ \cos\left\{\left(m-\frac{1}{2}\right)\omega_1\right\} & \dots & \cos\left\{\left(m-\frac{1}{2}\right)\omega_n\right\} \end{bmatrix} \qquad (2.14)$$

and $\mathbf{s}_r(t)$ is an $(n|1)$ vector given by

$$\mathbf{s}_r(t) = \begin{bmatrix} a_1 \cos\left\{\omega_1 t + \phi_1^+\right\} \\ \vdots \\ a_n \cos\left\{\omega_n t + \phi_n^+\right\} \end{bmatrix}, \tag{2.15}$$

where $\phi_k^+ = \phi_k - \frac{1}{2}\omega_k$ for $1 \le k \le n$. The noise snapshot vector $\boldsymbol{\epsilon}_r(t)$ in the modified model is given by

$$\boldsymbol{\epsilon}_r(t) = \frac{1}{2}\left\{\boldsymbol{\epsilon}_c(t) + \boldsymbol{\epsilon}_b(t)\right\}, \tag{2.16}$$

where analogous to the data vector

$$\boldsymbol{\epsilon}_b(t) \quad = \quad \begin{bmatrix} \epsilon(t-1) & \dots & \epsilon(t-m) \end{bmatrix}^\top. \tag{2.17}$$

Note here that in contrast to the complex-valued data model, the noise-free part of $\mathbf{y}_r(t)$ lies in an n dimensional subspace. Hence, we have now a more compact model, which is real-valued. We shall refer to this model as RDM. We introduce

$$\mathbf{P}_r(\tau) := \mathbf{E}\,\mathbf{s}_r(t+\tau)\mathbf{s}_r^\top(t), \tag{2.18}$$

also note that $\boldsymbol{\epsilon}_c(t)$ and $\boldsymbol{\epsilon}_b(t)$ are mutually independent random vectors for all t, which would mean $\mathbf{E}\boldsymbol{\epsilon}_r(t)\boldsymbol{\epsilon}_r^\top(t) = \frac{\sigma_1^2}{2}\mathbf{I}_m$. Using our usual notation $\mathbf{P}_r := \mathbf{P}_r(0)$, we get

$$\mathbf{R}_r := \mathbf{E}\,\mathbf{y}_r(t)\mathbf{y}_r^\top(t) = \mathbf{A}_r\mathbf{P}_r\mathbf{A}_r^* + \frac{\sigma_1^2}{2}\mathbf{I}_m. \tag{2.19}$$

As in case of the complex-valued model CDM, here also we can estimate $\mathbf{R}_r$ from the data. Next the estimate of $\mathbf{R}_r$ can be used in different manner to arrive at the estimate of the frequencies. In contrast to the CDM the matrices involved in RDM are all real-valued. On the other hand $\mathbf{A}_c$ is a Vandermonde matrix, but $\mathbf{A}_r$ has a rather complicated structure. The Vandermonde structure of $\mathbf{A}_c$ is exploited in the fast algorithms like ESPRIT [18] and MODE[2] [2], see also [10], which would not be as straightforward while working with $\mathbf{A}_r$ in the compact model.

3 Modified ESPRIT

ESPRIT – Estimation of signal parameters via rotational invariance techniques [16, 18] is one of the most popular subspace based techniques for sinusoidal frequency estimation (although the algorithm was first proposed in context of sensor array signal processing). In this section, we first briefly review the general notion of ESPRIT algorithm. Then we extend these ideas to the real-valued sinusoid model RDM. Consider a general low rank signal model

$$\mathbf{y}(t) = \mathbf{A}\mathbf{s}(t) + \boldsymbol{\epsilon}(t) \tag{3.1}$$

[2]MODE-Method of direction estimation was introduced in [24] for DOA estimation. The idea was extended in [2] to temporal spectrum estimation.

where $\mathbf{A}$ is a $m|p$ full column rank matrix, where $p < m$. In our context, $p = 2n$ for the complex-valued data model, and the real-valued compact model has $p = n$. Assuming $\mathbf{E}\mathbf{s}(t)\mathbf{s}^*(t) = \mathbf{P}$ and $\mathbf{E}\,\boldsymbol{\epsilon}(t)\boldsymbol{\epsilon}^*(t) = \sigma^2\mathbf{I}_m$. we have as before

$$\mathbf{R} := \mathbf{E}\mathbf{y}(t)\mathbf{y}^*(t) = \mathbf{A}\mathbf{P}\mathbf{A}^* + \sigma^2\mathbf{I}_m. \tag{3.2}$$

The key step in ESPRIT algorithm is to find out two $p_1|p$ ($p \le p_1 < m$) transformation matrices $\mathbf{T}^{(1)}$ and $\mathbf{T}^{(2)}$ such that

$$\mathbf{T}^{(2)}\mathbf{A} = \mathbf{T}^{(1)}\mathbf{A}\mathcal{D} \tag{3.3}$$

where $\mathcal{D}$ is a diagonal matrix and a simple invertible function of the parameter to be estimated. For example, for complex valued data model, by defining

$$\mathbf{T}_c^{(1)} := \begin{bmatrix} \mathbf{I}_{m-1} & \mathbf{0}_{m\times 1} \end{bmatrix}^\top, \quad \mathbf{T}_c^{(2)} := \begin{bmatrix} \mathbf{0}_{m\times 1} & \mathbf{I}_{m-1} \end{bmatrix}^\top \tag{3.4}$$

it follows that

$$\mathbf{T}_c^{(2)}\mathbf{A}_c = \mathbf{T}_c^{(1)}\mathbf{A}_c\mathcal{D}_c \tag{3.5}$$

where

$$\mathcal{D}_c = \operatorname{diag}\left(\mathrm{e}^{\mathrm{i}\omega_1}, \dots, \mathrm{e}^{\mathrm{i}\omega_n}, \mathrm{e}^{-\mathrm{i}\omega_1}, \dots, \mathrm{e}^{-\mathrm{i}\omega_n}\right) \tag{3.6}$$

In this case it has been quite straightforward to find out $\mathbf{T}_c^{(1)}$ and $\mathbf{T}_c^{(2)}$ because of the Vandermonde structure of $\mathbf{A}_c$. To see how (3.3) can be used consider the singular value decomposition

$$\mathbf{R} = \mathbf{S}\boldsymbol{\Lambda}\mathbf{S}^* + \mathbf{G}\boldsymbol{\Sigma}\mathbf{G}^* \tag{3.7}$$

where $\boldsymbol{\Lambda}$ is $p|p$ a diagonal matrix having the p largest singular values of $\mathbf{R}$ as its diagonal entries and $m|p$ matrix $\mathbf{S}$ is composed of the corresponding left singular vectors. Similarly, $\boldsymbol{\Sigma}$ is a diagonal matrix having the remaining $m-p$ singular values of $\mathbf{R}$ as its diagonal entries. Columns of the $m|m-p$ matrix $\mathbf{G}$ constitute the corresponding left singular vectors, *i.e*

$$\mathbf{S}^*\mathbf{S} = \mathbf{I}_p, \quad \mathbf{G}^*\mathbf{G} = \mathbf{I}_{m-p}, \quad \mathbf{S}^*\mathbf{G} = \mathbf{0}_{p\times(m-p)}. \tag{3.8}$$

In fact from (3.2) it can be shown that

$$\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_{m-p}, \tag{3.9}$$

and $\mathbf{G}$ is not unique. Any orthogonal transformation of $\mathbf{G}$ can replace $\mathbf{G}$ in (3.7). Though $\mathbf{G}$ is non unique, the corresponding singular subspace of $\mathbf{R}$ (which is the orthogonal complement of the signal subspace spanned by $\mathbf{S}$ and commonly known as noise subspace) is unique and so is the projection operator $\mathbf{G}\mathbf{G}^*$ onto the noise subspace. Now from (3.2), (3.7) and (3.8) it follows that

$$\mathbf{R}\mathbf{S} = \mathbf{S}\boldsymbol{\Lambda} = \mathbf{A}\mathbf{P}\mathbf{A}^*\mathbf{S} + \sigma^2\mathbf{S} \quad \Rightarrow \quad \mathbf{S}\left\{\boldsymbol{\Lambda} - \sigma^2\mathbf{I}_p\right\} = \mathbf{A}\mathbf{P}\mathbf{A}^*\mathbf{S} \tag{3.10}$$

Introduce the $p|p$ nonsingular matrix (it can be shown that each nonzero element of $\boldsymbol{\Lambda}$ is strictly greater than σ^2)

$$\mathbf{C} = \mathbf{P}\mathbf{A}^*\mathbf{S}\left\{\boldsymbol{\Lambda} - \sigma^2\mathbf{I}_p\right\}^{-1}, \quad \Rightarrow \quad \mathbf{S} = \mathbf{A}\mathbf{C}, \tag{3.11}$$

from (3.10). Hence, $\mathbf{S}$ forms an orthonormal basis of the column space of $\mathbf{A}$. Now from (3.11) and (3.3) it follows that

$$\mathbf{T}^{(2)}\mathbf{S} = \mathbf{T}^{(2)}\mathbf{AC} = \mathbf{T}^{(1)}\mathbf{A}\mathcal{D}\mathbf{C} = \mathbf{T}^{(1)}\mathbf{SC}^{-1}\mathcal{D}\mathbf{C} = \mathbf{T}^{(1)}\mathbf{S}\mathbf{\Phi} \tag{3.12}$$

where $\mathbf{\Phi} := \mathbf{C}^{-1}\mathcal{D}\mathbf{C}$. It is straightforward to see that the eigenvalues of $\mathbf{\Phi}$ are precisely the diagonal entries of $\mathcal{D}$. This observation leads us to ESPRIT algorithm, where an estimate $\widehat{\mathbf{S}}$ of $\mathbf{S}$ is obtained from the singular value decomposition of $\widehat{\mathbf{R}}$, an estimate of $\mathbf{R}$. Next, an estimate $\widehat{\mathbf{\Phi}}$ is obtained by solving (3.12) for $\mathbf{\Phi}$ with $\mathbf{S}$ replaced by $\widehat{\mathbf{S}}$. An eigen-decomposition of $\widehat{\mathbf{\Phi}}$ would then lead us to the frequency estimates for complex-valued signal model CDM, see (3.6). We also mention here that the estimate $\widehat{\mathbf{S}}$ deviates from $\mathbf{S}$ to some extent. Hence, while solving (3.12) for $\mathbf{\Phi}$, an exact solution may not be possible. It would be required to solve the equation in least squares (or total least squares) sense [3, 13, 20, 6].

From the above discussion it is clear that a modified ESPRIT estimator of the frequencies can be derived for the compact real valued data model RDM described in Section 2.2 provided two transformations $\mathbf{T}^{(1)}$ and $\mathbf{T}^{(2)}$ are derived for $\mathbf{A}_r$ such that (3.3) is satisfied. In that aim first we note that

$$\frac{1}{2}\left[\cos\left\{\left(j-\frac{3}{2}\right)\omega\right\} + \cos\left\{\left(j+\frac{1}{2}\right)\omega\right\}\right] = \cos\left\{\left(j-\frac{1}{2}\right)\omega\right\}\cos(\omega) \tag{3.13}$$

The special structure of $\mathbf{A}_r$ and (3.13) will be exploited in what follows next. Let us define $(m-2)|m$ Toeplitz matrices

$$\mathbf{T}_r^{(1)} = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{T}_r^{(2)} = \frac{1}{2}\begin{bmatrix} 1 & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 & 1 \end{bmatrix}. \tag{3.14}$$

Then it is easy to see using (3.13) and (2.14) that

$$\mathbf{T}_r^{(2)}\mathbf{A}_r = \mathbf{T}_r^{(1)}\mathbf{A}_r\mathcal{D}_r, \tag{3.15}$$

where

$$\mathcal{D}_r = \operatorname{diag}\left(\cos(\omega_1), \dots, \cos(\omega_n)\right). \tag{3.16}$$

With this observation it is now straightforward to derive an ESPRIT algorithm for the compact real-valued data model. Before spelling out the details, the following notation is introduced. Henceforth we shall add a subscript 'r' to denote the matrices concerning the compact real-valued data model RDM. For example, $\mathbf{\Phi}_r$ will denote the $\mathbf{\Phi}$ matrix for the real-valued data model. Similarly the matrices with subscript 'c' will denote the corresponding matrices for the complex-valued data model CDM. Using this notational convention, it is first required to estimate $\widehat{\mathbf{S}}_r$, which will then be used in (3.12) to solve for $\widehat{\mathbf{\Phi}}_r$. A diagonalization of $\widehat{\mathbf{\Phi}}_r$ will then lead us to $\widehat{\mathcal{D}}_r$ and finally an *arccos* operation on the diagonal elements of $\widehat{\mathcal{D}}_r$ will give the frequency estimates. We shall refer to this algorithm for real-valued sinusoids as R-Esprit algorithm[3]. The following asymptotic accuracy result holds for the R-Esprit algorithm.

Theorem 1: Denote the kth column of $\mathbf{I}_n$ by $\mathbf{e}_k$ and define

$$\begin{aligned}
\boldsymbol{\xi}_k^* &:= \mathbf{e}_k^* \left\{\mathbf{A}_r^* \mathbf{T}_r^{(1)*} \mathbf{T}_r^{(1)} \mathbf{A}_r\right\}^{-1} \mathbf{A}_r^* \mathbf{T}_r^{(1)*} \left\{\mathbf{T}_r^{(2)} - \mathbf{T}_r^{(1)} \cos(\omega_k)\right\} && (3.17)\\
\mathbf{B}_r &:= \mathbf{A}_r \left\{\mathbf{A}_r^* \mathbf{A}_r\right\}^{-1} = \left[\begin{array}{ccc} \boldsymbol{\nu}_1 & \ldots & \boldsymbol{\nu}_n \end{array}\right], && (3.18)\\
\mathbf{Q}_r(\tau) &:= \frac{1}{\sigma_1^2} \mathbf{E} \boldsymbol{\epsilon}_r(t+\tau) \boldsymbol{\epsilon}_r^*(t), && (3.19)
\end{aligned}$$

($\mathbf{Q}_r(\tau)$ is the normalized covariance matrix of $\boldsymbol{\epsilon}_r(t)$ if $\epsilon(t)$ is unit-variance). Denote the estimation error in ω_k by $\tilde{\omega}_k$ for $1 \leq k \leq n$. Then if the noise sequence $\epsilon(t)$ is Gaussian, the large sample ($N \gg m$) asymptotic second order statistics of the estimation error is given by

$$\begin{aligned}
\mathbf{E} \tilde{\omega}_k \tilde{\omega}_l &= \frac{4\sigma_1^4}{N-2m-1} \frac{1}{a_k^2 a_l^2 \sin(\omega_k) \sin(\omega_l)} \\
&\times \sum_{\tau=-2m+1}^{2m-1} \left\{\boldsymbol{\xi}_k^\top \mathbf{Q}_r(\tau) \boldsymbol{\xi}_l \; \boldsymbol{\nu}_k^\top \mathbf{Q}_r(\tau) \mathbf{b}_l + \boldsymbol{\xi}_k^\top \mathbf{Q}_r(\tau) \boldsymbol{\nu}_l \; \boldsymbol{\nu}_k^\top \mathbf{Q}_r(\tau) \boldsymbol{\xi}_l\right\} \quad (3.20)
\end{aligned}$$

Proof: For an expression of $\mathbf{Q}_r(\tau)$ see (B.7) Appendix B. The proof the theorem follows the same outline as the analysis of ESPRIT algorithm carried out in [25]. We use two preliminary results proved in the form of Lemma 1 and Lemma 2 in Appendix A. We also need the following expression, which is quite straightforward to prove [15]:

$$\mathbf{P}_r(\tau) := \mathbf{E} \mathbf{s}_r(t+\tau)\, \mathbf{s}_r^\top(t) = \frac{1}{2} \mathrm{diag}\left(a_1^2 \cos(\omega_1 \tau), \ldots, a_n^2 \cos(\omega_n \tau)\right). \quad (3.21)$$

Now let us refer to the general subspace model (3.1). Throughout this paper we use $\{\rho_k\}_{k=1}^p$ to denote the eigenvalues of $\mathcal{D}$ in (3.3). Since this proof is related to the R-Esprit algorithm, here we are concerned to the real-valued data model RDM, for which $\rho_k = \cos(\omega_k)$. Let $\hat{\omega}_k$ and $\hat{\rho}_k$ be the R-Esprit estimates of ρ_k and ω_k respectively. Denote $\tilde{\rho}_k = \hat{\rho}_k - \rho_k$ and $\tilde{\omega}_k = \hat{\omega}_k - \omega_k$. Similarly for any relevant quantity x we use $\hat{x}$ to denote its estimate and $\tilde{x}$ to denote the estimation error. Next we use the Lemma 1 in Appendix A. Note that the general definitions (A.1) and (A.3) in Appendix A get modified accordingly to (3.17) and (3.18), respectively, for RDM. Since $\mathbf{P}_r$ is diagonal, by using the corollary (A.14) to the Lemma 1 in Appendix A and (3.21) we get

$$\tilde{\omega}_k = -\tilde{\rho}_k / \sin(\omega_k) = -\frac{2\boldsymbol{\xi}_k^* \widehat{\mathbf{R}}_r \boldsymbol{\nu}_k}{a_k^2 \sin(\omega_k)}, \quad (3.22)$$

where we have used first order approximation for large N. Recall that for the real-valued data model all relevant matrices are real-valued. We shall use this fact frequently in this proof without mentioning it explicitly. Next we apply

[3]Note that $\mathcal{D}_c$ is unitary, hence multiplication by $\mathcal{D}_c$ corresponds to a rotation in $\mathbb{C}^{2n}$. But $\mathcal{D}_r$ is not an unitary matrix. Therefore R-Esprit no longer uses 'rotation invariance'. However, we prefer to name it R-Esprit because of its close similarity with the ESPRIT algorithm.

(A.20), Lemma 2 in Appendix A. In terms of the notations in Lemma 2, here $M = N - 2m + 1$ and $\tau_{\max} = 2m - 1$, see (B.7), Appendix B. We have

$$\begin{aligned}\mathbf{E}\tilde{\omega}_k\tilde{\omega}_l &= \frac{4\sigma_1^4(N-2m-1)^{-1}}{a_k^2 a_l^2 \sin(\omega_k)\sin(\omega_l)} \sum_{\tau=-2m+1}^{2m-1} \{\boldsymbol{\xi}_k^\top \mathbf{Q}_r(\tau)\boldsymbol{\xi}_l \; \boldsymbol{\nu}_k^\top \mathbf{Q}_r(\tau)\boldsymbol{\nu}_l \\ &\quad + \boldsymbol{\xi}_k^\top \mathbf{Q}_r(\tau)\boldsymbol{\nu}_l \; \boldsymbol{\nu}_k^\top \mathbf{Q}_r(\tau)\boldsymbol{\xi}_l + \boldsymbol{\xi}_k^\top \mathbf{Q}_r(\tau)\boldsymbol{\xi}_l \; \mathbf{e}_k^\top \mathbf{P}_r(\tau)\mathbf{e}_l\}\,. \end{aligned} \tag{3.23}$$

Note that $\mathbf{P}_r(\tau)$ is a diagonal matrix. Hence using (3.21) we get for the last term within brackets in the right hand side of (3.23)

$$\sum_{\tau=-2m+1}^{2m-1} \boldsymbol{\xi}_k^\top \mathbf{Q}_r(\tau)\boldsymbol{\xi}_l \; \mathbf{e}_k^\top \mathbf{P}_r^\top(\tau)\mathbf{e}_l = a_k^2 \boldsymbol{\xi}_k^\top \left\{ \sum_{\tau=-2m+1}^{2m-1} \mathbf{Q}_r(\tau)\cos(\omega_k \tau) \right\} \boldsymbol{\xi}_l \, \delta_{k,l} = 0, \tag{3.24}$$

where we have used (B.10) in Appendix B. From the remaining terms of (3.23) it is straightforward to get (3.20). ■

It is interesting to compare the performance of the R-Esprit algorithm to that of the conventional ESPRIT algorithm. We will present also an asymptotic accuracy result for ESPRIT algorithm. To the best of the knowledge of the authors, this result has never been proved in any published paper. We mention here that the asymptotic analysis of ESPRIT algorithm carried out in [25] assumes the standard framework where the data is complex-valued, *i.e.* not only the sinusoids are complex-valued but also the additive noise is complex-valued (which is natural to expect). On the contrary, the analysis we carry out here in the form of Theorem 2 assumes the data are real-valued as in (1.1) and ESPRIT algorithm is applied to the data pretending the data consist of a double number of complex-valued sinusoids. Therefore, unlike the case in Theorem 1 the quantities involved in the following theorem will be complex-valued in general, although the underlying data and the noise sequence are real-valued.

Theorem 2: Denote the kth column of $\mathbf{I}_{2n}$ by $\mathbf{e}_k$ and define

$$\boldsymbol{\zeta}_k^* := \mathbf{e}_k^\top \left\{\mathbf{A}_c^* \mathbf{T}_c^{(1)*}\mathbf{T}_c^{(1)}\mathbf{A}_c\right\}^{-1} \mathbf{A}_c^* \mathbf{T}_c^{(1)*}\left\{\mathbf{T}_c^{(2)} - \mathbf{T}_c^{(1)} \mathrm{e}^{\mathrm{i}\omega_k}\right\} \tag{3.25}$$

$$\mathbf{B}_c := \mathbf{A}_c\left\{\mathbf{A}_c^*\mathbf{A}_c\right\}^{-1} = \left[\begin{array}{ccc} \boldsymbol{\mu}_1 & \ldots & \boldsymbol{\mu}_{2n}\end{array}\right], \tag{3.26}$$

$$\mathbf{Q}_c(\tau) := \frac{1}{\sigma_1^2}\mathbf{E}\boldsymbol{\epsilon}_c(t+\tau)\boldsymbol{\epsilon}_c^\top(t). \tag{3.27}$$

Then if $\epsilon(t)$ is Gaussian, the asymptotic second order statistics of the estimation error $\tilde{\omega}_k$ of the ESPRIT estimate of ω_k (the notation $\tilde{\omega}_k$ should not be confused with the same in Theorem 1) is given by

$$\begin{aligned}\mathbf{E}\tilde{\omega}_k\tilde{\omega}_l &= \frac{8\sigma_1^4}{N+m-1}\;\frac{1}{a_k^2 a_l^2}\times \\ &\quad \sum_{\tau=-m+1}^{m-1} \mathrm{Re}\left[\mathrm{e}^{\mathrm{i}(\omega_k-\omega_l)}\left\{\boldsymbol{\zeta}_k^*\mathbf{Q}_c(\tau)\boldsymbol{\zeta}_l\,\boldsymbol{\mu}_k^\top\mathbf{Q}_c(\tau)\bar{\boldsymbol{\mu}}_l + \boldsymbol{\zeta}_k^*\mathbf{Q}_c(\tau)\bar{\boldsymbol{\mu}}_l\,\boldsymbol{\mu}_k^\top\mathbf{Q}_c(\tau)\boldsymbol{\zeta}_l\right\}\right. \\ &\quad \left. -\,\mathrm{e}^{\mathrm{i}(\omega_k+\omega_l)}\left\{\boldsymbol{\zeta}_k^*\mathbf{Q}_c(\tau)\bar{\boldsymbol{\zeta}}_l\,\boldsymbol{\mu}_k^\top\mathbf{Q}_c(\tau)\boldsymbol{\mu}_l + \boldsymbol{\zeta}_k^*\mathbf{Q}_c(\tau)\boldsymbol{\mu}_l\,\boldsymbol{\mu}_k^\top\mathbf{Q}_c(\tau)\bar{\boldsymbol{\zeta}}_l\right\}\right]\end{aligned} \tag{3.28}$$

Proof: The proof of this theorem will follow essentially the same steps as the previous one. For an expression of $\mathbf{Q}_c(\tau)$, see (B.4). Let us define

$$\mathcal{P}(\tau) := \frac{1}{4}\mathrm{diag}\left(a_1^2 \mathrm{e}^{\mathrm{i}\omega_1 \tau}, \ldots, a_n^2 \mathrm{e}^{\mathrm{i}\omega_n \tau}\right). \tag{3.29}$$

Then it is easy to show that

$$\mathbf{P}_c(\tau) := \mathbf{E}\mathbf{s}_c(t+\tau)\, \mathbf{s}_c^*(t) = \begin{bmatrix} \mathcal{P}(\tau) & \mathbf{0}_{n\times n} \\ \mathbf{0}_{n\times n} & \overline{\mathcal{P}}(\tau) \end{bmatrix} \tag{3.30}$$

where we have followed our usual notation to denote the complex conjugate of $\mathcal{P}(\tau)$ by $\overline{\mathcal{P}}(\tau)$. We introduce

$$\mathbf{P}_c^+(\tau) := \mathbf{E}\overline{\mathbf{s}}_c(t+\tau)\, \mathbf{s}_c^*(t) = \begin{bmatrix} \mathbf{0}_{n\times n} & \overline{\mathcal{P}}(\tau) \\ \mathcal{P}(\tau) & \mathbf{0}_{n\times n} \end{bmatrix}, \tag{3.31}$$

which is obvious from (3.30). Next we relate the perturbation $\tilde{\omega}_k$ to the perturbation $\tilde{\rho}_k$ in ρ_k (recall that for the complex valued data model $\{\rho_k\}_{k=1}^{2n}$ denote the eigenvalues of $\mathbf{\Phi}_c$). Now for $1 \le k \le n$

$$\rho_k = \mathrm{e}^{\mathrm{i}\omega_k} \quad \Rightarrow \quad \omega_k = \mathrm{Im}\left\{\log(\rho_k)\right\} \quad \Rightarrow \quad \tilde{\omega}_k = \mathrm{Im}\left\{\tilde{\rho}_k/\rho_k\right\} \tag{3.32}$$

by first order approximations. Using an identity in the algebra of complex numbers we have

$$\tilde{\omega}_k\tilde{\omega}_l = \frac{1}{2}\mathrm{Re}\left\{\frac{\tilde{\rho}_k\tilde{\rho}_l^*}{e^{\mathrm{i}(\omega_l-\omega_k)}} - \frac{\tilde{\rho}_k\tilde{\rho}_l}{e^{-\mathrm{i}(\omega_l+\omega_k)}}\right\}. \tag{3.33}$$

We will separately consider each of the terms in the right hand side of (3.33). Note that $\mathbf{P}_c$ is a diagonal matrix. Hence we can apply corollary (A.14) to the Lemma 1 in Appendix A to evaluate $\tilde{\rho}_k$. Note that the general definitions (A.1) and (A.3) in Appendix A get modified accordingly to (3.25) and (3.26), respectively, for CDM. Since $\mathbf{P}_c$ is diagonal, by using the corollary (A.14) to the Lemma 1 in Appendix A and (3.30) we get

$$\tilde{\rho}_k = \frac{4}{a_k^2}\boldsymbol{\zeta}_k^*\widehat{\mathbf{R}}_c\boldsymbol{\mu}_k. \tag{3.34}$$

Next we apply (A.20), Lemma 2 in Appendix A. In terms of the notations in Lemma 2, here $M = N - m + 1$ and $\tau_{\max} = m - 1$, see (B.4), Appendix B. We have

$$\begin{aligned} \mathbf{E}\tilde{\rho}_k\tilde{\rho}_l^* \;=\; & \frac{16\sigma_1^4(N+m-1)^{-1}}{a_k^2 a_l^2}\sum_{\tau=-m+1}^{m-1}\left\{\boldsymbol{\zeta}_k^*\mathbf{Q}_c(\tau)\boldsymbol{\zeta}_l\,\boldsymbol{\mu}_k^\top\mathbf{Q}_c(\tau)\bar{\boldsymbol{\mu}}_l\right. \\ & \left. + \boldsymbol{\zeta}_k^*\mathbf{Q}_c(\tau)\bar{\boldsymbol{\mu}}_l\,\boldsymbol{\mu}_k^\top\mathbf{Q}_c(\tau)\boldsymbol{\zeta}_l + \boldsymbol{\zeta}_k^*\mathbf{Q}_c(\tau)\boldsymbol{\zeta}_l\,\mathbf{e}_k^*\mathbf{P}_c^\top(\tau)\mathbf{e}_l\right\}, \end{aligned} \tag{3.35}$$

where we use $\bar{\boldsymbol{\mu}}$ to denote the complex conjugate of $\boldsymbol{\mu}$. The contribution of the last term in the right hand side of (3.35) can be shown to be identically zero. To see this refer to (3.29) and (3.30). We see that

$$\sum_{\tau=-m+1}^{m-1}\boldsymbol{\zeta}_k^*\mathbf{Q}_c(\tau)\boldsymbol{\zeta}_l\;\mathbf{e}_k^*\mathbf{P}_c^\top(\tau)\mathbf{e}_l = a_k^2\boldsymbol{\zeta}_k^*\left\{\sum_{\tau=-m+1}^{m-1}\mathbf{Q}_c(\tau)\mathrm{e}^{\mathrm{i}\omega_k\tau}\right\}\boldsymbol{\zeta}_l\,\delta_{k,l} = 0 \tag{3.36}$$

by (B.12) in Appendix B. Next consider the remaining term in (3.33). We use (A.14) again along with (A.19), Lemma 2. We have

$$\begin{aligned}\mathbf{E}\tilde{\rho}_k\tilde{\rho}_l &= \frac{16\sigma_1^4(N+m-1)^{-1}}{a_k^2a_l^2}\sum_{\tau=-m+1}^{m-1}\left\{\boldsymbol{\zeta}_k^*\mathbf{Q}_c(\tau)\bar{\boldsymbol{\zeta}}_l\,\boldsymbol{\mu}_k^\top\mathbf{Q}_c(\tau)\boldsymbol{\mu}_l\right.\\ &\left.+\boldsymbol{\zeta}_k^*\mathbf{Q}_c(\tau)\boldsymbol{\mu}_l\,\boldsymbol{\mu}_k^\top\mathbf{Q}_c(\tau)\bar{\boldsymbol{\zeta}}_l+\boldsymbol{\zeta}_k^*\mathbf{Q}_c(\tau)\bar{\boldsymbol{\zeta}}_l\,\mathbf{e}_k^*\mathbf{P}_c^+(\tau)\mathbf{e}_l\right\}. \end{aligned} \tag{3.37}$$

As before, for the last term in the right hand side of (3.37) using (3.31) we have

$$\sum_{\tau=-m+1}^{m-1}\boldsymbol{\zeta}_k^*\mathbf{Q}_c(\tau)\bar{\boldsymbol{\zeta}}_l\;\mathbf{e}_k^*\mathbf{P}_c^{+\top}(\tau)\mathbf{e}_l=a_k^2\boldsymbol{\zeta}_k^*\left\{\sum_{\tau=-m+1}^{m-1}\mathbf{Q}_c(\tau)\mathrm{e}^{-\mathrm{i}\omega_k\tau}\right\}\bar{\boldsymbol{\zeta}}_l\,\delta_{k,n+l}=0, \tag{3.38}$$

where we have used (B.13) in Appendix B. Hence combining (3.35), (3.36), (3.37) and (3.38) the theorem follows. ■

4 Complex-valued sinusoids

In this section we revisit the problem of estimating the frequencies of complex-valued sinusoids in noise. As mentioned earlier the subspace based super-resolution frequency estimation algorithms evolved primarily in the context of complex-valued sinusoids. Estimation of complex-valued sinusoidal frequencies has many applications particularly in the the field of telecommunication. Since any complex-valued sinusoidal sequence can be expressed as a combination of two real-valued sinusoidal sequences, corresponding to its real and imaginary part, the R-Esprit algorithm introduced in the previous section has a natural extension to the complex-valued signals. In the context of the current paper, it is a natural question to ask whether one can gain (in computational complexity or accuracy) using such an extension or not. Intuitively, one can expect such a gain because of three reasons. Firstly, for complex-valued signals the ESPRIT algorithm uses all complex operations. This can be reduced by an extended R-Esprit algorithm since all operations in R-Esprit algorithm will still involve only real numbers. Secondly, the number of data samples available for R-Esprit algorithm is about twice (as we will see shortly) the number of data samples available for ESPRIT, while the SNR remains the same. This should lead to improved accuracy. The final point is related to nuisance parameters. Note that for classical ESPRIT, each of the eigenvalues ρ_k of $\mathbf{\Phi}_c$ lies on the unit circle in the complex plane *i.e.* $\mathcal{D}_c$ is *apriori* known to be unitary. But this information is not used in the classical algorithm. Since the finite sample estimate $\widehat{\mathbf{\Phi}}_c$ deviates from the true $\mathbf{\Phi}_c$, there is no straightforward way to assure that the resulting $\widehat{\mathcal{D}}_c$ to be a unitary matrix[4]. So while estimating ρ_k, actually $2n$ real-valued parameters are estimated. On the other hand, this constraint is inbuilt in the R-Esprit. Let us introduce

$$z(t)=\sum_{k=1}^{n}a_k\mathrm{e}^{\mathrm{i}(\omega_kt+\phi_k)}+\epsilon(t) \tag{4.1}$$

[4]One way to incorporate such a constraint is to solve (3.12) in a structured least squares (or total least squares) sense, where $\widehat{\mathbf{\Phi}}_c$ is constrained to be a unitary matrix, so that all its eigenvalues are of unit magnitude, see [4] for example.

where in this case $\epsilon(t)$ is a complex-valued circular white noise sequence of mean zero, *i.e.*

$$\mathbf{E}\epsilon(t)\epsilon^*(t) = \sigma_2^2, \quad \mathbf{E}\epsilon(t)\overline{\epsilon}(t) = 0 \tag{4.2}$$

and ϕ_k is uniformly distributed in $[0, 2\pi)$. The amplitudes $\{a_k\}_{k=1}^n$ are real-valued. For such a model it is well known how to derive the ESPRIT estimate of the frequencies $\{\omega_k\}_{k=1}^n$. We shall not discuss that here. Rather our intention here is to extend the R-Esprit method for such a model. As in the case of real-valued data model RDM, we define

$$\begin{aligned} \mathbf{z}_c(t) &= \left[\begin{array}{cc} z(t) & \ldots z(t+m-1) \end{array} \right]^\top, \\ \mathbf{z}_b(t) &= \left[\begin{array}{cc} z(t-1) & \ldots z(t-m) \end{array} \right]^\top, \\ \mathbf{z}_x(t) &= \tfrac{1}{2}\left\{\mathbf{z}_c(t) + \mathbf{z}_b(t)\right\}. \end{aligned} \tag{4.3}$$

Similarly we define the noise snapshot vector $\boldsymbol{\epsilon}_x(t)$ as before, as the noise corrupted part of $\mathbf{z}_x(t)$, where $\boldsymbol{\epsilon}_x(t)$ is complex-valued in this case. Define

$$\boldsymbol{\epsilon}_1(t) := \mathrm{Re}\{\boldsymbol{\epsilon}_x(t)\}, \quad \boldsymbol{\epsilon}_2(t) := \mathrm{Im}\{\boldsymbol{\epsilon}_x(t)\}. \tag{4.4}$$

It follows readily that $\boldsymbol{\epsilon}_1(t)$ and $\boldsymbol{\epsilon}_2(t)$ are uncorrelated to each other and have identical second order statistics which is similar to that of $\boldsymbol{\epsilon}_r(t)$ in (2.13), *i.e.*

$$\mathbf{E}\boldsymbol{\epsilon}_1(t+\tau)\boldsymbol{\epsilon}_1^\top(t) = \mathbf{E}\boldsymbol{\epsilon}_2(t+\tau)\boldsymbol{\epsilon}_2^\top(t) = \frac{\sigma_2^2}{2}\mathbf{Q}_r(\tau), \tag{4.5}$$

since $\epsilon(t)$ is circular. Following the same steps in the derivation of RDM it is easy to derive that

$$\begin{aligned} \mathrm{Re}\{\mathbf{z}_x(t)\} &= \mathbf{A}_r\mathbf{s}_1(t) + \boldsymbol{\epsilon}_1(t) \\ \mathrm{Im}\{\mathbf{z}_x(t)\} &= \mathbf{A}_r\mathbf{s}_2(t) + \boldsymbol{\epsilon}_2(t) \end{aligned} \tag{4.6}$$

where $\mathbf{s}_1(t)$ and $\mathbf{s}_2(t)$ are given by

$$\mathbf{s}_1(t) = \left[\begin{array}{c} a_1 \cos\left\{\omega_1 t + \phi_1^+\right\} \\ \vdots \\ a_n \cos\left\{\omega_n t + \phi_n^+\right\} \end{array} \right], \quad \mathbf{s}_2(t) = \left[\begin{array}{c} a_1 \sin\left\{\omega_1 t + \phi_1^+\right\} \\ \vdots \\ a_n \sin\left\{\omega_n t + \phi_n^+\right\}. \end{array} \right] \tag{4.7}$$

where $\phi_k^+ = \phi_k - \frac{1}{2}\omega_k$ as before in the basic RDM description. It interesting to note that $\mathbf{s}_1(t)$ and $\mathbf{s}_2(t)$ have the identical second order statistics, *i.e.*

$$\mathbf{P}_x(\tau) := \mathbf{E}\left\{\mathbf{s}_1(t)\mathbf{s}_1^\top(t)\right\} = \mathbf{E}\left\{\mathbf{s}_2(t)\mathbf{s}_2^\top(t)\right\} = \frac{1}{2}\mathrm{diag}\left(a_1^2\cos(\omega_1\tau), \ldots, a_n^2\cos(\omega_n\tau)\right). \tag{4.8}$$

These observations leads us to the extended R-Esprit algorithm for complex valued sinusoids. Define

$$\mathbf{R}_1 := \mathbf{E}\left[\mathrm{Re}\{\mathbf{z}_x(t)\}\mathrm{Re}\{\mathbf{z}_x^\top(t)\}\right] \quad \mathbf{R}_2 := \mathbf{E}\left[\mathrm{Im}\{\mathbf{z}_x(t)\}\mathrm{Im}\{\mathbf{z}_x^\top(t)\}\right] \tag{4.9}$$

$$\mathbf{R}_x := \mathbf{R}_1 + \mathbf{R}_2 = 2\mathbf{A}_r\mathbf{P}_x\mathbf{A}_r^\top + \frac{\sigma_2^2}{2}\mathbf{I}_m, \tag{4.10}$$

where we have used (B.7) for $\mathbf{Q}_r(0)$ and as before $\mathbf{P}_x := \mathbf{P}_x(0)$. We stress here that $\mathbf{R}_1 = \mathbf{R}_2$ does not imply that $\widehat{\mathbf{R}}_1 = \widehat{\mathbf{R}}_2$, which are the corresponding estimates from the data. Therefore from (4.10) it is easy to see how R-Esprit can be extend for complex-valued sinusoids. The R-Esprit algorithm for complex-valued sinusoids will require an estimate $\widehat{\mathbf{R}}_x$ of $\mathbf{R}_x$. The remaining steps will be exactly same as the R-Esprit algorithm. Regarding the estimation accuracy, we have the following straightforward extension of Theorem 1.

Theorem 3: Assume $\epsilon(t)$ is complex Gaussian. Let $\{\boldsymbol{\xi}_k\}_{k=1}^n$ and $\{\boldsymbol{\nu}\}_{k=1}^n$ be defined as in (3.17) and (3.18), respectively, and let $\mathbf{e}_k$ denote the kth column of $\mathbf{I}_n$. Denote the estimation error in ω_k by $\tilde{\omega}_k$ for $1 \leq k \leq n$. Then for the extended R-Esprit algorithm the asymptotic second order statistics of the estimation error is given by

$$\begin{aligned} \mathbf{E}\tilde{\omega}_k\tilde{\omega}_l &= \frac{\sigma_2^4}{N-2m-1} \; \frac{1}{2a_k^2 a_l^2 \sin(\omega_k)\sin(\omega_l)} \\ &\times \sum_{\tau=-2m+1}^{2m-1} \left\{ \boldsymbol{\xi}_k^\top \mathbf{Q}_r(\tau)\boldsymbol{\xi}_l \; \boldsymbol{\nu}_k^\top \mathbf{Q}_r(\tau)\boldsymbol{\nu}_l + \boldsymbol{\xi}_k^\top \mathbf{Q}_r(\tau)\boldsymbol{\nu}_l \; \boldsymbol{\nu}_k^\top \mathbf{Q}_r(\tau)\boldsymbol{\xi}_l \right\} \end{aligned} \tag{4.11}$$

Proof: The proof is similar to that of Theorem 1. Following the same set arguments as in (3.22) and using (4.8) and (A.14) in this case we get the estimation error as

$$\tilde{\omega}_k = -\tilde{\rho}_k/\sin(\omega_k) = -\frac{\boldsymbol{\xi}_k^* \widehat{\mathbf{R}}_x \boldsymbol{\nu}_k}{a_k^2 \sin(\omega_k)} = -\frac{\boldsymbol{\xi}_k^* \left\{ \widehat{\mathbf{R}}_1 + \widehat{\mathbf{R}}_2 \right\} \boldsymbol{\nu}_k}{a_k^2 \sin(\omega_k)}, \tag{4.12}$$

Now using (A.20) in Lemma 2 and (4.5) we get

$$\begin{aligned} &\mathbf{E}\left\{ \left(\boldsymbol{\xi}_k^* \widehat{\mathbf{R}}_1 \boldsymbol{\nu}_k\right) \left(\boldsymbol{\xi}_l^* \widehat{\mathbf{R}}_1 \boldsymbol{\nu}_l\right) \right\} = \mathbf{E}\left\{ \left(\boldsymbol{\xi}_k^* \widehat{\mathbf{R}}_2 \boldsymbol{\nu}_k\right) \left(\boldsymbol{\xi}_l^* \widehat{\mathbf{R}}_2 \boldsymbol{\nu}_l\right) \right\} \\ = \; &\frac{\sigma_2^4/4}{N-2m-1} \sum_{\tau=-2m+1}^{2m-1} \left\{ \boldsymbol{\xi}_k^\top \mathbf{Q}_r(\tau)\boldsymbol{\xi}_l \; \boldsymbol{\nu}_k^\top \mathbf{Q}_r(\tau)\boldsymbol{\nu}_l + \boldsymbol{\xi}_k^\top \mathbf{Q}_r(\tau)\boldsymbol{\nu}_l \; \boldsymbol{\nu}_k^\top \mathbf{Q}_r(\tau)\boldsymbol{\xi}_l \right\}. \end{aligned} \tag{4.13}$$

Note that we have also used a similar argument as in (3.24) in the last equality. Next we show that

$$\mathbf{E}\left\{ \left(\boldsymbol{\xi}_k^* \widehat{\mathbf{R}}_1 \boldsymbol{\nu}_k\right) \left(\boldsymbol{\xi}_k^* \widehat{\mathbf{R}}_2 \boldsymbol{\nu}_k\right) \right\} = 0. \tag{4.14}$$

Note from (A.22) that

$$\boldsymbol{\xi}_k^* \widehat{\mathbf{R}}_j \boldsymbol{\nu}_k = \frac{1}{N-2m+1} \sum_{t=1}^{N-2m+1} \boldsymbol{\xi}_k^* \boldsymbol{\epsilon}_j(t) \left\{ \boldsymbol{\epsilon}_j^*(t)\boldsymbol{\nu}_k + \mathbf{s}_j^*(t)\mathbf{e}_k \right\} \tag{4.15}$$

for $j = 1, 2$, where $\mathbf{e}_k$ denote the kth column of $\mathbf{I}_n$. Note also that $\mathbf{E}\boldsymbol{\epsilon}_1(t+\tau)\boldsymbol{\epsilon}_2^\top(t) = 0$ for all τ. Also the third order moments of jointly Gaussian random variables vanish. Hence it follows that

$$\begin{aligned} &\mathbf{E}\left[\boldsymbol{\xi}_k^* \boldsymbol{\epsilon}_1(t) \left\{ \boldsymbol{\epsilon}_1^*(t)\boldsymbol{\nu}_k + \mathbf{s}_1^*(t)\mathbf{e}_k \right\} \boldsymbol{\xi}_l^* \boldsymbol{\epsilon}_2(t) \left\{ \boldsymbol{\epsilon}_2^*(t)\boldsymbol{\nu}_l + \mathbf{s}_2^*(t)\mathbf{e}_l \right\} \right] \\ = \; &\mathbf{E}\left[\boldsymbol{\xi}_k^* \boldsymbol{\epsilon}_1(t)\boldsymbol{\epsilon}_1^*(t)\boldsymbol{\nu}_k \right] \mathbf{E}\left[\boldsymbol{\xi}_l^* \boldsymbol{\epsilon}_2(t)\boldsymbol{\epsilon}_2^*(t)\boldsymbol{\nu}_l \right] = \frac{\sigma_2^4}{4} \left[\boldsymbol{\xi}_k^* \boldsymbol{\nu}_k \boldsymbol{\xi}_l^* \boldsymbol{\nu}_l \right] = 0 \end{aligned} \tag{4.16}$$

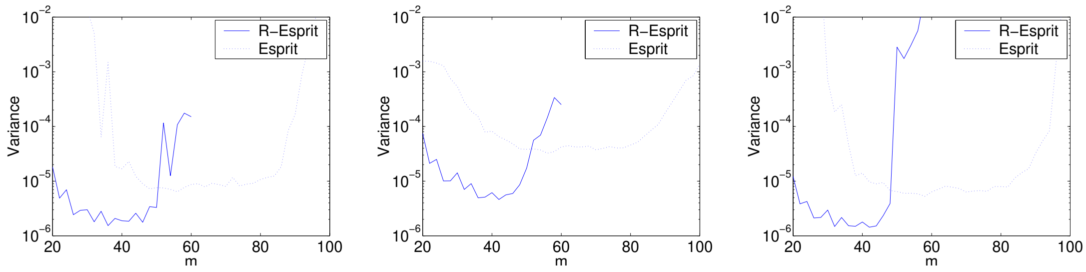
where in the second equality we used (4.5) and (B.7), while the last equality we have used the corollary 2 to the Lemma 1 in appendix A. Hence (4.14) follows from

(4.15) and (4.16). Consequently, (4.11) follows from (4.12), (4.13) and (4.14). ■

Note that it is also possible to apply R-Esprit algorithm individually to each of the real and imaginary parts of the observed complex-valued data. It is easy to see from Theorem 1 that such an estimate has a parameter variance twice that of the extended R-Esprit algorithm. This is expected because R-Esprit applied to the real part (or imaginary part) of the data used only half the number of data samples used by the extended R-Esprit algorithm.

5 Numerical simulation results

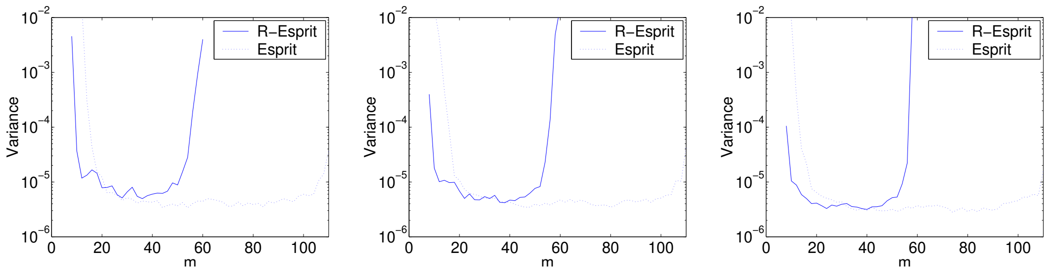
In this section we illustrate the advantage of R-Esprit algorithm in comparison to the conventional ESPRIT algorithm. We shall consider the case when the data is real-valued separately from the case when the data is complex-valued. We shall also display different aspects like small sample properties, large sample performance, performance with closely spaced frequencies with different examples. In the implementation of the ESPRIT algorithm we have used forward-backward averaging, which is known to give better estimates of the covariance matrices [8, 11] for small samples, which in turn results in a better accuracy of the parameter estimates [2].



(a) $\omega_1 = 0.60$ (b) $\omega_2 = 0.65$ (c) $\omega_3 = 0.70$

Figure 1: Comparison of the small sample ($N = 125$, $\sigma_1 = 0.1$) variances of the R-Esprit (solid line) and ESPRIT (dotted line) estimators for closely spaced frequencies.

Example 1: In this example, we consider the case when frequencies are very closely spaced and the data are real-valued. In Figure 1 we show the comparative performance of ESPRIT and R-Esprit algorithm with $\omega_1 = 0.60$, $\omega_2 = 0.65$ and $\omega_3 = 0.70$, while the amplitudes of the sinusoids are $a_1 = 1$, $a_2 = 1$ and $a_3 = 1.1$, respectively. The noise standard deviation $\sigma_1 = 0.1$. The number of data samples N was 125. In Figure 1 we have plotted for each frequency, the sample variance as a function of m, the dimension of the 'snapshot' vector, when the frequencies were estimated using R-Esprit and ESPRIT algorithms respectively. For each value of m the individual variances of the estimates were from 500 independent Monte-Carlo simulations. It is interesting to note that the minimum value of variance (with respect to m) of the estimates for the R-Esprit algorithm is about 10% of that obtained ESPRIT algorithm. Another interesting point is the comparative performance of the algorithms for lower values of m. We see that the it is possible to get very accurate estimates of the frequencies for quite small value of m. That would mean that R-Esprit algorithm is capable of yielding very accurate estimates with a computational complexity which is significantly lower than the amount computation required by ESPRIT in order to generate estimates of comparable accuracy. For a more quantitative comparison, the Matlab flops required by the ESPRIT algorithm to achieve its best performance is 3.5 Mflops, while to achieve the same accuracy R-Esprit needs 0.4 Mflops, which is only 12% of that for the ESPRIT algorithm.

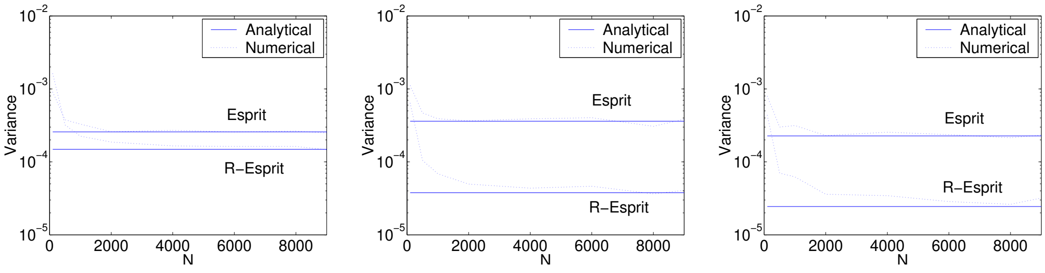


(a) $\omega_1 = 0.40$ (b) $\omega_2 = 0.65$ (c) $\omega_3 = 1.0$

Figure 2: Comparison of the small sample ($N = 125$, $\sigma_1 = 0.5$) variances of the R-Esprit (solid line) and ESPRIT (dotted line) estimators for sparsely spaced frequencies.

Example 2: Here we illustrate the comparative small sample performance of ESPRIT and the R-Esprit algorithm when the frequencies are not very close to each other. In this case we have $\omega_1 = 0.40$, $\omega_2 = 0.65$ and $\omega_3 = 1.0$, while the amplitudes of the sinusoids are $a_1 = 1$, $a_2 = 1$ and $a_3 = 1.1$, respectively. The noise standard deviation $\sigma_1 = 0.5$. The number of data samples N was 125. R-Esprit and ESPRIT algorithms were applied to 500 independent data sets. The result of estimation is displayed in Figure 2. In this case the minimum variance variance achieved by both the algorithms were of comparable magnitude. Note though that very accurate estimates of the frequencies were obtained by the R-Esprit algorithm for significantly smaller values of m compared to the ESPRIT algorithm, which is a clear indication of the significant computational gain of using R-Esprit.

Example 3: This experiment is to examine the large sample properties of the estimates. In this case we keep m fixed at 20 and vary N and study the variation of the variance of the estimates with N. We also compare the sample variance from the numerical simulation with the analytically predicted variance from Theorem 1 and Theorem 2. In this case we have $\omega_1 = 0.30$, $\omega_2 = 0.70$ and $\omega_3 = 1.0$, while the amplitudes of the sinusoids are $a_1 = 1$, $a_2 = 1$ and $a_3 = 1.1$, respectively. The noise standard deviation $\sigma_1 = 0.5$. For each N, the variance is obtained using the estimates from 500 independent realizations. The results are shown in Figure 3. Here we plot the normalized variance as a function of N. By normalized variance we mean the variance of $\tilde{\omega}_k\sqrt{(N-2m+1)}$ for R-Esprit method, while

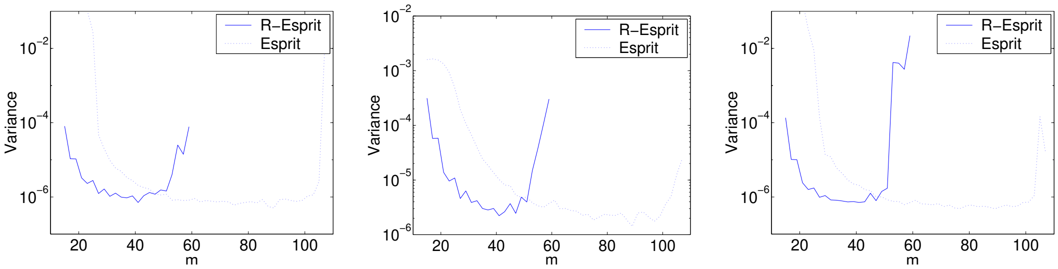


(a) $\omega_1 = 0.30$ (b) $\omega_2 = 0.7$ (c) $\omega_3 = 1.0$

Figure 3: Comparison of the large sample ($\sigma_1 = 0.5$) normalized variances of R-Esprit and ESPRIT. Analytical variances (solid lines) are compared with the numerical variances (dotted lines) for $m = 20$

.

the normalized variance for ESPRIT method is the variance of $\tilde{\omega}_k\sqrt{(N-m+1)}$. The asymptotic normalized variance as derived in Theorems 1 and 2, are constant with respect to N, if m is maintained at a constant value. In Figure 3 we see that numerical variances of the R-Esprit and ESPRIT estimates are well in accordance with the analytical mean square error predicted by Theorems 1 and 2. We can also note that for large sample case, the R-Esprit algorithm outperforms ESPRIT algorithm by significantly large margin.

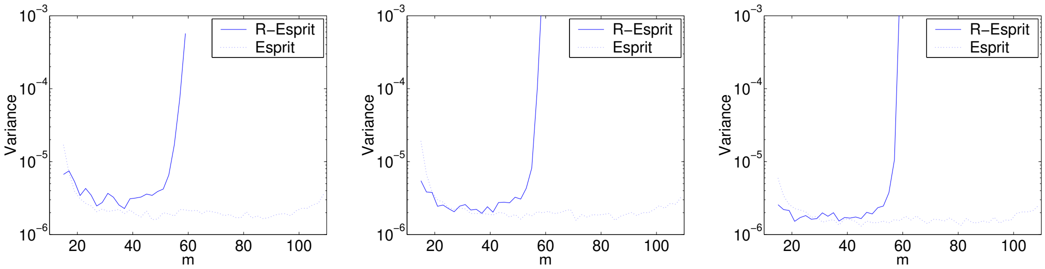


(a) $\omega_1 = 0.60$ (b) $\omega_2 = 0.65$ (c) $\omega_3 = 0.70$

Figure 4: Comparison of the small sample ($N = 125$, $\sigma_2 = 0.2$) variances of the extended R-Esprit (solid line) and ESPRIT (dotted line) estimators for closely spaced frequencies of complex-valued sinusoids.

Example 4: Next we illustrate the performance of the extended R-Esprit algorithm for complex-valued sine waves. As in the case of real-valued sine waves here also we consider three different cases. First we consider the small sample performance in case of closely spaced sine waves. Here also we consider three sine waves with frequencies $\omega_1 = 0.60$, $\omega_2 = 0.65$ and $\omega_3 = 0.70$, respectively, with associated amplitudes $a_1 = 1$, $a_2 = 1$ and $a_3 = 1.1$, respectively. The standard deviation of each of the real and imaginary parts of the noise sequence is 0.1, which means $\sigma_2 = 0.1\sqrt{2}$. 500 independent Monte Carlo simulations were performed. For each realization the number of data samples $N = 125$. The results are shown in Figure

4. In this case we can see that the minimum attainable parameter variance are of the similar magnitude for both R-Esprit and ESPRIT algorithm. But also in this case the R-Esprit performs significantly better than ESPRIT for lower value of m. The resulting computational gain is even greater in this case. Note that for ESPRIT for complex-valued data, all the relevant quantities are complex-valued and each complex operation is costlier than the corresponding real operation. The number of Matlab flops required to obtain the best accuracy for R-Esprit is 1.3 Mflops, which is about 10% of 12 Mflops required by ESPRIT algorithm to achieve the same accuracy.

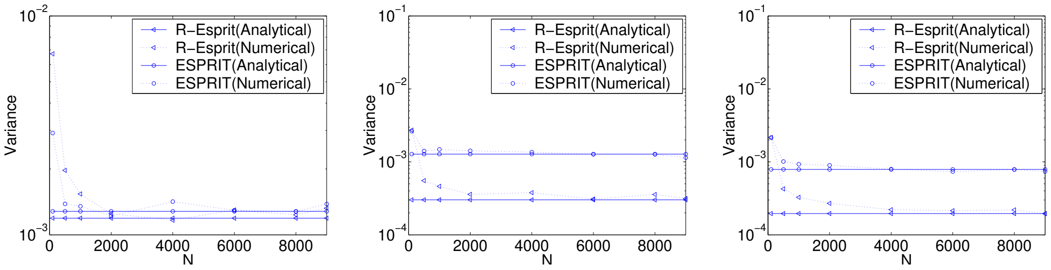


(a) $\omega_1 = 0.40$ (b) $\omega_2 = 0.65$ (c) $\omega_3 = 1.0$

Figure 5: Comparison of the small sample ($N = 125$, $\sigma_1 = 0.5$) variances of the R-Esprit (solid line) and ESPRIT (dotted line) estimators for sparsely spaced frequencies of complex-valued sinusoids.

Example 5: In this example we illustrate the comparative small sample performances of ESPRIT and the extended R-Esprit algorithms when the frequencies of complex-valued sin waves are not very close to each other. In this case we have $\omega_1 = 0.40$, $\omega_2 = 0.65$ and $\omega_3 = 1.0$, while the amplitudes of the sinusoids are $a_1 = 1$, $a_2 = 1$ and $a_3 = 1.1$ respectively. The noise standard deviation $\sigma_1 = 0.5\sqrt{2}$. The number of data samples N was 125. R-Esprit and ESPRIT algorithms were applied to 500 independent data sets. The result of estimation is displayed in Figure 5. Unlike the previous cases in this case we see that for smaller values of m R-Esprit and ESPRIT preform equally well and the best achievable performance of both the algorithms are also of comparable. But using R-Esprit we have a computational gain. For $m = 17$ the Matlab Flops required by R-Esprit is 0.5 Mflops, which is 60% of 0.8 Mflops required by ESPRIT.

Example 6: In our final example we illustrate the large sample properties of the R-Esprit and extended ESPRIT algorithm for complex-valued sine wave frequencies. In that aim we keep m fixed at 20 and vary N and study the variation of the variance of the estimates with N. Here we shall compare the analytically predicted normalized variance of the extended R-Esprit algorithm in Theorem 3 to the normalized numerical variance from the simulations. We shall also compare the normalized analytical and numerical variances of ESPRIT algorithm. Here we use the expression for the large sample mean square error of ESPRIT algorithm for complex-valued sine waves derived in [25]. In this case we have $\omega_1 = 0.30$, $\omega_2 = 0.70$ and $\omega_3 = 1.0$, while the amplitudes of the sinusoids are $a_1 = 1$, $a_2 = 1$



(a) $\omega_1 = 0.30$ (b) $\omega_2 = 0.7$ (c) $\omega_3 = 1.0$

Figure 6: Comparison of the large sample ($\sigma_1 = 0.5$) normalized variances of the extended R-Esprit and ESPRIT for complex-valued sin waves. Analytical variances(solid lines) are compared with numerical variances (dotted lines) for $m = 20$.

and $a_3 = 1.1$, respectively. The noise standard deviation $\sigma_1 = 0.5\sqrt{2}$. For each N, the variance is obtained using the estimates from 500 independent realizations. The results are displayed in Figure 6. Observations are similar to those in Figure 3. It can be seen that for large samples the R-Esprit outperforms ESPRIT, particularly for higher frequencies. The analytical variance predicted by Theorem 3 can be seen to be well in accordance with the numerical variance from the simulations.

6 Discussion

One of the two main observations from the numerical simulation results is that the R-Esprit algorithm is capable of generating very good estimates at considerably lower computational cost than ESPRIT. The second interesting observation is the significant improvement in the asymptotic accuracy of the estimates. Although the motivation behind developing the R-Esprit algorithm was to exploit the additional constraint if the data is real-valued, we saw that the R-Esprit algorithm is a better alternative even if the data is complex valued. The most striking advantage of R-Esprit algorithm is the associated fast decay of the mean square error with m. In other words, we can generate very accurate estimate using R-Esprit algorithm for considerably small values of m. In finite sample cases for closely spaced real-valued sine waves, the improvement in the accuracy using R-Esprit is one of the most significant results in this paper. Resolving closely spaced sinusoidal frequencies is the most challenging problem in the estimation of line spectra. As seen in Figure 1, R-Esprit is capable of resolving very closely placed sinusoids from small data records with considerably higher accuracy than the ESPRIT algorithm. We mention here that it is possible to improve the resolution capabilities of R-Esprit algorithm even more using the decimation approach described in [5].

Note that the m dimensional subspace description RDM used a window of length $2m$, see (2.2), (2.10) and (2.11). This is one of the main reasons why R-Esprit can give improved estimates for smaller m. On the other hand this is the reason behind the fast degradation of the performance of R-Esprit algorithm with increasing m, when small a record of data is available. For a more clear explanation let us refer to (3.12). Recall that the estimates are derived from the eigenvalues of $\widehat{\mathbf{\Phi}}$ (*i.e.* for R-Esprit it is $\widehat{\mathbf{\Phi}}_r$ and for ESPRIT it is $\widehat{\mathbf{\Phi}}_c$). Hence the accuracy of $\widehat{\mathbf{\Phi}}$ has a direct effect on the accuracy of the frequencies. The accuracy of $\widehat{\mathbf{\Phi}}$ depends upon two contradictory factors. An increase in m improves the accuracy of $\widehat{\mathbf{\Phi}}$ in the sense that a larger number of equations are available in (3.12) to solve for $\widehat{\mathbf{\Phi}}$. On the other hand for small samples an increase in m means lowering the number of data samples to compute $\widehat{\mathbf{R}}$, which results in poor estimates $\widehat{\mathbf{R}}$ and hence $\widehat{\mathbf{S}}$. For small m, the first factor dominates so that the estimation accuracy improves rapidly with an increase in m. For larger m the second factor dominates so that the estimation accuracy degrades with increase in m. This phenomenon for a limited number of samples can seen in Figures 1, 2, 4 and 5 for both R-Esprit and ESPRIT algorithms. Since for the R-Esprit algorithm, we have only $N-2m+1$ data samples to compute $\widehat{\mathbf{R}}_r$ (the real-valued data case) compared to $N-m+1$ data samples for ESPRIT algorithm, the degradation of the performance of R-Esprit algorithm with increase in m occurs for smaller values of m.

We can notice in Figures 3 and 6 that the improvement in the accuracy of R-Esprit algorithm compared to the accuracy in ESPRIT algorithm is not as significantly impressive at lower frequencies as in case of the higher frequencies. This can be explained by the accuracy results in Theorems 1 and 3. We see that the mean square error at frequency ω_k is inversely proportional to $\sin^2(\omega_k)$. This implies that the accuracy of R-Esprit algorithm degrades at frequencies close to 0 and π. The accuracy of $\widehat{\mathbf{\Phi}}$ may be improved by incorporating weighting in (3.12), resulting in an optimally weighted least squares or total least squares solution of (3.12). We did not consider such a possibility in this paper. See [1] for such an approach in the context of sensor array signal processing. Finally we mention here that in this paper we have dealt with LS-ESPRIT approach, in Theorems 1-3, *i.e.* we assume that (3.12) is solved in least squares sense. As another alternative one can solve (3.12) in a total least squares sense, which in general brings in an improvement in small sample properties of the estimates. However, in many earlier papers on signal processing and system identification it has been shown that TLS and LS behave equivalently as far as large sample behaviour is concerned [26, 19]. See [14] for a comprehensive description of the large sample behaviour of the TLS-ESPRIT.

7 Conclusions

In this work we have proposed an ESPRIT like algorithm, namely R-Esprit for estimation of the frequencies of sine waves observed in noise. The algorithm is motivated by the real-valued data case where the conventional subspace based algorithms do not account for the *realness* of the data. In order to incorporate this additional information in the algorithm, we propose a compact low rank signal model RDM, where the dimension of the signal subspace is reduced to n from $2n$ in case of the conventional complex-valued data model for n real-valued sine waves. An ESPRIT like algorithm is also developed, which exploits RDM to obtain the estimates of the frequencies. Next, R-Esprit is extended for the case of complex-valued sine waves, where the real and imaginary parts of the complex-valued data are used as input to the R-Esprit algorithm. Asymptotic analysis of the R-Esprit algorithm is presented both for real-valued and the complex-valued data case. An analysis of ESPRIT for the real-valued data case is also carried out. The algorithms are compared using numerical simulations. The results from the numerical simulations show that the R-Esprit algorithm results in a significant gain in computational complexity compared to ESPRIT, without any compromise in the accuracy. For closely spaced real-valued sine waves R-Esprit outperforms ESPRIT in accuracy by a significantly large margin. It is also shown that the asymptotic performance of R-Esprit is better than ESPRIT algorithm. The asymptotic analysis results derived in this paper are also validated through the simulation study.

A Preliminary Results

In this section we state and prove two preliminary lemmas which is useful for analysis of Esprit type algorithms. The first one is regarding the large sample perturbation of the eigenvalues of $\mathcal{D}$. The first lemma follows immediately.

Lemma 1: Consider the general model structure (3.1) introduced in the context of the general description of the ESPRIT algorithm in Section 3. Denote the eigenvalues of $\mathbf{\Phi}$ (*i.e.* the diagonal entries of $\mathcal{D}$, see(3.12)) by $\{\rho_k\}_{k=1}^{p}$ and let $\mathbf{e}_k$ denote the kth column of $\mathbf{I}_p$. Define

$$\begin{aligned} \boldsymbol{\eta}_k^* &:= \mathbf{e}_k^* \left\{\mathbf{A}^*\mathbf{T}^{(1)*}\mathbf{T}^{(1)}\mathbf{A}\right\}^{-1} \mathbf{A}^*\mathbf{T}^{(1)*}\left\{\mathbf{T}^{(2)} - \mathbf{T}^{(1)}\rho_k\right\}, &\text{(A.1)}\\ \mathbf{X} &:= \left[\begin{array}{ccc} \boldsymbol{\eta}_1 & \dots & \boldsymbol{\eta}_p \end{array}\right], &\text{(A.2)}\\ \mathbf{B} &:= \mathbf{A}\left\{\mathbf{A}^*\mathbf{A}\right\}^{-1} = \left[\begin{array}{ccc} \mathbf{b}_1 & \dots & \mathbf{b}_p \end{array}\right]. &\text{(A.3)} \end{aligned}$$

Then for large N, the first order perturbation $\tilde{\rho}_k$ of the estimate $\hat{\rho}_k$ from the true value ρ_k is given by

$$\tilde{\rho}_k = \left[\mathbf{X}^*\widehat{\mathbf{R}}\mathbf{B}\mathbf{P}^{-1}\right]_{kk}, \quad 1 \le k \le p. \tag{A.4}$$

Proof: The essential ingredients for the proof of this Lemma is given in [25]. We repeat a summary of the proof here for convenience of the readers. The goal of this lemma is to relate the perturbation $\hat{\rho}_k$ to the data $\{y(t)\}_{t=1}^{N}$ through the estimated covariance matrix $\widehat{\mathbf{R}}$. In that aim well known results on the first order perturbation expansion of the eigenvalues of matrices, see [21, 27], we have the perturbation $\tilde{\rho}_k$ due to a perturbation $\widetilde{\mathbf{\Phi}}$ in $\mathbf{\Phi}$ as

$$\tilde{\rho}_k = \mathbf{e}_k^*\mathbf{C}\widetilde{\mathbf{\Phi}}\mathbf{C}^{-1}\mathbf{e}_k. \tag{A.5}$$

Next we consider the perturbation $\widetilde{\mathbf{\Phi}}$ due to a perturbation $\widetilde{\mathbf{S}}$. Recall that $\widehat{\mathbf{\Phi}}$ is a least squares solution of (3.12) with $\mathbf{S}$ replaced by $\widehat{\mathbf{S}}$. Introduce

$$\begin{aligned} \mathbf{S}_1 &:= \mathbf{T}^{(1)}\mathbf{S}, \quad \mathbf{A}_1 := \mathbf{T}^{(1)}\mathbf{A}, &\text{(A.6)}\\ \mathbf{S}_2 &:= \mathbf{T}^{(2)}\mathbf{S}, \quad \mathbf{A}_2 := \mathbf{T}^{(2)}\mathbf{A}. &\text{(A.7)} \end{aligned}$$

Then using well known results in the theory of least squares estimation and instrumental variables methods [20, 12] the perturbation $\widetilde{\mathbf{\Phi}}$ is given by

$$\begin{aligned}
\widetilde{\mathbf{\Phi}} &= \left\{\mathbf{S}_1^*\mathbf{S}_1\right\}^{-1}\mathbf{S}_1^*\left\{\widetilde{\mathbf{S}}_2-\widetilde{\mathbf{S}}_1\mathbf{\Phi}\right\} \\
&= \mathbf{C}^{-1}\left\{\mathbf{A}_1^*\mathbf{A}_1\right\}^{-1}\mathbf{A}_1^*\left\{\mathbf{T}^{(2)}\widetilde{\mathbf{S}}-\mathbf{T}^{(1)}\widetilde{\mathbf{S}}\mathbf{\Phi}\right\}. \qquad (A.8)
\end{aligned}$$

Hence combining (A.5) and (A.8) we get

$$\begin{aligned}
\tilde{\rho}_k &= \mathbf{e}_k^*\left\{\mathbf{A}_1^*\mathbf{A}_1\right\}^{-1}\mathbf{A}_1^*\left\{\mathbf{T}^{(2)}\widetilde{\mathbf{S}}-\mathbf{T}^{(1)}\widetilde{\mathbf{S}}\mathbf{\Phi}\right\}\mathbf{C}^{-1}\mathbf{e}_k \\
&= \mathbf{e}_k^*\left\{\mathbf{A}_1^*\mathbf{A}_1\right\}^{-1}\mathbf{A}_1^*\left\{\mathbf{T}^{(2)}\widetilde{\mathbf{S}}\mathbf{C}^{-1}-\mathbf{T}^{(1)}\widetilde{\mathbf{S}}\mathbf{C}^{-1}\mathcal{D}\right\}\mathbf{e}_k \\
&= \mathbf{e}_k^*\left\{\mathbf{A}_1^*\mathbf{A}_1\right\}^{-1}\mathbf{A}_1^*\left\{\mathbf{T}^{(2)}-\mathbf{T}^{(1)}\rho_k\right\}\widetilde{\mathbf{S}}\mathbf{C}^{-1}\mathbf{e}_k \\
&= \boldsymbol{\eta}_k^*\widetilde{\mathbf{S}}\mathbf{C}^{-1}\mathbf{e}_k, \qquad (A.9)
\end{aligned}$$

where we have used the definition of $\boldsymbol{\eta}_k$ in (A.1). Next we need to relate the perturbation $\widetilde{\mathbf{S}}$ in $\mathbf{S}$ of $\mathbf{R}$ to $\widehat{\mathbf{R}}$. An interesting observation regarding $\boldsymbol{\eta}_k$ is very crucial for that purpose. For all $1 \le k \le p$

$$\begin{aligned}
\boldsymbol{\eta}_k^*\mathbf{A} &= \mathbf{e}_k^*\left\{\mathbf{A}_1^*\mathbf{A}_1\right\}^{-1}\mathbf{A}_1^*\left\{\mathbf{A}_2-\mathbf{A}_1\rho_k\right\} \\
&= \mathbf{e}_k^*\left\{\mathcal{D}-\rho_k)\mathbf{I}_p\right\} = \mathbf{0}_{1\times m} \quad \Rightarrow \\
\boldsymbol{\eta}_k^* &= \boldsymbol{\alpha}_k^*\mathbf{G}^*, \qquad (A.10)
\end{aligned}$$

for some $(m-p)|1$ dimensional vector $\boldsymbol{\alpha}_k$. In the aim of relating $\widetilde{\mathbf{S}}$ to $\widetilde{\mathbf{R}}$ we can now use the following well known first order perturbation result, which has been proved in many previous papers on this subject and direction of arrival estimation (see for example [25]).

$$\mathbf{G}^*\widehat{\mathbf{R}}\mathbf{S} \approx \mathbf{G}^*\widetilde{\mathbf{S}}\mathbf{\Lambda}+\sigma^2\widetilde{\mathbf{G}}^*\mathbf{S} \approx \mathbf{G}^*\widetilde{\mathbf{S}}\left\{\mathbf{\Lambda}-\sigma^2\mathbf{I}_p\right\}. \qquad (A.11)$$

Now combination of (A.9), (A.10) and (A.11) gives

$$\tilde{\rho}_k = \boldsymbol{\eta}_k^*\widehat{\mathbf{R}}\mathbf{S}\left\{\mathbf{\Lambda}-\sigma^2\mathbf{I}_p\right\}^{-1}\mathbf{C}^{-1}\mathbf{e}_k \qquad (A.12)$$

Now from (3.11) and (A.3) we have

$$\begin{aligned}
\mathbf{P}^{-1} &= \mathbf{A}^*\mathbf{S}\left\{\mathbf{\Lambda}-\sigma^2\mathbf{I}_p\right\}^{-1}\mathbf{C}^{-1} \quad \Rightarrow \\
\mathbf{B}\mathbf{P}^{-1} &= \mathbf{A}\{\mathbf{A}^*\mathbf{A}\}^{-1}\mathbf{A}^*\mathbf{A}\mathbf{C}\left\{\mathbf{\Lambda}-\sigma^2\mathbf{I}_p\right\}^{-1}\mathbf{C}^{-1} \\
&= \mathbf{S}\left\{\mathbf{\Lambda}-\sigma^2\mathbf{I}_p\right\}^{-1}\mathbf{C}^{-1}. \qquad (A.13)
\end{aligned}$$

Combining (A.13) with (A.12) we get (A.4). ■

Corollary 1: Let $\mathbf{P}$ be a diagonal matrix (which is the case in the current context) such that $\mathbf{P} = \mathrm{diag}(\alpha_1,\ldots,\alpha_p)$. Then from (A.4) it is straightforward that

$$\tilde{\rho}_k = \frac{\boldsymbol{\eta}_k^*\widehat{\mathbf{R}}\mathbf{b}_k}{\alpha_k} \qquad (A.14)$$

$\blacksquare$

Corollary 2: From (A.10) and (A.3) it follows that $\boldsymbol{\eta}_k^* \mathbf{b}_l = 0$, for $1 \leq k, l \leq p$. $\blacksquare$

The next lemma is required to derive the mean square errors of the parameter estimates described in Theorems 1, 2 and 3.

Lemma 2: Consider the general model structure (3.1), where $\boldsymbol{\epsilon}(t)$ is real-valued. Let

$$
\begin{aligned}
\mathbf{P}(\tau) &:= \mathbf{E}\mathbf{s}(t+\tau)\mathbf{s}^*(t), && \text{(A.15)}\\
\overline{\mathbf{P}}(\tau) &:= \mathbf{E}\overline{\mathbf{s}}(t+\tau)\mathbf{s}^*(t), && \text{(A.16)}\\
\mathbf{Q}(\tau) &:= \mathbf{E}\boldsymbol{\epsilon}(t+\tau)\boldsymbol{\epsilon}^\top(t). && \text{(A.17)}
\end{aligned}
$$

Further, assume that $\mathbf{Q}(\tau) = \mathbf{0}_{m\times m}$ for $|\tau| > \tau_{\max}$ and that the covariance matrix $\mathbf{R}$ in (3.2) is estimated as

$$
\widehat{\mathbf{R}} = \frac{1}{M}\sum_{t=1}^{M} \mathbf{y}(t)\mathbf{y}^*(t). \tag{A.18}
$$

Then for large $M \gg m$ (m may also be large)

$$
\begin{aligned}
\mathbf{E}\left\{\left(\boldsymbol{\eta}_k^*\widehat{\mathbf{R}}\mathbf{b}_k\right)\left(\boldsymbol{\eta}_l^*\widehat{\mathbf{R}}\mathbf{b}_l\right)\right\} &= \frac{1}{M}\sum_{\tau=-\tau_{\max}}^{\tau_{\max}} \{\boldsymbol{\eta}_k^*\mathbf{Q}(\tau)\bar{\boldsymbol{\eta}}_l\ \mathbf{b}_k^\top\mathbf{Q}(\tau)\mathbf{b}_l \\
&\quad +\boldsymbol{\eta}_k^*\mathbf{Q}(\tau)\mathbf{b}_l\ \mathbf{b}_k^\top\mathbf{Q}(\tau)\bar{\boldsymbol{\eta}}_l + \boldsymbol{\eta}_k^*\mathbf{Q}(\tau)\bar{\boldsymbol{\eta}}_l\ \mathbf{e}_k^*\overline{\mathbf{P}}(\tau)\mathbf{e}_l\}, && \text{(A.19)}\\
\mathbf{E}\left\{\left(\boldsymbol{\eta}_k^*\widehat{\mathbf{R}}\mathbf{b}_k\right)\left(\boldsymbol{\eta}_l^*\widehat{\mathbf{R}}\mathbf{b}_l\right)^*\right\} &= \frac{1}{M}\sum_{\tau=-\tau_{\max}}^{\tau_{\max}} \boldsymbol{\eta}_k^*\mathbf{Q}(\tau)\boldsymbol{\eta}_l\ \mathbf{b}_k^\top\mathbf{Q}(\tau)\bar{\mathbf{b}}_l \\
&\quad +\boldsymbol{\eta}_k^*\mathbf{Q}(\tau)\bar{\mathbf{b}}_l\ \mathbf{b}_k^\top\mathbf{Q}(\tau)\boldsymbol{\eta}_l + \boldsymbol{\eta}_k^*\mathbf{Q}(\tau)\boldsymbol{\eta}_l\ \mathbf{e}_k^*\mathbf{P}^\top(\tau)\mathbf{e}_l\}. && \text{(A.20)}
\end{aligned}
$$

Proof: From (3.1) the estimate of the covariance matrix is given by

$$
\widehat{\mathbf{R}} = \frac{1}{M}\sum_{t=1}^{M}\mathbf{y}(t)\mathbf{y}^*(t) = \frac{1}{M}\sum_{t=1}^{M}\{\mathbf{A}\mathbf{s}(t)+\boldsymbol{\epsilon}(t)\}\{\mathbf{A}\mathbf{s}(t)+\boldsymbol{\epsilon}(t)\}^*. \tag{A.21}
$$

The value of the scaling integer M depends on the estimation approach. For R-Esprit algorithm $M = N - 2m + 1$, while for ESPRIT algorithm $M = N - m + 1$. Using (A.10) we have

$$
\begin{aligned}
\boldsymbol{\eta}_k^*\widehat{\mathbf{R}}\mathbf{b}_k &= \frac{1}{M}\sum_{t=1}^{M}\boldsymbol{\eta}_k^*\boldsymbol{\epsilon}(t)\{\mathbf{A}\mathbf{s}(t)+\boldsymbol{\epsilon}(t)\}^*\mathbf{b}_k \\
&= \frac{1}{M}\sum_{t=1}^{M}\boldsymbol{\eta}_k^*\boldsymbol{\epsilon}(t)\{\boldsymbol{\epsilon}^*(t)\mathbf{b}_k + \mathbf{s}^*(t)\mathbf{e}_k\} && \text{(A.22)}
\end{aligned}
$$

Note that $\mathbf{s}(t_1)$ and $\boldsymbol{\epsilon}(t_2)$ are independent random variables for all t_1 and t_2, also the third order moments of a Gaussian random vector vanish. Using results on the fourth order moments of jointly Gaussian random vectors [7] we get

$$
\begin{aligned}
&\mathbf{E}\left\{\left(\boldsymbol{\eta}_k^*\widehat{\mathbf{R}}\mathbf{b}_k\right)\left(\boldsymbol{\eta}_l^*\widehat{\mathbf{R}}\mathbf{b}_l\right)\right\} \\
= &\ \frac{1}{M^2}\sum_{t_1=1}^{M}\sum_{t_2=1}^{M}\mathbf{E}\left\{\boldsymbol{\eta}_k^*\boldsymbol{\epsilon}(t_1)\boldsymbol{\epsilon}^*(t_1)\mathbf{b}_k\boldsymbol{\eta}_l^*\boldsymbol{\epsilon}(t_2)\boldsymbol{\epsilon}^*(t_2)\mathbf{b}_l\right. \\
&\qquad\qquad\qquad\left. +\boldsymbol{\eta}_k^*\boldsymbol{\epsilon}(t_1)\mathbf{s}^*(t_1)\mathbf{e}_k\boldsymbol{\eta}_l^*\boldsymbol{\epsilon}(t_2)\mathbf{s}^*(t_2)\mathbf{e}_l\right\} \\
= &\ \frac{1}{M^2}\sum_{t_1=1}^{M}\sum_{t_2=1}^{M}\{\boldsymbol{\eta}_k^*\mathbf{Q}(0)\mathbf{b}_k\boldsymbol{\eta}_l^*\mathbf{Q}(0)\mathbf{b}_l+\boldsymbol{\eta}_k^*\mathbf{Q}(t_1-t_2)\bar{\boldsymbol{\eta}}_l\mathbf{b}_k^\top\mathbf{Q}(t_1-t_2)\mathbf{b}_l \\
&+\boldsymbol{\eta}_k^*\mathbf{Q}(t_1-t_2)\mathbf{b}_l\boldsymbol{\eta}_l^*\mathbf{Q}(t_2-t_1)\mathbf{b}_k+\boldsymbol{\eta}_k^*\mathbf{Q}(t_1-t_2)\bar{\boldsymbol{\eta}}_l\mathbf{e}_k^*\overline{\mathbf{P}}(t_1-t_2)\mathbf{e}_l\}
\end{aligned}
\tag{A.23}
$$

Note that $\mathbf{Q}(0)=\sigma^2\mathbf{I}_p$, also $\boldsymbol{\eta}_k^*\mathbf{b}_k=0$ from corollary 2 to Lemma 1. Hence the first term in the right hand side of (A.23) vanishes. For the remaining terms Substituting $\tau=t_1-t_2$ and taking $M\to\infty$ we get (A.19). Note that we used $\mathbf{Q}(\tau)=\mathbf{Q}^\top(-\tau)$. Similarly,

$$
\begin{aligned}
&\mathbf{E}\left\{\left(\boldsymbol{\eta}_k^*\widehat{\mathbf{R}}\mathbf{b}_k\right)\left(\boldsymbol{\eta}_l^*\widehat{\mathbf{R}}\mathbf{b}_l\ \right)^*\right\} \\
= &\ \frac{1}{M^2}\sum_{t_1=1}^{M}\sum_{t_2=1}^{M}\mathbf{E}\left\{\boldsymbol{\eta}_k^*\boldsymbol{\epsilon}(t_1)\boldsymbol{\epsilon}^*(t_1)\mathbf{b}_k\mathbf{b}_l^*\boldsymbol{\epsilon}(t_2)\boldsymbol{\epsilon}^*(t_2)\boldsymbol{\eta}_l\right. \\
&\qquad\qquad\qquad\left. +\boldsymbol{\eta}_k^*\boldsymbol{\epsilon}(t_1)\mathbf{s}^*(t_1)\mathbf{e}_k\mathbf{e}_l^*\mathbf{s}(t_2)\boldsymbol{\epsilon}^*(t_2)\boldsymbol{\eta}_l\right\} \\
= &\ \frac{1}{M^2}\sum_{t_1=1}^{M}\sum_{t_2=1}^{M}\{\boldsymbol{\eta}_k^*\mathbf{Q}(0)\mathbf{b}_k\,\mathbf{b}_l^*\mathbf{Q}(0)\boldsymbol{\eta}_l+\boldsymbol{\eta}_k^*\mathbf{Q}(t_1-t_2)\boldsymbol{\eta}_l\,\mathbf{b}_k^\top\mathbf{Q}(t_1-t_2)\bar{\mathbf{b}}_l \\
&+\boldsymbol{\eta}_k^*\mathbf{Q}(t_1-t_2)\bar{\mathbf{b}}_l\,\mathbf{b}_k^\top\mathbf{Q}(t_1-t_2)\boldsymbol{\eta}_l+\boldsymbol{\eta}_k^*\mathbf{Q}(t_1-t_2)\boldsymbol{\eta}_l\,\mathbf{e}_k^*\mathbf{P}^\top(t_1-t_2)\mathbf{e}_l\}
\end{aligned}
\tag{A.24}
$$

Arguing as before, (A.20) follows from (A.24). ∎

B Noise covariance function

In this appendix we characterize the matrices $\mathbf{Q}_r(\tau)$ and $\mathbf{Q}_c(\tau)$. We also derive some properties which are useful in the accuracy analysis. If we assume $\sigma=1$ from (3.19) and (3.27) it follows that

$$\mathbf{Q}_r(\tau)=\mathbf{E}\boldsymbol{\epsilon}_r(t+\tau)\boldsymbol{\epsilon}_r^\top(t),\qquad \mathbf{Q}_c(\tau)=\mathbf{E}\boldsymbol{\epsilon}_c(t+\tau)\boldsymbol{\epsilon}_c^\top(t),\tag{B.1}$$

So for convenience we assume $\sigma=1$ in this appendix. Examine the matrix sequence $\{\mathbf{Q}_r(\tau)\}$ introduced in (3.19). Recall that

$$\mathbf{Q}_r(\tau)=\mathbf{E}\boldsymbol{\epsilon}_r(t+\tau)\boldsymbol{\epsilon}_r^*(t)=\frac{1}{4}\mathbf{E}\left\{\boldsymbol{\epsilon}_c(t+\tau)+\boldsymbol{\epsilon}_b(t+\tau)\right\}\left\{\boldsymbol{\epsilon}_c(t)+\boldsymbol{\epsilon}_b(t)\right\}^\top\tag{B.2}$$

It is clear from the construction of $\boldsymbol{\epsilon}_c(t)$ and $\boldsymbol{\epsilon}_b(t)$ in (2.7) and (2.17) that $\mathbf{Q}_r(\tau)=\mathbf{0}_{m\times m}$ for $|\tau|\geq 2m$. Introduce $m|m$ Toeplitz matrices $\{\mathbf{J}_\tau\}_{\tau=-m-1}^{m-1}$ such that

$$[\mathbf{J}_\tau]_{ij}=\begin{cases}1, & j-i=\tau\\ 0 & \text{otherwise}\end{cases}\tag{B.3}$$

Then it is straightforward to show that for $|\tau| < m-1$

$$\mathbf{Q}_c(\tau) := \mathbf{E}\boldsymbol{\epsilon}_c(t+\tau)\boldsymbol{\epsilon}_c^*(t) = \mathbf{J}_\tau, \quad \mathbf{E}\boldsymbol{\epsilon}_b(t+\tau)\boldsymbol{\epsilon}_b^*(t) = \mathbf{J}_{-\tau}, \tag{B.4}$$

In order to examine the cross-correlation between $\boldsymbol{\epsilon}_c(t)$ and $\boldsymbol{\epsilon}_b(t)$ we see that

$$\mathbf{E}\boldsymbol{\epsilon}_b(t+\tau)\boldsymbol{\epsilon}_c^*(t) = \begin{cases} \mathbf{K}_\tau, & 0 < \tau < 2m \\ \mathbf{0}_{m\times m} & \text{otherwise,} \end{cases} \tag{B.5}$$

where $\{\mathbf{K}_\tau\}_{\tau=1}^{2m-1}$ are Hankel matrices such that

$$[\mathbf{K}_\tau]_{ij} = \begin{cases} 1, & i+j-1=\tau \\ 0, & \text{otherwise.} \end{cases} \tag{B.6}$$

Combining (B.2), (B.4) and (B.5) it follows that

$$\mathbf{Q}_r(\tau) = \begin{cases} \frac{1}{2}\mathbf{J}_0 = \frac{1}{2}\mathbf{I}_m, & \tau = 0 \\ \frac{1}{4}\left\{\mathbf{J}_\tau + \mathbf{J}_{-\tau} + \mathbf{K}_{|\tau|}\right\} & 0 < |\tau| < m \\ \frac{1}{4}\mathbf{K}_{|\tau|} & m \le |\tau| < 2m \\ \mathbf{0}_{m\times m} & \text{otherwise.} \end{cases} \tag{B.7}$$

Let us define

$$\mathbf{X}_r^{(k)} := \sum_{\tau=-2m+1}^{2m-1} \mathbf{Q}_r(\tau)\cos(\omega_k\tau). \tag{B.8}$$

Then it can be easily verified that the element at jth row and lth column of $\mathbf{X}_r^{(k)}$ is given by

$$\begin{aligned} \left[\mathbf{X}_r^{(k)}\right]_{jl} &= \frac{1}{2}\cos\{(j+l-1)\omega_k + (\cos\{(j-l)\omega_k\}) \\ &= \cos\left\{\left(j-\frac{1}{2}\right)\omega_k\right\}\cos\left\{\left(l-\frac{1}{2}\right)\omega_k\right\}. \end{aligned} \tag{B.9}$$

Let $\mathbf{a}_k$ is the kth column of $\mathbf{A}_r$. Then from (2.14) it is straightforward that $\mathbf{X}_r^{(k)} = \mathbf{a}_k\mathbf{a}_k^\top$. Note that $\boldsymbol{\eta}_k$ is orthogonal to the columns of $\mathbf{A}$ from (A.10). Hence comparing definitions (3.17) and (A.1) it is clear that $\boldsymbol{\xi}_k$ is orthogonal to the columns of $\mathbf{A}_r$. Hence it follows readily that

$$\boldsymbol{\xi}_k^\top\left\{\sum_{\tau=-2m+1}^{2m-1}\mathbf{Q}_r(\tau)\cos(\omega_k\tau)\right\}\boldsymbol{\xi}_l = 0, \quad 1 \le k, l \le n. \tag{B.10}$$

A similar identity holds for CDM also. Let us define

$$\mathbf{X}_c^{(k)} =:= \sum_{\tau=-m+1}^{m-1}\mathbf{Q}_c(\tau)\mathrm{e}^{\mathrm{i}\omega_k\tau} \tag{B.11}$$

Then it can be easily verified that the element at jth row and lth column of $\mathbf{X}_c^{(k)}$ is given by

$$\left[\mathbf{X}_c^{(k)}\right]_{jl} = \mathrm{e}^{\mathrm{i}(j-l)\omega_k}$$

Let $\boldsymbol{a}_k$ denote the kth column of $\mathbf{A}_c$. Then from (2.4) it is straightforward that $\mathbf{X}_c^{(k)} = \boldsymbol{a}_k \boldsymbol{a}_k^*$. Now comparing definitions (3.25) and (A.1) it is clear that $\boldsymbol{\zeta}_k$ is orthogonal to the columns of $\mathbf{A}_c$. Hence it follows readily that

$$\boldsymbol{\zeta}_k^* \left\{ \sum_{\tau=-m+1}^{m-1} \mathbf{Q}_c(\tau) \mathrm{e}^{\mathrm{i}\omega_k \tau} \right\} \boldsymbol{\zeta}_l = 0, \quad 1 \leq k, l \leq n. \tag{B.12}$$

Similarly,

$$\boldsymbol{\zeta}_k^* \left\{ \sum_{\tau=-m+1}^{m-1} \mathbf{Q}_c(\tau) \mathrm{e}^{-\mathrm{i}\omega_k \tau} \right\} \overline{\boldsymbol{\zeta}}_l = \boldsymbol{\zeta}_k^* \boldsymbol{a}_{k+n} \boldsymbol{a}_{k+n}^* \overline{\boldsymbol{\zeta}}_l = 0, \quad 1 \leq k, l \leq n. \tag{B.13}$$

References

[1] A. Eriksson and P. Stoica. Optimally weighted ESPRIT for direction estimation. *Signal Processing*, 38:223–229, 1994.

[2] A Eriksson, P Stoica, and T Söderström. “Markov-based eigenanalysis method for frequency estimation”. *IEEE Transactions on Signal Processing*, 42:3:568–594, March 1994.

[3] G. H. Golub and C. F. Van Loan. *Matrix Computations, 2nd ed.* John Hopkins University Press, Baltimore, MD, USA, 1989.

[4] M. Haardt. Structured least squares to improve the preformance of ESPRIT-type algorithms. *IEEE Transactions on Signal Processing*, 45(3):792–799, March 1997.

[5] B. Haldar and T. Kailath. Efficient estimation of closely spaced sinusoidal frequencies using subspace based methods. *IEEE Signal Processing Letters*, 4:2:49–51, 1997.

[6] S. Van Huffel and J. Vandewalle. *The Total Least Squares Problem: Computational Aspects and Analysis.* SIAM, Philadelphia, 1991.

[7] P. Janssen and P. Stoica. On the expectation of the product of four matrix-valued gaussian random variables. *IEEE Trans. Automatic Control*, AC-33:867–870, 1988.

[8] M Jansson and P Stoica. Forward-only and forward-backward sample covariances - acomparative study. *Signal Processing*, 77:235–245, 1999.

[9] S. M. Kay. *Modern Spectral Estimation.* Prentice-Hall, Upper Saddle River, NJ, 1988.

[10] M. Kristensson, M. Jansson, and B. Ottersten. Further results and insights on subspace based sinusoidal frequency estimation. *IEEE Transactions on Signal Processing*, 49(12):2962–2974, December 2001.

[11] H. Li, J. Li, and P. Stoica. Performance analysis of forward-backward matched-filterbank spectral estimators. *IEEE Trans. Signal Processing*, 46:1954–1966, 1998.

[12] L. Ljung. *System Identification - Theory for the User, 2nd edition.* Prentice Hall, Upper Saddle River, NJ, USA, 1999.

[13] J. M. Mendel. *Lessons in Estimation Theory for Signal Processing, Communications and Control.* Prentice Hall, Englewood Cliffs, New Jersey, 1995.

[14] B. Ottersten, M. Viberg, and T. Kailath. Performance analysis of the total least squares ESPRIT algorithm. *IEEE Trans. on Signal Processing*, 39:4:1122–1135, May 1991.

[15] A. Papoulis. *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, New York, 1991.

[16] A. Paulraj, R. Roy, and T. Kailath. A subspace rotation approach to signal parameter estimation. *Proceedings of IEEE*, 74:7:1044–1045, July 1986.

[17] B. D. Rao and K. V. S. Hari. Weighted subspace methods and spatial smoothing: analysis and comparison. *IEEE Transactions on Signal Processing*, 41:2:788 –803, February 1993.

[18] R. Roy and T. Kailath. ESPRIT-Estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37:7:984–995, July 1989.

[19] T. Söderström and K. Mahata. On instrumental variable and total least squares approaches for identification of noisy systems. *International Journal of Control*, 75(6):381–389, April 2002.

[20] T. Söderström and P. Stoica. *System Identification.* Prentice Hall International, Hemel Hempstead, UK, 1989.

[21] G W Stewart and Ji-guang Sun. *Matrix Perturbation Theory.* Academic Press, Boston, MA, USA, 1990.

[22] P Stoica and A Eriksson. MUSIC estimation of real-valued sinewave frequencies. *Signal Processing*, 42:4:139–146, April 1995.

[23] P. Stoica and R. Moses. *Introduction to Spectral Analysis.* Prentice-Hall, Upper Saddle River, NJ, 1997.

[24] P. Stoica and K. C. Sharman. Maximum likelihood methods for direction-of-arival estimation. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38(7):1132–1143, July 1990.

[25] P Stoica and T Söderström. "Statistical analysis of MUSIC and subspace rotation estimates of sinusoidal frequencies". *IEEE Transactions on Signal Processing*, 39:1836–1847, August 1991.

[26] P. Stoica and M. Viberg. Weighted LS and TLS approaches yield asymptotically equivalent results. *Signal Processing*, 43:255–259, 1995.

[27] Ji-guang Sun. Stability and Accuracy. Perturbation Analysis of Algebraic Eigenproblems (book manuscript). Technical Report UMINF 98-07, Umeå University, Department of Computing Science, August 1998.