Research Article

Sinusoidal Order Estimation Using Angles between Subspaces

Mads Græsbøll Christensen,[1] Andreas Jakobsson (EURASIP Member),[2] and Søren Holdt Jensen (EURASIP Member)[3]

[1] *Department of Media Technology, Aalborg University, Niels Jernes Vej 14, 9220 Aalborg, Denmark*
[2] *Department of Mathematical Statistics, Lund University, 221 00 Lund, Sweden*
[3] *Department of Electronic Systems, Aalborg University, Niels Jernes Vej 12, 9220 Aalborg, Denmark*

Correspondence should be addressed to Mads Græsbøll Christensen, mgc@imi.aau.dk

We consider the problem of determining the order of a parametric model from a noisy signal based on the geometry of the space. More specifically, we do this using the nontrivial angles between the candidate signal subspace model and the noise subspace. The proposed principle is closely related to the subspace orthogonality property known from the MUSIC algorithm, and we study its properties and compare it to other related measures. For the problem of estimating the number of complex sinusoids in white noise, a computationally efficient implementation exists, and this problem is therefore considered in detail. In computer simulations, we compare the proposed method to various well-known methods for order estimation. These show that the proposed method outperforms the other previously published subspace methods and that it is more robust to the noise being colored than the previously published methods.

1. Introduction

Estimating the order of a model is a central, yet commonly overlooked, problem in parameter estimation, with the majority of literature assuming prior knowledge of the model order. In many cases, however, the order cannot be known a priori and may change over time. This is the case, for example, in speech and audio signals. Many parameter estimation methods, like the maximum likelihood and subspace methods, require that the order is known to work properly. The consequence of choosing an erroneous order, aside from the size of the parameter set being wrong, is that the found parameters may be biased or suffer from a huge variance. The most commonly used methods for estimating the model order are perhaps the minimum description length (MDL) [1, 2], the Akaike information criterion (AIC) [3], and the maximum a posteriori (MAP) rule of [4]. These methods are based on certain asymptotic approximations and on statistical models of the observed signal, like the noise being white and Gaussian distributed. We refer the interested reader to [4, 5] for an overview of such statistical methods. A notable feature of the MAP

rule of [4] is that it shows that linear and nonlinear parameters should be penalized differently, something that not recognized by many prior methods (on this topic, see also [6]). In this paper, we are concerned with a more specific, yet important, case, namely, that of finding the number of complex sinusoids buried in noise. This problem is treated in great detail from a statistical point of view in [4] and is also exemplified in [5] and other notable approaches include those of [7–13]. A different class of methods is subspace methods, which is also the topic of interest here. In subspace methods, the eigenvectors of the covariance matrix are divided into a set that spans the space of the signal of interest, called the signal subspace, and its orthogonal complement, the noise subspace. These subspaces and their properties can then be used for various estimation and identification tasks. Subspace methods have a rich history in parameter estimation and signal enhancement. Especially for the estimation of sinusoidal frequencies and finding the direction of arrival of sources in array processing, these methods have proven successful during the past three decades. The most common subspace methods for parameter estimation are perhaps the MUSIC (MUltiple SIgnal Classification) method

[14, 15] and the ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) method of [16] while the earliest example of such methods is perhaps Pisarenko's method [17]. In the context of subspace methods, the typical way of finding the dimensions of the signal and noise subspaces is based on statistical principles where the likelihood function of the observation vector is combined with one of the aforementioned order selection rules with the likelihood function depending on the ratio between the arithmetic and geometric means of the eigenvalues [18, 19]. Recently, the underlying principles of ESPRIT and MUSIC have been extended to the problem of order estimation by exploiting the properties of the eigenvectors rather than the eigenvalues. Compared to the order estimation techniques based on the eigenvalues, one can interpret these methods as being based on the geometry of the space rather than the distribution of energy. Specifically, two related subspace methods based on ESPRIT have been proposed, namely, to the ESTimation ERror (ESTER) method [20] and the Subspace-based Automatic Model Order Selection (SAMOS) method [21]. Similarly, it was shown in [22] that the orthogonality principle of MUSIC can be used for finding the number of harmonics for a set of harmonically related sinusoids when normalized appropriately. See also [23] for a comparison of this method with the ESTER and SAMOS methods. An attractive property of the subspace-based order estimation criteria is that they do not require prior knowledge of the probability density function (pdf) of the observation noise but only a consistent covariance matrix estimate. This means that the subspace methods will work in situations where the statistical methods may fail due to the assumed pdf not being a good approximation of the observed data. Furthermore, it can be quite difficult to derive a method like the MAP rule of [4] for complicated signal models.

Mathematically, the specific problem considered herein can be stated as follows. A signal consisting of complex sinusoids having frequencies $\{\omega_l\}$ is corrupted by additive noise, $\epsilon(n)$, for $n = 0, \ldots, N - 1$,

$$x(n) = \sum_{l=1}^{L} A_l e^{j(\omega_l n + \phi_l)} + \epsilon(n), \qquad (1)$$

where $A_l > 0$ and ϕ_l are the amplitude and the phase of the lth sinusoid. Here, $\epsilon(n)$ is assumed to be white complex symmetric zero-mean noise. The problem considered is then how to estimate the model order L. The model in (1) may seem a bit restrictive, but the proposed method can in fact be used for more general problems. Firstly, the proposed method is valid for a large class of signal models; however, for the case of complex exponentials a computationally efficient implementation of our method exists. This is also the case for damped sinusoids where the principles of unitary ESPRIT may be applied [24]. Secondly, for the case of colored noise, the proposed method is also applicable by the use of prewhitening.

In this paper, we study the problem of finding the model order using the angles between a candidate signal subspace and the signal subspace in depth. In the process of finding the model order, nonlinear model parameters are also found.

The concept of angles between subspaces has previously been applied within the field of signal processing to, among other things, analysis of subspace-based enhancement algorithms, for example, [25, 26], and multipitch estimation [27]. For complex sinusoids, the measure based on angles between subspaces reduces to a normalization of the well-known cost function first proposed for frequency and direction-of-arrival estimation in [14] for a high number of observations. We analyze, discuss, and compare the measure and its properties to other commonly used measures of the angles between subspaces and show that the proposed measure provides an upper bound for some other more complicated measures. These other measures turn out to be less useful for our application, and, in simulations, we compare the proposed method to other methods for finding the number of complex sinusoids. Our results show that the method has comparable performance to commonly used methods and is generally best among the subspace-based methods. It is also demonstrated, however, that the method is more robust to model violations, like colored noise. As an aside, our results also establish the MUSIC criterion for parameter estimation [14] as an approximation to the angles between the noise and candidate model subspaces.

The remaining part of this paper is organized as follows. First, we recapitulate the covariance matrix model that forms the basis for the subspace methods and briefly describe the MUSIC method in Section 2. In Section 3, we then move on to derive the new measure based on angles between subspaces. We relate this measure to other similar measures and proceed to discuss its properties and application to the problem interest. The statistical performance of the method is then evaluated in simulations studies in Section 4 and compared to a number of related parametric and nonparametric methods and, in Section 5, the results are discussed. Finally, we conclude on our work in Section 6.

2. Fundamentals

We start out this section by presenting some fundamental definitions, relations, and results. First, we define $\mathbf{x}(n)$ as a signal vector, referred to as a subvector, containing $M < N$ samples of the observed signal, that is,

$$\mathbf{x}(n) = \begin{bmatrix} x(n) & x(n+1) & \cdots & x(n+M-1) \end{bmatrix}^T \qquad (2)$$

with $(\cdot)^T$ denoting the transpose. Assuming that the phases of the sinusoids are independent and uniformly distributed on the interval $(-\pi, \pi)$, the covariance matrix $\mathbf{R} \in \mathbb{C}^{M \times M}$ of the signal in (1) can be written as [5]

$$\mathbf{R} = \mathrm{E}\left\{\mathbf{x}(n)\mathbf{x}^H(n)\right\} = \mathbf{A}\mathbf{P}\mathbf{A}^H + \sigma^2 \mathbf{I}_M, \qquad (3)$$

where $\mathrm{E}\{\cdot\}$ and $(\cdot)^H$ denote the statistical expectation and the conjugate transpose, respectively. We here require that $L < M$. Moreover, we note that for the above to hold, the noise need not be Gaussian. The matrix $\mathbf{P}$ is diagonal and contains the squared amplitudes, that is,

$\mathbf{P} = \text{diag}\left(\begin{bmatrix} A_1^2 & \cdots & A_L^2 \end{bmatrix}\right)$, and $\mathbf{A} \in \mathbb{C}^{M \times L}$ is a Vandermonde matrix defined as

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}(\omega_1) & \cdots & \mathbf{a}(\omega_L) \end{bmatrix}, \tag{4}$$

where $\mathbf{a}(\omega) = \begin{bmatrix} 1 e^{j\omega} & \cdots & e^{j\omega(M-1)} \end{bmatrix}^T$. Also, σ^2 denotes the variance of the additive noise, $\epsilon(n)$, and $\mathbf{I}_M$ is the $M \times M$ identity matrix. Assuming that the frequencies $\{\omega_l\}$ are distinct, the columns of $\mathbf{A}$ are linearly independent and $\mathbf{A}$ and $\mathbf{A}\mathbf{P}\mathbf{A}^H$ have rank L. Let

$$\mathbf{R} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^H \tag{5}$$

be the eigenvalue decomposition (EVD) of the covariance matrix. Then, $\mathbf{Q}$ contains the M orthonormal eigenvectors of $\mathbf{R}$, that is, $\mathbf{Q} = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_M \end{bmatrix}$ and $\boldsymbol{\Lambda}$ is a diagonal matrix containing the corresponding eigenvalues, λ_k, with

$$\lambda_1 \geq \cdots \geq \lambda_L \geq \lambda_{L+1} = \cdots = \lambda_M = \sigma^2. \tag{6}$$

The subspace-based methods are based on a partitioning of the eigenvectors into a set belonging to the signal subspace spanned by the columns of $\mathbf{A}$ and its orthogonal complement known as the noise subspace. Let $\mathbf{S}$ be formed from the eigenvectors corresponding to the L most significant eigenvalues, that is,

$$\mathbf{S} = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_L \end{bmatrix}. \tag{7}$$

We denote the space spanned by the columns of $\mathbf{S}$ as $\mathcal{R}(\mathbf{S})$ and henceforth refer to it as the signal subspace. Similarly, let $\mathbf{G}$ be formed from the eigenvectors corresponding to the $M - L$ least significant eigenvalues, that is,

$$\mathbf{G} = \begin{bmatrix} \mathbf{q}_{L+1} & \cdots & \mathbf{q}_M \end{bmatrix}, \tag{8}$$

where $\mathcal{R}(\mathbf{G})$ is referred to as the noise subspace. Using the EVD in (5), the covariance matrix model in (3) can now be written as $\mathbf{Q}(\boldsymbol{\Lambda} - \sigma^2\mathbf{I}_M)\mathbf{Q}^H = \mathbf{A}\mathbf{P}\mathbf{A}^H$. Introducing $\boldsymbol{\Lambda}_S = \text{diag}\left(\begin{bmatrix} \lambda_1 - \sigma^2 & \cdots & \lambda_L - \sigma^2 \end{bmatrix}\right)$, we can write this as

$$\mathbf{S}\boldsymbol{\Lambda}_S\mathbf{S}^H = \mathbf{A}\mathbf{P}\mathbf{A}^H. \tag{9}$$

From the last equation, it can be seen that the columns of $\mathbf{A}$ span the same space as the columns of $\mathbf{S}$ and that $\mathbf{A}$ therefore also must be orthogonal to $\mathbf{G}$, that is,

$$\mathbf{A}^H\mathbf{G} = \mathbf{0}. \tag{10}$$

In practice, the eigenvectors are of course unknown and are replaced by estimates. Here, we will estimate the covariance matrix as

$$\hat{\mathbf{R}} = \frac{1}{N-M+1}\sum_{n=0}^{N-M}\mathbf{x}(n)\mathbf{x}^H(n), \tag{11}$$

which is a consistent estimate for ergodic processes and the maximum likelihood estimate for Gaussian noise. The eigenvector estimates obtained from this matrix are then

also consistent and the covariance matrix model (3) and the orthogonality property (10) therefore hold asymptotically.

Since the covariance matrix and eigenvectors are estimated from a finite set of vectors, the orthogonality property in (10) only holds approximately. In the MUSIC algorithm [14, 15], the set of distinct frequencies $\{\omega_l\}$ are found by minimizing the Frobenius norm, denoted $\| \cdot \|_F$, of (10), that is,

$$\{\hat{\omega}_l\} = \arg\min_{\{\omega_l\}}\left\|\mathbf{A}^H\mathbf{G}\right\|_F^2. \tag{12}$$

Since the squared Frobenius norm is additive over the columns of $\mathbf{A}$, we can find the individual sinusoidal frequencies for $l = 1, \dots, L$ as

$$\hat{\omega}_l = \arg\min_{\omega_l}\left\|\mathbf{a}^H(\omega_l)\mathbf{G}\right\|_F^2 \tag{13}$$

with the requirements that the frequencies are distinct and fulfill the two following conditions:

$$\frac{\partial\left\|\mathbf{a}^H(\omega_l)\mathbf{G}\right\|_F^2}{\partial\omega_l} = 0, \qquad \frac{\partial^2\left\|\mathbf{a}^H(\omega_l)\mathbf{G}\right\|_F^2}{\partial\omega_l^2} > 0. \tag{14}$$

The reciprocal form of the cost function in (13) is sometimes referred to as spectral MUSIC and $1/\|\mathbf{a}^H(\omega_l)\mathbf{G}\|_F^2$ as the pseudospectrum from which the L frequencies are obtained as the peaks. We mention in passing that it is possible to solve (13) using numeric rooting methods [28] or FFTs. Regarding the statistical properties of MUSIC, the effects of order estimation errors, that is, the effect of choosing an erroneous $\mathbf{G}$ in (13), on the parameter estimates obtained using MUSIC have been studied in [29] in a slightly different context and it was concluded that the MUSIC estimator is more sensitive to underestimation of L than overestimation. The more common case of L being known has been treated in great detail, with the statistical properties of MUSIC having been studied in [30–34].

3. Angles between Subspaces

3.1. Definition and Basic Results. The orthogonality property states that for the true parameters, the matrix $\mathbf{A}$ is orthogonal to the noise subspace eigenvectors in $\mathbf{G}$. For estimation purposes, we need a measure of this. The concept of orthogonality is of course closely related to the concept of angles, and how to define angles in multidimensional spaces is what we will now investigate further.

The principal (nontrivial) angles $\{\theta_k\}$ between the two subspaces $\mathcal{A} = \mathcal{R}(\mathbf{A})$ and $\mathcal{G} = \mathcal{R}(\mathbf{G})$ are defined recursively for $k = 1, \dots, K$ as (see, e.g., [35])

$$\cos(\theta_k) = \max_{\mathbf{u}\in\mathcal{A}}\max_{\mathbf{v}\in\mathcal{G}}\frac{\mathbf{u}^H\mathbf{v}}{\|\mathbf{u}\|_2\|\mathbf{v}\|_2} \triangleq \mathbf{u}_k^H\mathbf{v}_k. \tag{15}$$

The quantity K is the minimal dimension of the two subspaces, that is, $K = \min\{L, M - L\}$, which is the number of nontrivial angles between the two subspaces. Moreover, the directions along which the angles are defined

are orthogonal, that is, $\mathbf{u}^H \mathbf{u}_i = 0$ and $\mathbf{v}^H \mathbf{v}_i = 0$ for $i = 1, \ldots, k - 1$.

We will now rewrite (15) into something more useful, and in doing this, we will make extensive use of projection matrices. The (orthogonal) projection matrix for a subspace $\mathcal{X}$ spanned by the columns of a matrix $\mathbf{X}$ is defined as $\mathbf{\Pi}_X = \mathbf{X}(\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H$. Such projections matrices are Hermitian, that is, $\mathbf{\Pi}_X^H = \mathbf{\Pi}_X$ and have the properties $\mathbf{\Pi}_X^m = \mathbf{\Pi}_X$ for $m = 1, 2, \ldots$ and $\|\mathbf{\Pi}_X\|_F^2 = \dim(\mathcal{X})$ where $\dim(\cdot)$ is the dimension of the subspace. Let $\mathbf{\Pi}_G$ be the projection matrix for subspace $\mathcal{G}$, and the $\mathbf{\Pi}_A$ the projection matrix for subspace $\mathcal{A}$. Using the two projection matrices, we can write the vector $\mathbf{u} \in \mathcal{A}$ as $\mathbf{\Pi}_A \mathbf{y}$ and $\mathbf{v} \in \mathcal{G}$ as $\mathbf{\Pi}_G \mathbf{z}$ with $\mathbf{y}, z \in \mathbb{C}^M$. This allows us to express (15) as

$$\cos(\theta_k) = \max_{\mathbf{y} \in \mathbb{C}^M} \max_{\mathbf{z} \in \mathbb{C}^M} \frac{\mathbf{y}^H \mathbf{\Pi}_A \mathbf{\Pi}_G \mathbf{z}}{\|\mathbf{y}\|_2 \|\mathbf{z}\|_2} \tag{16}$$
$$\triangleq \mathbf{y}_k^H \mathbf{\Pi}_A \mathbf{\Pi}_G \mathbf{z}_k = \sigma_k$$

for $k = 1, \ldots, K$. Again, we require that $\mathbf{y}^H \mathbf{y}_i = 0$ and $\mathbf{z}^H \mathbf{z}_i = 0$ for $i = 1, \ldots, k - 1$, that is, that the vectors are orthogonal. Futhermore, the denominator ensures that the vectors have unit norm. It then follows that $\{\sigma_k\}$ are the singular values of the matrix product $\mathbf{\Pi}_A \mathbf{\Pi}_G$, and that the two sets of vectors $\{\mathbf{y}\}$ and $\{\mathbf{z}\}$ are the left and right singular vectors, respectively. Regarding the mapping of the singular values to actual angles, a difficult problem, we refer the interested reader to [36] for a numerically stable algorithm.

The set of principal angles obey the following inequality:

$$0 \leq \theta_1 \leq \cdots \leq \theta_K \leq \frac{\pi}{2}. \tag{17}$$

Next, the singular values are related to the Frobenius norm of the product $\mathbf{\Pi}_A \mathbf{\Pi}_G$ as

$$\|\mathbf{\Pi}_A \mathbf{\Pi}_G\|_F^2 = \mathrm{Tr}\{\mathbf{\Pi}_A \mathbf{\Pi}_G\} = \sum_{k=1}^{K} \sigma_k^2, \tag{18}$$

and therefore also to the angles between the subspaces, that is,

$$\sum_{k=1}^{K} \cos^2(\theta_k) = \|\mathbf{\Pi}_A \mathbf{\Pi}_G\|_F^2. \tag{19}$$

3.2. A Simplified Measure. We will now show how the concepts introduced in the previous section can be simplified for use in estimation. The Frobenius norm of the product $\mathbf{\Pi}_A \mathbf{\Pi}_G$ can be expressed as

$$\|\mathbf{\Pi}_A \mathbf{\Pi}_G\|_F^2 = \mathrm{Tr}\left\{\mathbf{\Pi}_A \mathbf{\Pi}_G \mathbf{\Pi}_G^H \mathbf{\Pi}_A^H\right\} = \mathrm{Tr}\left\{\mathbf{\Pi}_A \mathbf{\Pi}_G^H\right\} \tag{20}$$

$$= \mathrm{Tr}\left\{\mathbf{A}\left(\mathbf{A}^H \mathbf{A}\right)^{-1} \mathbf{A}^H \mathbf{G} \mathbf{G}^H\right\}. \tag{21}$$

This expression can be seen to be complicated since it involves matrix inversion and it does not decouple the problem of estimating the parameters of the column of $\mathbf{A}$. Additionally, it is not related to the MUSIC cost function in

a simple way. It can, though, be simplified in the following way. The columns of $\mathbf{A}$ consist of complex sinusoids, and for any distinct set of frequencies these are asymptotically orthogonal, meaning that

$$\lim_{M \to \infty} M \mathbf{\Pi}_A = \lim_{M \to \infty} M \mathbf{A}\left(\mathbf{A}^H \mathbf{A}\right)^{-1} \mathbf{A}^H \tag{22}$$
$$= \mathbf{A} \mathbf{A}^H.$$

We can now simplify (21) and manipulate it into a familiar form, that is,

$$\|\mathbf{\Pi}_A \mathbf{\Pi}_G\|_F^2 = \mathrm{Tr}\left\{\mathbf{A}\left(\mathbf{A}^H \mathbf{A}\right)^{-1} \mathbf{A}^H \mathbf{G} \mathbf{G}^H\right\}$$
$$\approx \frac{1}{M} \mathrm{Tr}\left\{\mathbf{A}^H \mathbf{G} \mathbf{G}^H \mathbf{A}\right\} = \frac{1}{M} \left\|\mathbf{A}^H \mathbf{G}\right\|_F^2, \tag{23}$$

which, except for the scaling $1/M$, is the reciprocal of the original MUSIC cost function as introduced in [14]. From (19) and (23), we get

$$\frac{1}{M} \left\|\mathbf{A}^H \mathbf{G}\right\|_F^2 \approx \sum_{k=1}^{K} \cos^2(\theta_k). \tag{24}$$

This shows that the original MUSIC cost function can be explained and understood in the context of angles between subspaces. At this point, it must be emphasized that this interpretation only holds for signal models consisting of vectors that are orthogonal or asymptotically orthogonal. Consequently, it holds for sinusoids, for example, but not for damped sinusoids.

We now arrive at a convenient measure of the extent to which the orthogonality property in (10) holds, which is the average over all the principal (nontrivial) angles between $\mathcal{A}$ and $\mathcal{G}$:

$$\frac{1}{K} \sum_{k=1}^{K} \cos^2(\theta_k) = \frac{1}{K} \sum_{k=1}^{K} \sigma_k^2 \approx \frac{1}{MK} \left\|\mathbf{A}^H \mathbf{G}\right\|_F^2 \triangleq J \tag{25}$$

with $K = \min\{L, M - L\}$. This measure is only zero when all angles are $\pi/2$, that is, when the subspaces $\mathcal{A}$ and $\mathcal{B}$ are orthogonal in all directions. Additionally, the intersection of the subspaces is the range of the set of principal vectors for which $\cos(\theta_k) = 1$. Due to the normalization $1/K$, the measure can be seen to be bounded as

$$0 \leq \frac{1}{K} \sum_{k=1}^{K} \cos^2(\theta_k) \leq 1. \tag{26}$$

This bound is also asymptotically valid for the rightmost expression in (25) and is otherwise an approximation for finite lengths. To put the derived measure a bit into perspective, it can, in fact, be brought into a form similar as the aforementioned and well-known statistical methods (MDL, AIC, etc.) by taking the logarithm of (25), that is,

$$\ln J = \ln \left\|\mathbf{A}^H \mathbf{G}\right\|_F^2 - \ln(MK), \tag{27}$$

which consists of two familiar terms: a "goodness of fit" measure and an order-dependent penalty function, which in this case is a nonlinear function of the model order, unlike, for example, MDL and AIC.

3.3. Relation to Other Measures. We will now proceed to relate the derived measure to some other measures. Interestingly, the Frobenius norm of the difference between the two projection matrices can be expressed as

$$\|\mathbf{\Pi}_A - \mathbf{\Pi}_G\|_F^2 = \text{Tr}\{\mathbf{\Pi}_A + \mathbf{\Pi}_G - 2\mathbf{\Pi}_A\mathbf{\Pi}_G\}$$
$$= M - 2\|\mathbf{\Pi}_A\mathbf{\Pi}_G\|_F^2, \quad (28)$$

which shows that minimizing (18) is the same as maximizing the Frobenius norm of the difference between the two projection matrices. This puts the original MUSIC cost function into perspective, as it was originally motivated in [14] as the distance between the subspaces.

In [22], it was proposed to measure the orthogonality using the following normalized Frobenius norm of the matrix product $\mathbf{A}^H\mathbf{G}$:

$$\frac{\|\mathbf{A}^H\mathbf{G}\|_F^2}{LM(M - L)}, \quad (29)$$

which was derived from the Cauchy-Schwarz inequality. A new derivation of the measure in (29) is provided in the appendix in which it is shown that this too can be interpreted as an average over cosine to angles, more specifically, between each vector pair. However, the definition of the angles differs from that of the angles between subspaces, and, as a result, the normalizations differ as well. Clearly, we have that

$$ML(M - L) \geq M\min\{L, M - L\} \quad (30)$$

and thus

$$\frac{\|\mathbf{A}^H\mathbf{G}\|_F^2}{ML(M - L)} \leq \frac{\|\mathbf{A}^H\mathbf{G}\|_F^2}{M\min\{L, M - L\}}. \quad (31)$$

That the two approaches lead to different normalizations may seem like a minor detail, but this is in fact also the fundamental difference between the AIC, MDL, MAP, and so forth, order selection rules. These all provide a different order dependent scaling of the likelihood function. At the very least, the new normalization is mathematically more tractable than the old one. In Figure 1, the two normalizations, namely, $ML(M - L)$ and $M\min\{L, M - L\}$, are shown as a function of L for $M = 50$. Note that the curves have been scaled by their respective maximum values. Interestingly, both the measures defined in (29) and (25), respectively, are consistent with finding the frequencies using (13) in the sense that the frequencies that minimize (13) also minimize either of these measures for a given order L.

The measure in (25) can also be related to some other measures that have been defined in relation to angles between subspaces, like the projection 2-norm [37]. The distance or gap between subspaces, is defined for $L = M/2$ as [35–37]

$$\text{dist}(\mathcal{A}, \mathcal{G}) = \|\mathbf{\Pi}_A - \mathbf{\Pi}_G\|_2 \quad (32)$$

and is related to the concept of angles between subspaces in the sense that (see, e.g., [35])

$$\text{dist}(\mathcal{A}, \mathcal{G}) = \sin(\theta_K) = \sqrt{1 - \cos(\theta_K)}, \quad (33)$$
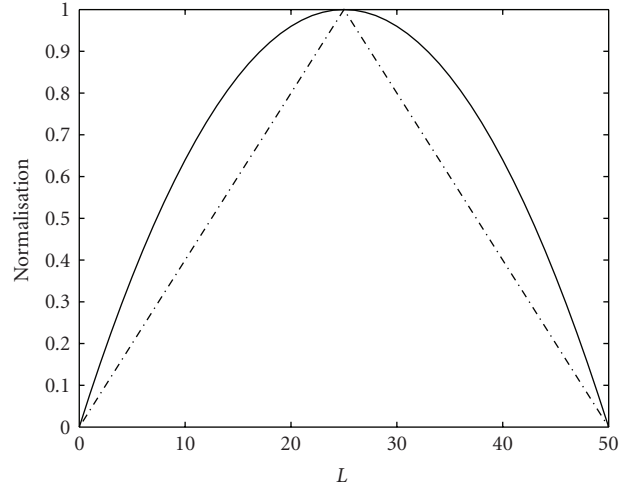

FIGURE 1: Normalization factors (scaled for convenience) as a function of L for the measure in [22] (solid) and based on the theory of angles between subspaces (dash-dotted).

which is given by the Kth singular value of the matrix product $\mathbf{\Pi}_A\mathbf{\Pi}_G$ as $\theta_K = \arccos(\sigma_K)$. Another measure of interest is the minimum principal angle which by the definition in (15) is a function of the maximum singular value as $\theta_1 = \arccos(\sigma_1)$ and is given by the induced matrix 2-norm, that is,

$$\|\mathbf{\Pi}_A\mathbf{\Pi}_G\|_2^2 = \sigma_1^2. \quad (34)$$

In the study of angles between subspaces, there has also been some interest in a different definition of the angle between two subspaces based on exterior algebra. Specifically, this so-called higher dimensional angle θ is related to the principal angles as [38, 39]

$$\cos^p(\theta) = \prod_{k=1}^{K}\cos^p(\theta_k) = \prod_{k=1}^{K}\sigma_k^p \quad (35)$$

for $p = 1, 2, \ldots$, which for $p = 1$ can be interpreted as the volume of a certain matrix [38]. In [40], θ was shown to be an angle in the usual Euclidean sense.

Equations (33), (34), and (35) are not very convenient measures for our purpose since they cannot be calculated from the individual columns of $\mathbf{A}$ but rather depend on all of them. This means that optimization of any of these measures would require multidimensional nonlinear optimization over the frequencies $\{\omega_l\}$.

We will now investigate how the various measures relate to each other, and in doing so, we will arrive at some interesting bounds. First, we note that the arithmetic mean of the singular values can be related to the geometric mean and (35) as

$$\frac{1}{K}\sum_{k=1}^{K}\sigma_k^2 \geq \left(\prod_{k=1}^{K}\sigma_k^2\right)^{1/K} \geq \prod_{k=1}^{K}\sigma_k^2, \quad (36)$$

where the right-most expression follows from $\sigma_k \leq 1$. We can now establish the following set of inequalities that relate

the various measures based on angles between subspaces to the Frobenius norm:

$$\prod_{k=1}^{K}\sigma_k^2 \le \sigma_K^2 \le \sigma_1^2 \le \sum_{k=1}^{K}\sigma_k^2. \tag{37}$$

It follows that the Frobenius norm can be seen as a majorizing function for the other measures. Therefore, finding the frequencies using (12) can be seen to minimize the upper bound of the other measures. Similarly, we obtain the following set of inequalities for the normalized measure involving the average over the squared cosine terms in (19), that is,

$$\prod_{k=1}^{K}\sigma_k^2 \le \sigma_K^2 \le \frac{1}{K}\sum_{k=1}^{K}\sigma_k^2 \le \sigma_1^2. \tag{38}$$

In this case, the normalized Frobenius norm is still an upper bound for two of the measures, but it is lower than or equal to the measure in (19). In this sense, the measure in (19) can be seen as a majorizing function for the measures in (33) and (35). It can be seen from (38) that the measures are identical when all singular values $\{\sigma_k\}$ are either one or zero, that is, when the subspaces have a K dimensional intersection or are orthogonal in all directions. The only measure, however, that ensures orthogonality in all directions for a value of zero, is the proposed measure in (25). Clearly, this is a desirable property for our application.

3.4. Application to Sinusoidal Order Estimation. As can be seen, (10) can only be expected to hold when the eigenvectors of **R** are partitioned into a signal and a noise subspace such that the rank of the signal subspace is equal to the true number of sinusoids. Based on the proposed orthogonality measure, the order is found by evaluating the measure for various candidate orders $1 \le L \le M-1$ and then picking the order for which the measure is minimized, that is,

$$\widehat{L} = \arg\min_L \min_{\{\omega_l\}} \frac{||\mathbf{A}^H\mathbf{G}||_F^2}{MK} \tag{39}$$

$$= \arg\min_L \sum_{l=1}^{L} \min_{\omega_l} \frac{||\mathbf{a}^H(\omega_l)\mathbf{G}||_F^2}{MK} \tag{40}$$

with $K = \min\{L, M-L\}$. As before, the frequencies should be distinct and satisfy (14). The set of candidate orders does not include zero (as no angles can be measured then), meaning that the measure cannot be used for determining whether only noise is present. This is also the case for the related ESTER and SAMOS methods.

3.4.1. Consistency. Regarding the consistency of the proposed method, it can easily be verified that the covariance matrix model and the orthogonality property hold for the noise-free case. We will here make the following simple argument for the consistency of the method for noisy signals based on [31]: since a consistent estimate of the covariance matrix is used, the eigenvector estimates are consistent too and the

covariance matrix model in (3) holds asymptotically in N and M (which is here assumed to be chosen proportional to N) [31, 32]. Therefore, the orthogonality criterion in (10) holds as N tends to infinity. Provided that the sinusoids are linearly independent but not all orthogonal, (10) holds only for the combination of the true set of frequencies $\{\omega_l\}$ and order L. Regarding the finite length performance of MUSIC, it is well known to perform well for high SNR and N being consistent but suboptimal [31, 32] while exhibiting thresholding behavior below certain SNR or number of samples N. This thresholding behavior can largely be attributed to the occurrence of "subspace swapping" [41, 42].

3.4.2. Computational Complexity. The major contributor to the computational complexity of a direct implementation of (40) is the EVD of the covariance matrix, and this is also the case for the ESTER and SAMOS methods and [18, 19]. This can be lessened by the use of recursive computation of the covariance matrix eigenvectors over time, also known as subspace tracking. However, for our method and the ESTER and SAMOS methods, it is critical that a subspace tracker is chosen that tracks the eigenvectors and not just an arbitrary basis of the subspace. The reasons is that a subpartitioning of an arbitrary basis is not necessarily the same as a subpartitioning of the eigenvectors and the methods may therefore fail to provide accurate order estimates. Examples of subspace trackers that are suited for this purpose are, for example, [43–45] (see [46] for more on this). Aside from the EVD, our method also requires nonlinear optimization for finding the frequencies. This is by no means a particular property of our methodl; indeed most other methods for finding the order of the model in (1), including [4, 10–13], require this as well, with the methods of [19, 20] being notable exceptions. For (40), this can be done either by FFTs (see [22, 46]) or by polynomial rooting methods [28]. In the FFT-based implementation of [22], the Fourier transform of the eigenvectors is calculated once per segment and this information is simply reused in the subsequent optimization. The complexity is therefore similar to that of spectral [14] or root MUSIC [28], two methods that have a rich history in spectral estimation and array processing. In practice, the complexity can be reduced considerably by applying certain approximations, that is, by either (1) using the min-norm solution, which can be calculated recursively over the orders, instead of the full noise subspace [5, 47], or by (2) finding approximate solutions using a number of the least significant eigenvectors that are known with certainty to belong to the noise subspace (usually an upper bound on number of possible sinusoids can be identified from the application).

3.4.3. Comparison to ESTER and SAMOS. There appears to be a number of advantages to our method compared to the related methods ESTER and SAMOS that are also based on the eigenvectors. It can be seen from (40) that the method can find orders in a wider range than both the ESTER and SAMOS methods, with those methods being able to find orders in the intervals $1 \le L \le M-2$ and

$1 \leq L \leq (M-1)/2$, respectively. The class of shift-invariant signal models also includes damped sinusoids and the ESTER and SAMOS methods hold also for this model and so does the orthogonality property of MUSIC. At first sight it may appear that an efficient implementation of the nonlinear optimization in (40) does not exist. However, either the rooting approach of [28] may be used or the principle of unitary ESPRIT can be applied by using a forward-backward estimate of the covariance matrix whereby the FFT-based implementation is applicable (see [24]). We here stress that an additional advantage of the MUSIC-based method presented here is that it is more general than those based on the shift-invariance property [20, 21]; that is, the relation (10) can be used for a more general class of signal models. It is, however, not certain that there exits an efficient implementation of the nonlinear optimization required by this approach.

4. Experimental Results

4.1. Details and Reference Methods. We now proceed to evaluate the performance of the proposed estimator (denoted MUSIC (new) in the figures) under various conditions using Monte Carlo simulations comparing to a number of other methods that have appeared in literature. The reference methods are listed in Table 1. It should be noted that the model selection criteria of the MDL [13] and the MAP [4] methods are in fact identical for this problem, although derived from different perspectives. The difference between these two methods is then, essentially, that one uses high-resolution estimates of the frequencies while the other uses the computationally simple periodogram. Note that it is possible to refine the initial frequency estimates obtained from the periodogram in several ways, for example, [48, 49], but to retain the computational simplicity, we refrain from doing this here.

In the experiments, signals are generated according to the model in (1) with Gaussian noise. Furthermore, all amplitudes are set to unity, that is, $A_l = 1$ for all l and the signal-to-noise ratio (SNR) is defined as SNR $= 10 \log_{10}(\sum_{l=1}^{L} A_l^2/\sigma^2)$ [dB]. Note that similar results have been obtained for other amplitude distributions. For example, the general conclusions are the same for a Rayleigh pdf, but in the interest of brevity we will focus on the simple case of unit amplitudes. The sinusoidal phases and frequencies are generated according to a uniform pdf in the interval $(-\pi, \pi]$ which will result in spectrally overlapping sinusoids sometimes. For each combination of the parameters, 500 Monte Carlo simulations were run. Unless otherwise stated, we will use $L = 5$ and $M = N/2$.

4.2. Statistical Evaluation. First, we will evaluate the performance in terms of the percentage of correctly estimated orders under various conditions. We start out by varying the number of observations N while keeping the SNR fixed at 20 dB and then we will keep N fixed at 200 while varying the SNR. The partitioning of the EVD into signal and noise subspaces in (7) and (8) depends on the sorting
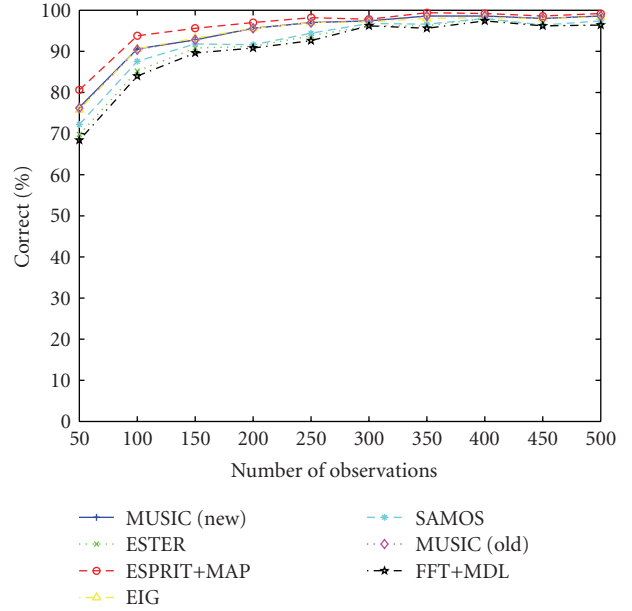

FIGURE 2: Percentage of correctly estimated model orders as a function of the number of observations for an SNR of 20 dB.
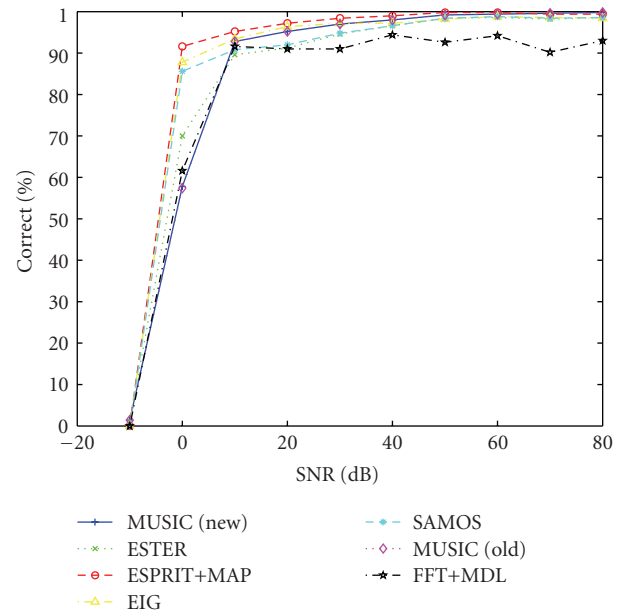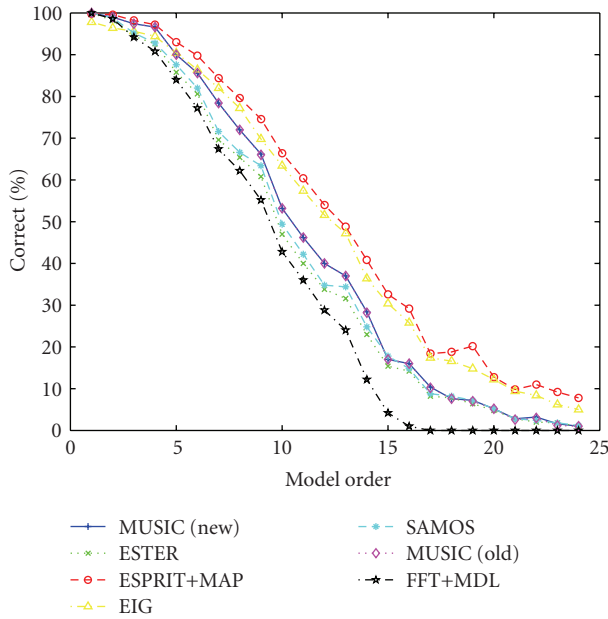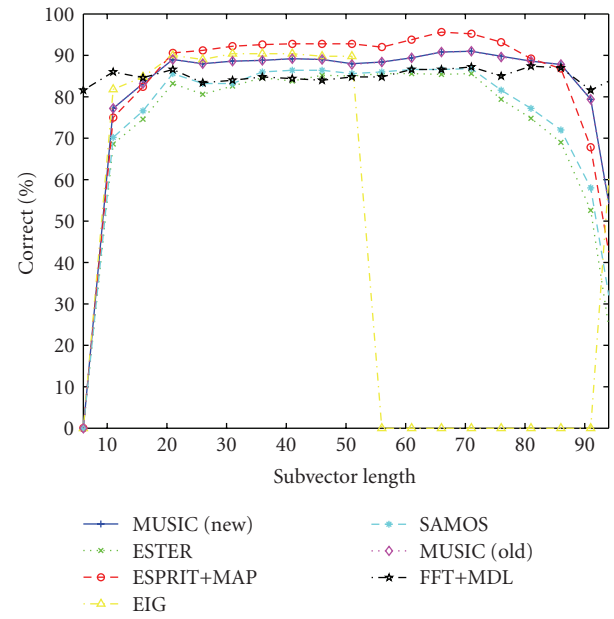

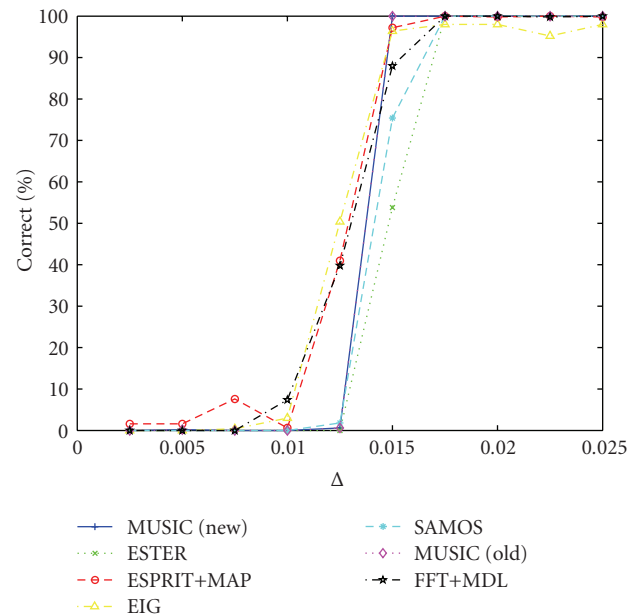FIGURE 3: Percentage of correctly estimated model orders versus the SNR for $N = 200$.

of the eigenvalues resulting in the right ordering of the eigenvectors. As a result, the performance of the methods is expected to depend on the SNR. The results are shown in Figures 2 and 3. Next, we evaluate the performance as a function of the true model order for $N = 100$ and SNR $= 20$ dB. Note that the choice of M also limits the number of possible sinusoids that can be found using MUSIC since $M > L$. The results are depicted in Figure 4. An experiment to investigate the dependency of the performance on the choice of M while keeping $N = 100$ constant has also been

TABLE 1: List of reference methods used in the experiments with short descriptions and references to literature.

Name	Reference	Description
ESTER	[20]	Subspace-based method based on the shift-invariance property of the signal model
ESPRIT+MAP	[4, 16]	Frequencies estimated using ESPRIT, amplitudes using least-squares, model selection using the MAP criterion
EIG	[19]	Method based on the ratio between the arithmetic and geometric means of the eigenvalues
SAMOS	[21]	Same as ESTER except for measure
MUSIC (old)	[22, 23]	Same as the proposed method except for the normalization
FFT+MDL	[1, 12, 13]	Statistical method based on MDL, with parameters estimated using the periodogram

FIGURE 4: Percentage of correctly estimated model orders as a function of the true order with SNR = 20 dB and $N = 100$.

FIGURE 5: Percentage of correctly estimated model orders as a function of subvector length with SNR = 20 dB and $N = 100$.

conducted with an SNR of 20 dB. The results are shown in Figure 5. The reason that the method of [19] fails here is that the covariance matrix is rank deficient for $M > N/2$. This can of course easily be fixed by modifying the range over which the geometric and arithmetic means of the eigenvalues are calculated. Since the gap between the signal and noise subspace eigenvalues depends not only on the SNR but also on how closely spaced the sinusoids are in frequency, the importance of the difference in frequency between the sinusoids will now be investigated. We do this by distributing the frequencies evenly as $2\pi\Delta l$ and then vary Δ for $L = 5$ sinusoids, $N = 100$, $M = 25$, and an SNR of 20 dB. All other experimental conditions are as described earlier. The results are shown in Figure 6. In a final experiment, we illustrate the applicability of the estimators in the presence of colored Gaussian noise. The percentages of correctly estimated orders are shown in Figure 7 as a function of the SNR. To generate the colored noise, a second-order autoregressive process was used having the transfer function $H(z) = 1/(1 - 0.25z^{-1} + 0.5z^{-2})$. Other than the noise color, the experimental conditions are the same as for Figure 3, that is, with $N = 200$. Note that for a fair comparison, the white noise model selection criterion has been used for all the

FIGURE 6: Percentage of correctly estimated model orders as a function of the difference between frequencies distributed uniformly as $2\pi\Delta l$ with SNR = 20 dB and $N = 100$.
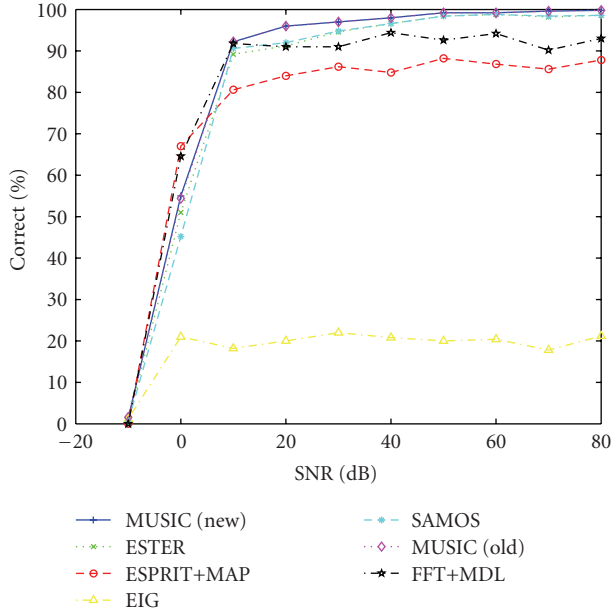
FIGURE 7: Percentage of correctly estimated model orders as a function of the SNR for colored Gaussian noise for $N = 200$.

methods. In other words, this experiment can be seen as an evaluation of the sensitivity to the white noise assumption. It is of course possible to modify the methods to take the colored noise into account in various ways, one way that can be applied to all the methods being prewhitening [18], but all such ways require that the statistics of the noise be known.

5. Discussion

From the experiments the following general observations can be made. First of all, it can be observed that, with one exception, all the methods exhibit the same dependencies on the tested variables, although they sometimes exhibit quite different thresholding behavior. The one exception is for colored Gaussian noise. It can be seen from these figures that the proposed estimator has the desirable properties that the performance improves as the SNR and/or the number of observations increases and that the model order can be determined with high probability for a high SNR and/or a high number of observations, and this is generally the case of all the tested methods. MUSIC can also be observed to consistently outperform the other subspace methods based on the eigenvectors, namely, ESTER and SAMOS. Curiously, the new MUSIC criterion performs similarly to the old one in all the simulations, which indicates that the orthogonality criterion does not depend strongly on the normalization. The MAP criterion of [4] combined with ESPRIT and the method based on the eigenvalues [19] can be seen to generally perform the best, outperforming the measure based on angles between subspaces when the noise is white Gaussian. This is, most likely, due to these methods making use of the assumption that the noise is not only white but also Gaussian; this assumption is not used in the proposed method. Despite their good performance for

white Gaussian noise, both aforementioned methods appear to be rather sensitive to the white noise assumption and their performance is rather poor for colored noise. The poor performance of the eigenvalue-based method of [19] for colored noise is no surprise. In fact, for colored noise, the method of [19] can be shown to overestimate the model order with probability 1 [50, 51]. That the MAP criterion in combination with ESPRIT outperforms the method of [13] can only be attributed to the former method resulting in superior parameter estimates to the periodogram, which will fail to resolve adjacent sinusoids for a low number of samples. We observe from Figure 4 that the performance of all the methods deteriorates as the number of parameters approaches M. That the MAP-based method fails in this case cannot be solely attributed to the MAP rule since it relies on sinusoidal parameter estimates being accurate. However, the MAP rule was derived in [4] based on the assumption that the likelihood function is highly peaked around the parameters estimates, which is usually the case when N is high relative to the number of parameters. We have observed from order estimation error histograms that while the orders are not estimated correctly for high orders, the estimated order is still generally close to the true one and may thus still be useful. From Figure 5, it appears that the methods are not very sensitive to the choice of M as long as it is not chosen too low or too high, that is, not too close to either L or N.

6. Conclusion and Future Work

In this paper, we have considered the problem of finding the number of complex sinusoids in white noise, and a new measure for solving this problem has been derived based on angles between the noise subspace and the candidate model. The measure is essentially the mean of the cosine to all non-trivial angle squared, which is asymptotically closely related to the original MUSIC cost function as defined for direction-of-arrival and frequency estimation. The derivations in this paper put order estimation using the orthogonality property of MUSIC on a firm mathematical ground. Numerical simulations show that the correct order can be determined for a high number of observations and/or a high signal-to-noise ratio (SNR) with a high probability. Additionally, experiments show that the performance of the proposed method exhibits the same functional dependencies on the SNR, the number of observations, and the model order as statistical methods. The experiments showed that the proposed method outperforms other previously published subspace methods and that the method is more robust to the noise being colored than all the other methods. Future work includes a rigorous statistical analysis of the proposed method along the lines of [33].

Appendix

Alternative Derivation of the Old Measure

We will now derive the normalized MUSIC cost function first proposed in [22] for finding the number of sinusoids.

Note that this derivation differs from the one in [22]. The following can be established for the acute angle $0 \leq \theta_{l,m} \leq \pi/2$ between two vectors $\mathbf{a}(\omega_l)$ and $\mathbf{q}_m$:

$$\cos^2\theta_{l,m} = \frac{\left|\mathbf{a}^H(\omega_l)\mathbf{q}_m\right|^2}{\|\mathbf{a}(\omega_l)\|_2^2\|\mathbf{q}_m\|_2^2}. \tag{A.1}$$

Averaging over $\cos^2\theta_{l,m}$ for all vector pairs, we get

$$\begin{aligned}
J &= \frac{1}{L(M-L)}\sum_{l=1}^{L}\sum_{m=L+1}^{M}\cos^2\theta_{l,m} \\
&= \frac{1}{L(M-L)}\sum_{l=1}^{L}\sum_{m=L+1}^{M}\frac{\left|\mathbf{a}^H(\omega_l)\mathbf{q}_m\right|^2}{\|\mathbf{a}(\omega_l)\|_2^2\|\mathbf{q}_m\|_2^2}.
\end{aligned} \tag{A.2}$$

Noting that all the columns of $\mathbf{A}$ and $\mathbf{G}$ have the same norms, this can be written as

$$\begin{aligned}
J &= \sum_{l=1}^{L}\sum_{m=L+1}^{M}\frac{\left|\mathbf{a}^H(\omega_l)\mathbf{q}_m\right|^2}{L\|\mathbf{a}(\omega_l)\|_2^2(M-L)\|\mathbf{q}_m\|_2^2} \\
&= \frac{\|\mathbf{A}^H\mathbf{G}\|_F^2}{\|\mathbf{A}\|_F^2\|\mathbf{G}\|_F^2} = \frac{\|\mathbf{A}^H\mathbf{G}\|_F^2}{LM(M-L)},
\end{aligned} \tag{A.3}$$

and, clearly, we have the following inequalities:

$$0 \leq \frac{\|\mathbf{A}^H\mathbf{G}\|_F^2}{LM(M-L)} \leq 1, \tag{A.4}$$

which also follow from the Cauchy-Schwartz inequality. The orthogonality measure in (A.3) has the desirable properties that it facilitates optimization over the individual columns of $\mathbf{A}$ and is invariant to the dimensions of the matrices. This measure is different than the original measure proposed in [14] due to the scaling of the cost function. Note that the MUSIC cost function originally was introduced as the reciprocal of the Euclidean distance between the signal model vectors and the signal subspace.

Acknowledgments

References

[1] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.

[2] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, pp. 461–464, 1978.

[3] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[4] P. M. Djuric, "Asymptotic MAP criteria for model selection," *IEEE Transactions on Signal Processing*, vol. 46, no. 10, pp. 2726–2735, 1998.

[5] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Pearson Prentice Hall, Englewood Cliffs, NJ, USA, 2005.

[6] E. G. Larsson and Y. Selen, "Linear regression with a sparse parameter vector," *IEEE Transactions on Signal Processing*, vol. 55, no. 2, pp. 451–460, 2007.

[7] J.-J. Fuchs, "Estimating the number of sinusoids in additive white noise," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 12, pp. 1846–1853, 1988.

[8] J.-J. Fuchs, "Estimation of the number of signals in the presence of unknown correlated sensor noise," *IEEE Transactions on Signal Processing*, vol. 40, no. 5, pp. 1053–1061, 1992.

[9] A. T. James, "Test of equality of latent roots of a covariance matrix," in *Multivariate Analysis*, pp. 205–218, 1969.

[10] B. G. Quinn, "Estimating the number of terms in a sinusoidal regression," *Journal of Time Series Analysis*, vol. 10, no. 1, pp. 71–75, 1989.

[11] X. Wang, "An AIC type estimator for the number of cosinusoids," *Journal of Time Series Analysis*, vol. 14, no. 4, pp. 434–440, 1993.

[12] L. Kavalieris and E. J. Hannan, "Determining the number of terms in a trigonometric regression," *Journal of Time Series Analysis*, vol. 15, no. 6, pp. 613–625, 1994.

[13] E. J. Hannan, "Determining the number of jumps in a spectrum," in *Developments in Time Series Analysis*, pp. 127–138, Chapman and Hall, London, UK, 1993.

[14] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[15] G. Bienvenu, "Influence of the spatial coherence of the background noise on high resolution passive methods," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '79)*, pp. 306–309, April 1979.

[16] R. Roy and T. Kailath, "ESPRIT—estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[17] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophysical Journal of the Royal Astronomical Society*, vol. 33, pp. 347–366, 1973.

[18] G. Bienvenu and L. Kopp, "Optimality of high resolution array processing using the eigensystem approach," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 31, no. 5, pp. 1235–1248, 1983.

[19] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.

[20] R. Badeau, B. David, and G. Richard, "A new perturbation analysis for signal enumeration in rotational invariance techniques," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 450–458, 2006.

[21] J.-M. Papy, L. De Lathauwer, and S. van Huffel, "A shift invariance-based order-selection technique for exponential data modelling," *IEEE Signal Processing Letters*, vol. 14, no. 7, pp. 473–476, 2007.

[22] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Joint high-resolution fundamental frequency and order estimation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 5, pp. 1635–1644, 2007.

[23] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Sinusoidal order estimation using the subspace orthogonality and shift-invariance properties," in *Proceedings of the 41st Asilomar Conference on Signals, Systems and Computers (ACSSC '07)*, pp. 651–655, Pacific Grove, Calif, USA, November 2007.

[24] M. Haardt and J. A. Nossek, "Unitary ESPRIT: how to obtain increased estimation accuracy with a reduced computational

burden," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1232–1242, 1995.

[25] P. C. Hansen and S. H. Jensen, "Prewhitening for rank-deficient noise in subspace methods for noise reduction," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3718–3726, 2005.

[26] P. C. Hansen and S. H. Jensen, "Subspace-based noise reduction for speech signals via diagonal and triangular matrix decompositions: survey and analysis," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 92953, 24 pages, 2007.

[27] M. G. Christensen and A. Jakobsson, *Multi-Pitch Estimation*, vol. 5 of *Synthesis Lectures on Speech & Audio Processing*, Morgan & Claypool, San Rafael, Calif, USA, 2009.

[28] A. J. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '83)*, vol. 1, pp. 336–339, Boston, Mass, USA, April 1983.

[29] H. Saarnisaari, "Robustness of the MUSIC algorithm to errors in estimation the dimensions of the subspaces: delay estimation in DS/SS in the presence of interference," in *Proceedings of IEEE Military Communications Conference (MILCOM '99)*, vol. 2, pp. 851–854, 1999.

[30] A. Eriksson, P. Stoica, and T. Töderström, "Asymptotical analysis of MUSIC and ESPRIT frequency estimates," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '93)*, vol. 4, pp. 556–559, Minneapolis, Minn, USA, April 1993.

[31] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 5, pp. 720–741, 1989.

[32] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound: further results and comparisons," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 12, pp. 2140–2150, 1990.

[33] B. G. Quinn and E. J. Hannan, *The Estimation and Tracking of Frequency*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, UK, 2001.

[34] D. R. Farrier, D. J. Jeffries, and R. Mardani, "Theoretical performance prediction of the music algorithm," *IEE Proceedings F: Communications, Radar and Signal Processing*, vol. 135, no. 3, pp. 216–224, 1988.

[35] G. H. Golub and C. F. V. Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, Md, USA, 3rd edition, 1996.

[36] A. V. Knyazev and M. E. Argentati, "Principal angles bewteen subspaces in an a-based sclar product: algorithms and pertubation estimates," *SIAM Journal on Scientific Computing*, vol. 23, no. 6, pp. 2009–2041, 2002.

[37] P.-A. Absil, A. Edelman, and P. Koev, "On the largest principal angle between random subspaces," *Linear Algebra and Its Applications*, vol. 414, no. 1, pp. 288–294, 2006.

[38] J. Miao and A. Ben-Israel, "On principal angles between subspaces in $\mathbb{R}^n$," *Linear Algebra and Its Applications*, vol. 171, pp. 81–98, 1992.

[39] J. Miao and A. Ben-Israel, "Product cosines of angles between subspaces," *Linear Algebra and Its Applications*, vol. 237-238, pp. 71–81, 1996.

[40] S. Jiang, "Angles between euclidean subspaces," *Geometriae Dedicata*, vol. 63, no. 2, pp. 113–121, 1996.

[41] M. Hawkes, A. Nehorai, and P. Stoica, "Performance breakdown of subspace-based methods: prediction and cure," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '01)*, vol. 6, pp. 4005–4008, Salt Lake, Utah, USA, May 2001.

[42] J. K. Thomas, L. L. Scharf, and D. W. Tufts, "The probability of a subspace swap in the SVD," *IEEE Transactions on Signal Processing*, vol. 43, no. 3, pp. 730–736, 1995.

[43] B. Yang, "Projection approximation subspace tracking," *IEEE Transactions on Signal Processing*, vol. 43, no. 1, pp. 95–107, 1995.

[44] D. J. Rabideau, "Fast, rank adaptive subspace tracking and applications," *IEEE Transactions on Signal Processing*, vol. 44, no. 9, pp. 2229–2244, 1996.

[45] P. Strobach, "Low-rank adaptive filters," *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 2932–2947, 1996.

[46] J. R. Jensen, J. K. Nielsen, M. G. Christensen, S. H. Jensen, and T. Larsen, "On fast implementation of harmonic MUSIC for known and unknown model orders," in *Proceedings of the 16th European Signal Processing Conferenc (EURASIP '08)*, Lausanne, Switzerland, August 2008.

[47] R. Kumaresan and D. W. Tufts, "Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 30, no. 6, pp. 833–840, 1982.

[48] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 4, pp. 477–489, 1988.

[49] B. G. Quinn and J. M. Fernandes, "A fast efficient technique for the estimation of frequency," *Biometrika*, vol. 78, no. 3, pp. 489–497, 1991.

[50] W. Xu and M. Kaveh, "Analysis of the performance and sensitivity of eigendecomposition-based detectors," *IEEE Transactions on Signal Processing*, vol. 43, no. 6, pp. 1413–1426, 1995.

[51] A. P. Liavas and P. A. Regalia, "On the behavior of information theoretic criteria for model order selection," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1689–1695, 2001.