A new perturbation analysis for signal enumeration in rotational invariance techniques

Roland Badeau, Bertrand David, Gael Richard

A new perturbation analysis for signal enumeration in rotational invariance techniques

Roland Badeau, *Member, IEEE,* Bertrand David, and Gaël Richard, *Member, IEEE*

Abstract— The ESPRIT algorithm is a subspace-based high resolution method used in source localization and spectral analysis, which provides very accurate estimates of the signal parameters. However, the underlying theory assumes a known model order, which is usually not the case in many applications. In particular, it is well known that under-evaluating the model order biases the estimation. In this paper, we analyze the perturbation induced by an erroneous model order, and we present an error bound for the estimated parameters. Based on this theoretical framework, we propose a new method for selecting an appropriate modeling order, which consists in minimizing the error bound. This approach is applied to both synthetic and musical signals and its performance is compared to that of existing methods, such as the Information Theoretic Criteria (ITC).

Index Terms— ESPRIT, rotational invariance, model order selection, signal enumeration, perturbation theory.

I. Introduction

ESTIMATING a line spectrum is an important task for many applications, such as speech signal analysis and synthesis [1] and musical signal modification [2]. Although the Fourier transform remains a prominent tool for frequency estimation, the ESPRIT algorithm [3] overcomes the resolution limit of the Fourier analysis and provides straight estimates of the model parameters. This method relies on the rotational invariance property of the signal subspace spanned by the sinusoids. Its drawback is that the model order is supposed to be known, which is not the case in practice.

Many methods were proposed in the literature for estimating the number of sinusoids in white noise. The most classical ones are the maximum likelihood method [4] and the Information Theoretic Criteria (ITC) [5], among which the Akaike Information Criterion (AIC) [6] and the Maximum Description Length (MDL) by Schwartz [7] and Rissanen [8]. Another consistent procedure in the framework of the ITC is the Efficient Detection Criteria (EDC) [9], which proves to be robust to non-additive white noise [10]. The various ITC rely on the similarity of the eigenvalues within the noise subspace, and not on the existence of a gap between the signal and noise subspaces [11]. A criterion for model order selection based on this gap, which looks for a *maximally stable* decomposition, has been developed in [12]. Other methods proposed for model

order selection include the Wishart matrices [13] and the cross-validation [14] approaches. However in the presence of a correlated noise, these methods tend to overestimate the model order. Consequently, specific methods have been designed to address the colored noise case, including new information theoretic criteria [15], [16], a technique based on a band noise covariance matrix model [17], and a maximum a posteriori criterion [18].

In other respects, we show in this paper how applying the ESPRIT high resolution method with an erroneous model order perturbs the estimation of the sinusoids. Note that in the literature, most papers rather focus on the perturbations induced by the additive noise. For example, the asymptotic second-order properties of ESPRIT were studied in the Direction of Arrival (DOA) [19] and in the frequency estimation [20] context, for a finite Signal to Noise Ratio (SNR). Reciprocally, a similar study was carried out for a finite number of data samples under a large SNR hypothesis [21]. In [22], a class of modeling errors were analyzed. However, to the best of our knowledge, no perturbation analysis of the ESPRIT estimates in the case of erroneous modeling order has ever been published (in the case of the MUSIC algorithm, a study is available in [23]).

Note that all the above mentioned performance analyses of the ESPRIT algorithm, as well as the subspace perturbation approach in [24], rely on first order approximations. Conversely, we present in this paper error bounds for the frequency estimates, which are derived without approximation and which can be easily computed. Furthermore, they are more precise than those presented in [25]. Based on this result, we propose a new model order selection method which consists in minimizing the perturbation. Contrary to the other methods proposed in the literature, which select the model order by analyzing the spectral properties of the additive noise, our approach focuses on the signal itself. Although it relies on a noiseless model, we observed that it outperforms the classical ITC, even in low SNR scenarios.

The paper is organized as follows. Section II summarizes the principles of the ESPRIT high resolution method. In section III, the perturbation of the poles induced by an erroneous model order is analyzed. Then our new model order selection method, referred to as the ESTER method, is introduced in section IV, where a fast implementation is proposed. In section V, the relevance of our criterion as an error bound is examined, and the performance of the ESTER method is compared to that of some existing methods. Finally, the main conclusions of this paper are summarized in section VI.

Roland Badeau, Bertrand David, and Gaël Richard are with the Department of Signal and Image Processing, Ecole Nationale Supérieure des Télécommunications (ENST), Paris, France. E-mail: [roland.badeau, bertrand.david, gael.richard]@enst.fr.

II. THE ESPRIT METHOD FOR SPECTRAL ANALYSIS

The noiseless Exponential Sinusoidal Model (ESM), also known as the Exponentially Damped Sinusoidal (EDS) model [26], defines the discrete signal as

$$x(t) = \sum_{k=1}^{r} \alpha_k z_k^t$$

where r is the order of the model, $\alpha_k \in \mathbb{C} - \{0\}$ are the complex amplitudes, and $z_k \in \mathbb{C} - \{0\}$ are the complex poles.

Let $n > r$, $l > r$, and define the $n \times l$ Hankel data matrix

$$\boldsymbol{X} = \begin{bmatrix} x(0) & x(1) & \cdots & x(l-1) \\ x(1) & x(2) & \cdots & x(l) \\ \vdots & \vdots & \cdots & \vdots \\ x(n-1) & x(n) & \cdots & x(n+l-2) \end{bmatrix}.$$

which involves $N = n + l - 1$ samples of the signal. This matrix can be factorized in the form $\boldsymbol{X} = \boldsymbol{V}^n \boldsymbol{A} \boldsymbol{V}^{l^T}$, where $\boldsymbol{V}^n$ is a $n \times r$ Vandermonde matrix

$$\boldsymbol{V}^n = \begin{bmatrix} 1 & \cdots & 1 \\ z_1 & \cdots & z_r \\ \vdots & \vdots & \vdots \\ z_1^{n-1} & \cdots & z_r^{n-1} \end{bmatrix},$$

$\boldsymbol{A} = \text{diag}(\alpha_1 \ldots \alpha_r)$, and $\boldsymbol{V}^l$ is a $l \times r$ Vandermonde matrix [27]. If the r poles $\{z_1, \ldots, z_r\}$ are distinct, $\boldsymbol{X}$ has a r-dimensional range space, spanned by the full-rank matrix $\boldsymbol{V}^n$. This range space fully characterizes the signal poles. It is thus referred to as the *signal subspace*.

Let $\boldsymbol{V}_{\downarrow}^n$ be the matrix extracted from $\boldsymbol{V}^n$ by deleting the last row. Similarly, let $\boldsymbol{V}_{\uparrow}^n$ be the matrix extracted from $\boldsymbol{V}^n$ by deleting the first row. Then the Vandermonde matrix $\boldsymbol{V}^n$ satisfies the so-called *rotational invariance property*:

$$\boldsymbol{V}_{\uparrow}^n = \boldsymbol{V}_{\downarrow}^n \boldsymbol{D} \tag{1}$$

where $\boldsymbol{D} = \text{diag}(z_1, \ldots, z_r)$.

In practice, $\boldsymbol{V}^n$ is unknown, but the signal subspace can be obtained by computing the singular value decomposition (SVD) of $\boldsymbol{X}$ (or via subspace tracking techniques [28]–[30] in an adaptive context). Indeed, if $\{\boldsymbol{w}(1), \ldots, \boldsymbol{w}(n)\}$ are the left singular vectors associated to the singular values $\sigma_1 \geq \ldots \geq \sigma_n \geq 0$ sorted in decreasing order, then the signal subspace is spanned by the $n \times r$ orthonormal matrix $\boldsymbol{W}(r) = [\boldsymbol{w}(1), \ldots, \boldsymbol{w}(r)]$ (the $n - r$ last singular values being equal to 0). Since $\boldsymbol{W}(r)$ and $\boldsymbol{V}^n$ span the same subspace, there is a non-singular matrix $\boldsymbol{G}$ of dimension $r \times r$ such that

$$\boldsymbol{V}^n = \boldsymbol{W}(r) \boldsymbol{G}. \tag{2}$$

By deleting the last row in equation (2) we obtain

$$\boldsymbol{W}_{\downarrow}(r) = \boldsymbol{V}_{\downarrow}^n \boldsymbol{G}^{-1}. \tag{3}$$

Similarly, deleting the first row in equation (2) and substituting equation (1) yields

$$\boldsymbol{W}_{\uparrow}(r) = \boldsymbol{V}_{\downarrow}^n \boldsymbol{D} \boldsymbol{G}^{-1}. \tag{4}$$

Substituting equation (3) into equation (4) leads to the rotational invariance property of the matrix $\boldsymbol{W}(r)$:

$$\boldsymbol{W}_{\uparrow}(r) = \boldsymbol{W}_{\downarrow}(r) \boldsymbol{\Phi}(r)$$

where $\boldsymbol{\Phi}(r)$ is defined by its eigenvalue decomposition: $\boldsymbol{\Phi}(r) \triangleq \boldsymbol{G} \boldsymbol{D} \boldsymbol{G}^{-1}$. Finally, the ESPRIT algorithm [3] consists of the following steps:

- computing $\boldsymbol{W}(r)$,
- computing $\boldsymbol{\Phi}(r) = \boldsymbol{W}_{\downarrow}(r)^{\dagger} \boldsymbol{W}_{\uparrow}(r)$ (where the symbol $\dagger$ denotes the Moore-Penrose pseudo-inverse),
- extracting the poles z_k as the eigenvalues of $\boldsymbol{\Phi}(r)$.

III. IMPACT OF AN ERRONEOUS MODEL ORDER

In practice, the model order r is unknown. We assume below that the ESPRIT algorithm is applied with an erroneous model order p and we focus on how the estimation of the poles is affected. For all $p \in \{1 \ldots n\}$, let $\boldsymbol{W}(p) \triangleq [\boldsymbol{w}(1), \ldots, \boldsymbol{w}(p)]$ and

$$\boldsymbol{\Phi}(p) \triangleq \boldsymbol{W}_{\downarrow}(p)^{\dagger} \boldsymbol{W}_{\uparrow}(p). \tag{5}$$

The estimated poles are defined as the eigenvalues of $\boldsymbol{\Phi}(p)$.

A. Over-estimation of the model order

If $p \geq r$, the following proposition shows that the r true poles belong to the whole set of eigenvalues of $\boldsymbol{\Phi}(r)$.

Proposition III.1. *Suppose that $r \leq p < n$ and $\boldsymbol{W}_{\downarrow}(p)$ is full rank. Then $\forall k \in \{1, \ldots, r\}$, z_k is an eigenvalue of $\boldsymbol{\Phi}(p)$.*

Proof. Let $\boldsymbol{v}_k$ be the right eigenvector of $\boldsymbol{\Phi}(r)$ associated to the eigenvalue z_k and consider the p-dimensional vector $\overline{\boldsymbol{v}}_k \triangleq \begin{bmatrix} \boldsymbol{v}_k \\ \boldsymbol{0} \end{bmatrix}$. Note that $\boldsymbol{W}(p) \overline{\boldsymbol{v}}_k = \boldsymbol{W}(r) \boldsymbol{v}_k$. Consequently, $\boldsymbol{W}_{\uparrow}(p) \overline{\boldsymbol{v}}_k = \boldsymbol{W}_{\uparrow}(r) \boldsymbol{v}_k = \boldsymbol{W}_{\downarrow}(r) \boldsymbol{\Phi}(r) \boldsymbol{v}_k = z_k \boldsymbol{W}_{\downarrow}(r) \boldsymbol{v}_k = z_k \boldsymbol{W}_{\downarrow}(p) \overline{\boldsymbol{v}}_k$. Since $\boldsymbol{W}_{\downarrow}(p)$ is full rank, left multiplying the previous equality by $\boldsymbol{W}_{\downarrow}(p)^{\dagger}$ yields $\boldsymbol{\Phi}(p) \overline{\boldsymbol{v}}_k = z_k \overline{\boldsymbol{v}}_k$. $\square$

B. Under-estimation of the model order

If $p < r$, it is well known that the eigenvalues of $\boldsymbol{\Phi}(p)$ do not match the poles in the general case. More precisely, let $\widehat{z}$ be an eigenvalue of $\boldsymbol{\Phi}(p)$. In this section, it will be shown that $\widehat{z}$ approximates one of the eigenvalues of $\boldsymbol{\Phi}(r)$, and that an error bound can be easily computed. First, we need to define the *upper condition number* of the signal subspace[1]:

$$\kappa_2 = \inf_{\boldsymbol{\Lambda} \in \text{diag}(\mathbb{R}^{+r})} \frac{\sigma_{\max}(\boldsymbol{V}^n \boldsymbol{\Lambda})}{\sigma_{\min}(\boldsymbol{V}_{\downarrow}^n \boldsymbol{\Lambda})} \tag{6}$$

where $\sigma_{\max}(.)$ denotes the largest singular value of a matrix, and $\sigma_{\min}(.)$ denotes the smallest one. This condition number characterizes the noiseless signal itself, and does not depend on p. It is an unknown constant for our problem, which does not need to be calculated. It is involved in the following theorem, whose proof can be found in the appendix.

---

[1]In [25], κ_2 was defined as the upper condition number of the *Vandermonde matrix* $\boldsymbol{V}^n$, equal to $\frac{\sigma_{\max}(\boldsymbol{V}^n)}{\sigma_{\min}(\boldsymbol{V}_{\downarrow}^n)}$. The new definition of κ_2 in equation (6) yields better error bounds, due to the presence of the infimum.

Notation: $\mathbb{R}^+$ denotes the set of all positive real numbers.

Theorem III.2 (a priori error bound[2]). *For all $\widehat{z} \in \mathbb{C}$, there is an eigenvalue z_k of $\boldsymbol{\Phi}(r)$ for which*

$$|\widehat{z} - z_k| \leq \kappa_2 \, \sigma_{\min}\left(\boldsymbol{W}_{\uparrow}(p) - \widehat{z}\,\boldsymbol{W}_{\downarrow}(p)\right). \tag{7}$$

Note that the bound $\sigma_{\min}\left(\boldsymbol{W}_{\uparrow}(p) - \widehat{z}\,\boldsymbol{W}_{\downarrow}(p)\right)$ can be computed without knowing the exact value of r. Corollary III.3, which follows from theorem III.2, has a certain similarity with the well known Bauer-Fike theorem [31, pp. 365], [32, pp. 321]. It gives an error bound valid for all the eigenvalues of $\boldsymbol{\Phi}(p)$. Again, this bound can be computed without knowing the exact value of r. It involves the spectral norm of a matrix (or 2-norm), also denoted $\|.\|_2$, defined as $\|\boldsymbol{M}\|_2 \triangleq \max_{\|\boldsymbol{u}\|_2=1} \|\boldsymbol{M}\,\boldsymbol{u}\|_2 = \sigma_{\max}(\boldsymbol{M})$.

Corollary III.3 (a posteriori error bound[3]). *For each eigenvalue $\widehat{z}$ of $\boldsymbol{\Phi}(p)$, there is an eigenvalue z_k of $\boldsymbol{\Phi}(r)$ for which*

$$|\widehat{z} - z_k| \leq \kappa_2 \, \|\boldsymbol{E}(p)\|_2 \tag{8}$$

where

$$\boldsymbol{E}(p) = \boldsymbol{W}_{\uparrow}(p) - \boldsymbol{W}_{\downarrow}(p)\,\boldsymbol{\Phi}(p). \tag{9}$$

Proof. Let $\widehat{z}$ be an eigenvalue of $\boldsymbol{\Phi}(p)$ and $\widehat{\boldsymbol{v}}$ a unitary eigenvector associated with $\widehat{z}$. Let $\boldsymbol{e}(p) \triangleq \left(\boldsymbol{W}_{\uparrow}(p) - \widehat{z}\,\boldsymbol{W}_{\downarrow}(p)\right)\widehat{\boldsymbol{v}}$. Since $\widehat{\boldsymbol{v}}$ is unitary, $\sigma_{\min}\left(\boldsymbol{W}_{\uparrow}(p) - \widehat{z}\,\boldsymbol{W}_{\downarrow}(p)\right) \leq \|\boldsymbol{e}(p)\|_2$. In other respects, $\boldsymbol{e}(p) = \left(\boldsymbol{W}_{\uparrow}(p) - \boldsymbol{W}_{\downarrow}(p)\,\boldsymbol{\Phi}(p)\right)\widehat{\boldsymbol{v}}$, thus $\|\boldsymbol{e}(p)\|_2 \leq \|\boldsymbol{E}(p)\|_2$. Consequently,

$$\sigma_{\min}\left(\boldsymbol{W}_{\uparrow}(p) - \widehat{z}\,\boldsymbol{W}_{\downarrow}(p)\right) \leq \|\boldsymbol{E}(p)\|_2. \tag{10}$$

Finally, substituting equation (10) into (7) yields equation (8). $\qquad\square$

Remark. *Let $p < n - 1$. We know that if $p = r$, $\boldsymbol{E}(p) = 0$. Conversely, if $\boldsymbol{E}(p) = 0$, then the matrices $\boldsymbol{W}_{\downarrow}(p)$ and $\boldsymbol{W}_{\uparrow}(p)$ span the same subspace, which means that the rotational invariance property is satisfied at order p. Thus p complex exponentials can be extracted from the observed signal, and the corresponding complex poles can be estimated by means of the ESPRIT algorithm. Since the signal does not contain more than r complex exponentials, we expect that $p \leq r$. The case $p < r$ can happen if the signal parameters satisfy some particular relationships. In practice, r is always the greatest value of p for which $\boldsymbol{E}(p) = 0$.*

IV. SELECTION OF AN APPROPRIATE MODELING ORDER BASED ON THE ESTIMATION ERROR

The practical interest of corollary III.3 is that $\|\boldsymbol{E}(p)\|_2$ (which will be referred to as the *a posteriori error bound*) can be computed for all $p \in \{1 \dots p_{\max}\}$, where $1 \leq p_{\max} < n - 1$. If $p_{\max}$ happens to be lower than r, the a posteriori error bound gives a quantitative criterion for selecting an

---

[2]In comparison, the a priori error bound proposed in [25] was equal to $\frac{\sigma_{\max}(\boldsymbol{V}^n)}{\sigma_{\min}(\boldsymbol{V}_{\downarrow}^n)} \left\|\boldsymbol{W}_{\uparrow}(p)\,\widehat{\boldsymbol{v}} - \widehat{z}\,\boldsymbol{W}_{\downarrow}(p)\,\widehat{\boldsymbol{v}}\right\|_2$, where $\widehat{\boldsymbol{v}}$ was an arbitrary unitary vector. Note that the condition number κ_2 defined in equation (6) is lower than $\frac{\sigma_{\max}(\boldsymbol{V}^n)}{\sigma_{\min}(\boldsymbol{V}_{\downarrow}^n)}$. Moreover $\sigma_{\min}\left(\boldsymbol{W}_{\uparrow}(p) - \widehat{z}\,\boldsymbol{W}_{\downarrow}(p)\right) \leq \left\|\boldsymbol{W}_{\uparrow}(p)\,\widehat{\boldsymbol{v}} - \widehat{z}\,\boldsymbol{W}_{\downarrow}(p)\,\widehat{\boldsymbol{v}}\right\|_2$ for all unitary vector $\widehat{\boldsymbol{v}}$. Thus the a posteriori error bound in equation (7) is lower than that proposed in [25].

[3]This a posteriori error bound is lower than that proposed in [25], because of the lower value of κ_2.

appropriate modeling order, such that the estimation error bound is minimum. If $p_{\max}$ happens to be greater than r, then r is the greatest value of $p \in \{1 \dots p_{\max}\}$ for which the a posteriori error bound is zero. In any case, detecting the maxima of the inverse error function $J : p \mapsto \frac{1}{\|\boldsymbol{E}(p)\|_2^2}$ in the range $\{1 \dots p_{\max}\}$ is a relevant approach for selecting the modeling order. Below, this function will be referred to as the ESTimation ERror (ESTER) criterion. In presence of noise, we observed that a robust way of selecting the modeling order consists in detecting the greatest value of p for which the function $J(p)$ reaches a local maximum which is greater than a fraction of its global maximum (typically one tenth of the global maximum). Examples of the function J are represented in section V. Proposition IV.1 shows that its values are in the interval $[1, +\infty[$ (the proof is given in the appendix).

Proposition IV.1. *For all $p \in \{1, \dots, n\}$, $\|\boldsymbol{E}(p)\|_2 \leq 1$.*

Note that this ESTER criterion measures the rotational invariance of $\boldsymbol{W}(p)$, since by definition the rotational invariance property is satisfied exactly if $\|\boldsymbol{E}(p)\|_2 = 0$. The drawback of the ESTER method is that a direct implementation would lead to a very computationally demanding algorithm. First, the singular vectors $\boldsymbol{w}(p)$ have to be computed for all $p \in \{1 \dots p_{\max}\}$, which requires $O(N \log_2(N) p_{\max} + n p_{\max}^2)$ operations, by means of a variant of the orthogonal iteration EVD algorithm presented in [26]. Then the matrix $\boldsymbol{E}(p)$ must be calculated for all $p \in \{1, \dots, p_{\max}\}$. Such a computation would involve $3np^2$ MAC for each p, so that the overall complexity would be $n p_{\max}^3$ MAC[4]. This computational cost is to be compared to that of the ITC criteria illustrated in section V. In particular, the complexity of the AIC, MDL [5] and EDC [9] criteria is linear in $p_{\max}$. However, contrary to the ESTER method, these ITC require the full SVD of the data matrix, whose complexity is $O(N^3)$. Besides, the complexity of the criteria proposed in [16] for addressing the colored noise case also have a complexity equal to $O(N^3)$. Consequently, the relative complexities of the ESTER method and the various ITC depend on $p_{\max}$.

To make the ESTER method faster, we developed a recursive implementation, presented in table I, which only involves $6np + O(p^2)$ MAC for each p, so that its overall complexity is $3 n p_{\max}^2 + O(p_{\max}^3)$ (plus the computation of the singular vectors). In particular, it can be noticed that computing the matrices $\boldsymbol{E}(p)$ for all $p \in \{1 \dots p_{\max}\}$ in this way is not more computationally demanding than computing $\boldsymbol{E}(p_{\max})$ directly. Sections IV-A and IV-B present fast methods for computing $\boldsymbol{\Phi}(p)$ and $\boldsymbol{E}(p)$ recursively.

A. Recursive computation of $\boldsymbol{\Phi}(p)$

A direct calculation of $\boldsymbol{\Phi}(p)$ for all $p \in \{1 \dots p_{\max}\}$ from equation (5) would involve $2np^2 + O(p^3)$ MAC for each p, and the overall complexity would be $\frac{2}{3} n p_{\max}^3 + O(p_{\max}^4)$. This section aims at computing the $p \times p$ matrix $\boldsymbol{\Phi}(p)$ recursively, in

---

[4]In this paper, operations counts are expressed in terms of multiply / accumulate (MAC) operations. Note that $p_{\max}$ is supposed to be much lower than n.

order to reduce the complexity. Equation (5) can be rewritten in the form

$$\boldsymbol{\Phi}(p) = \boldsymbol{\Omega}(p)\,\boldsymbol{\Psi}(p) \qquad (11)$$

where $\boldsymbol{\Omega}(p)$ and $\boldsymbol{\Psi}(p)$ are the $p \times p$ matrices

$$\boldsymbol{\Omega}(p) \triangleq \left(\boldsymbol{W}_{\downarrow}(p)^H\,\boldsymbol{W}_{\downarrow}(p)\right)^{-1} \qquad (12)$$

$$\boldsymbol{\Psi}(p) \triangleq \boldsymbol{W}_{\downarrow}(p)^H\,\boldsymbol{W}_{\uparrow}(p). \qquad (13)$$

The matrix $\boldsymbol{\Omega}(p)$ can be easily calculated. Indeed, since $\boldsymbol{W}(p)$ is orthonormal, $\boldsymbol{W}(p)^H\,\boldsymbol{W}(p) = \boldsymbol{I}_p$. In particular, this equation yields $\boldsymbol{W}_{\downarrow}(p)^H\,\boldsymbol{W}_{\downarrow}(p) = \boldsymbol{I}_p - \boldsymbol{\nu}(p)\,\boldsymbol{\nu}(p)^H$, where $\boldsymbol{\nu}(p)$ is the p-dimensional vector such that $\boldsymbol{\nu}(p)^H$ is the last row of $\boldsymbol{W}(p)$. Finally, the matrix inversion lemma [31, pp. 18-19] shows that

$$\boldsymbol{\Omega}(p) = \boldsymbol{I}_p + \frac{1}{1 - \|\boldsymbol{\nu}(p)\|^2}\,\boldsymbol{\nu}(p)\,\boldsymbol{\nu}(p)^H. \qquad (14)$$

Moreover, $\boldsymbol{\Psi}(p)$ can be recursively updated. Indeed, equation (13) yields

$$\boldsymbol{\Psi}(p) = \left[\begin{array}{c|c} \boldsymbol{\Psi}(p-1) & \boldsymbol{\psi}_r(p) \\ \hline \boldsymbol{\psi}_l(p)^H & \psi_{lr}(p) \end{array}\right] \qquad (15)$$

where $\boldsymbol{\psi}_r(p) \triangleq \boldsymbol{W}_{\downarrow}(p-1)^H\,\boldsymbol{w}_{\uparrow}(p)$, $\boldsymbol{\psi}_l(p) \triangleq \boldsymbol{W}_{\uparrow}(p-1)^H\,\boldsymbol{w}_{\downarrow}(p)$ and $\psi_{lr}(p) \triangleq \boldsymbol{w}_{\downarrow}(p)^H\,\boldsymbol{w}_{\uparrow}(p)$. It can be noticed that the computation of $\boldsymbol{\Psi}(p)$ from $\boldsymbol{\Psi}(p-1)$ requires only $2np$ MAC.

Finally, $\boldsymbol{\Phi}(p)$ can be computed from $\boldsymbol{\Psi}(p)$. Indeed, substituting equation (14) into equation (11) yields

$$\boldsymbol{\Phi}(p) = \boldsymbol{\Psi}(p) + \frac{1}{1 - \|\boldsymbol{\nu}(p)\|^2}\,\boldsymbol{\nu}(p)\,\boldsymbol{\varphi}(p)^H \qquad (16)$$

where

$$\boldsymbol{\varphi}(p) \triangleq \boldsymbol{\Psi}(p)^H\,\boldsymbol{\nu}(p). \qquad (17)$$

It can be noticed that the computation of $\boldsymbol{\Phi}(p)$ from $\boldsymbol{\Psi}(p)$ requires only $p^2 + O(p)$ MAC, plus the computation of $\boldsymbol{\varphi}(p)$. This last operation normally requires p^2 MAC, but lemma IV.2 suggests a recursive implementation, which involves only $O(p)$ MAC.

Lemma IV.2. *Let $\mu(p)$ be the complex number such that*

$$\boldsymbol{\nu}(p) = \left[\begin{array}{c} \boldsymbol{\nu}(p-1) \\ \mu(p) \end{array}\right] \qquad (18)$$

Then $\boldsymbol{\varphi}(p)$ satisfies the recursion

$$\boldsymbol{\varphi}(p) = \left[\begin{array}{c} \boldsymbol{\varphi}(p-1) + \mu(p)\,\boldsymbol{\psi}_l(p) \\ \boldsymbol{\psi}_r(p)^H\boldsymbol{\nu}(p-1) + \mu(p)\,\psi_{lr}(p)^* \end{array}\right] \qquad (19)$$

Proof. The assertion can be shown by substituting equations (15) an (18) into equation (17). $\qquad\square$

Finally, the recursive computation of $\boldsymbol{\Phi}(p)$ consists in computing $\boldsymbol{\Psi}(p)$ from $\boldsymbol{\Psi}(p-1)$ with equation (15), then computing $\boldsymbol{\varphi}(p)$ from $\boldsymbol{\varphi}(p-1)$ with equation (19), then computing $\boldsymbol{\Phi}(p)$ from $\boldsymbol{\Psi}(p)$ with equation (16). This method requires $2np + O(p^2)$ MAC at each step. Therefore, its overall computational cost is $np_{\max}^2 + O(p_{\max}^3)$ MAC.

B. Recursive computation of $\boldsymbol{E}(p)$

Here we suppose that all the $\boldsymbol{\Phi}(p)$ have been computed. A direct calculation of $\boldsymbol{E}(p)$ for all $p \in \{1 \ldots p_{\max}\}$ from equation (9) would involve np^2 MAC for each p, and the overall complexity would be $\frac{1}{3}\,np_{\max}^3$. This section aims at computing $\boldsymbol{E}(p)$ recursively, in order to reduce the complexity.

Substituting equation (16) into equation (9) shows that

$$\boldsymbol{E}(p) = \boldsymbol{\Xi}(p) - \frac{1}{1 - \|\boldsymbol{\nu}(p)\|^2}\,(\boldsymbol{W}_{\downarrow}(p)\,\boldsymbol{\nu}(p))\,\boldsymbol{\varphi}(p)^H \qquad (20)$$

where

$$\boldsymbol{\Xi}(p) \triangleq \boldsymbol{W}_{\uparrow}(p) - \boldsymbol{W}_{\downarrow}(p)\,\boldsymbol{\Psi}(p). \qquad (21)$$

Note that the computation of $\boldsymbol{E}(p)$ from $\boldsymbol{\Xi}(p)$ requires $2np$ MAC. Then substituting equation (15) into equation (21) yields a recursion for the $(n-1) \times p$ matrix $\boldsymbol{\Xi}(p)$:

$$\boldsymbol{\Xi}(p) = \left[\;\boldsymbol{\Xi}(p-1) - \boldsymbol{w}_{\downarrow}(p)\,\boldsymbol{\psi}_l(p)^H\;\middle|\;\boldsymbol{\xi}(p)\;\right] \qquad (22)$$

where $\boldsymbol{\xi}(p) \triangleq \boldsymbol{w}_{\uparrow}(p) - \boldsymbol{W}_{\downarrow}(p-1)\,\boldsymbol{\psi}_r(p) - \boldsymbol{w}_{\downarrow}(p)\,\psi_{lr}(p)$. The computation of $\boldsymbol{\Xi}(p)$ from $\boldsymbol{\Xi}(p-1)$ involves $2np$ MAC. Finally, the recursive computation of $\boldsymbol{E}(p)$ consists in computing $\boldsymbol{\Xi}(p)$ from $\boldsymbol{\Xi}(p-1)$ with equation (22), then computing $\boldsymbol{E}(p)$ from $\boldsymbol{\Xi}(p)$ with equation (20). This method requires $4np$ MAC at each step. Thus its overall computational cost is $2np_{\max}^2$ MAC. As a result, it can be noticed that computing the matrices $\boldsymbol{\Phi}(p)$ and $\boldsymbol{E}(p)$ for all $p \in \{1 \ldots p_{\max}\}$ is not more computationally demanding than just computing them for $p = p_{\max}$. In both cases, the overall complexity is $3np_{\max}^2 + O(p_{\max}^3)$.

The complete pseudo code for computing $\boldsymbol{E}(p)$ for all $p \in \{1 \ldots p_{\max}\}$ is presented in table I. Note that the calculation of the matrices $\boldsymbol{\Psi}(p)$ and $\boldsymbol{\Phi}(p)$ is not even required.

TABLE I
RECURSIVE COMPUTATION OF $\boldsymbol{E}(p)$

	Cost
Initialization	
compute $\boldsymbol{w}(p)$ for all $p = 1 \ldots p_{\max}$, $\boldsymbol{\varphi}(0) = [\,]$, $\boldsymbol{\Xi}(0) = [\,]$	
For $p = 1$ to $p_{\max}$	
Update of the auxiliary matrix $\boldsymbol{\Psi}(p)$	
$\boldsymbol{\psi}_r(p) = \boldsymbol{W}_{\downarrow}(p-1)^H\,\boldsymbol{w}_{\uparrow}(p)$	np
$\boldsymbol{\psi}_l(p) = \boldsymbol{W}_{\uparrow}(p-1)^H\,\boldsymbol{w}_{\downarrow}(p)$	np
$\psi_{lr}(p) = \boldsymbol{w}_{\downarrow}(p)^H\,\boldsymbol{w}_{\uparrow}(p)$	n
Update of the auxiliary matrix $\boldsymbol{\Xi}(p)$	
$\boldsymbol{\xi}(p) = \boldsymbol{w}_{\uparrow}(p) - \boldsymbol{W}_{\downarrow}(p-1)\,\boldsymbol{\psi}_r(p) - \boldsymbol{w}_{\downarrow}(p)\,\psi_{lr}(p)$	np
$\boldsymbol{\Xi}(p) = \left[\;\boldsymbol{\Xi}(p-1) - \boldsymbol{w}_{\downarrow}(p)\,\boldsymbol{\psi}_l(p)^H\;\middle|\;\boldsymbol{\xi}(p)\;\right]$	np
Computation of $\boldsymbol{E}(p)$ from $\boldsymbol{\Xi}(p)$	
$\boldsymbol{\varphi}(p) = \left[\begin{array}{c} \boldsymbol{\varphi}(p-1) + \mu(p)\,\boldsymbol{\psi}_l(p) \\ \boldsymbol{\psi}_r(p)^H\boldsymbol{\nu}(p-1) + \mu(p)\,\psi_{lr}(p)^* \end{array}\right]$	$2p$
$\boldsymbol{E}(p) = \boldsymbol{\Xi}(p) - \frac{1}{1-\|\boldsymbol{\nu}(p)\|^2}\,(\boldsymbol{W}_{\downarrow}(p)\,\boldsymbol{\nu}(p))\,\boldsymbol{\varphi}(p)^H$	$2np$

V. SIMULATION RESULTS

Section V-A illustrates the relevance of our error bounds. Then the ESTER method is applied to synthetic signals (sections V-B, V-C) and to a musical signal (section V-D).

A. Relevance of the a priori and a posteriori error bounds

In this section, the relevance of the a priori and a posteriori error bounds is illustrated. The test signal is a sum of $r = 20$ undamped complex exponentials of the same amplitude $\alpha_k = 1$, whose frequencies are randomly distributed in $\left[-\frac{1}{2}, \frac{1}{2}\right]$. The upper condition number of the signal subspace satisfies $1 \leq \kappa_2 \leq 1.05$. The singular vectors forming the matrices $\boldsymbol{W}(p)$ have been obtained by computing the SVD of a Hankel data matrix of $n = 512$ rows and $l = 512$ columns, containing the $n + l - 1 = 1023$ samples of the whole signal. For all $p \in \{1 \ldots p_{\max} = r\}$, the eigenvalues $\{\widehat{z}_{(p,m)}\}_{m \in \{1 \ldots p\}}$ of the matrix $\boldsymbol{\Phi}(p)$ have been computed.

In figure 1-a, the solid line represents the eigenvalue errors obtained for $p = 7 < r$, *i.e.*

$$\left\{ \min_{k \in \{1 \ldots r\}} |\widehat{z}_{(7,m)} - z_k| \right\}_{m \in \{1 \ldots 7\}}$$

sorted in increasing order. The dotted line represents the corresponding a priori error bounds, *i.e.*

$$\left\{ \kappa_2 \, \sigma_{\min} \left(\boldsymbol{W}_\uparrow(7) - \widehat{z}_{(7,m)} \, \boldsymbol{W}_\downarrow(7) \right) \right\}_{m \in \{1 \ldots 7\}}$$

It can be noticed that the dotted line is above the solid line (as expected), and more importantly that the variations of the dotted line follow those of the solid line, which suggests that the a priori error bound is relevant.
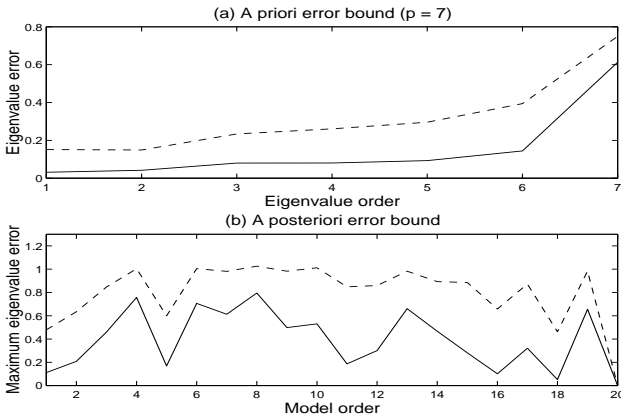

Fig. 1. A priori and a posteriori error bounds
(a) a priori error bound *vs.* eigenvalue error at order $p = 7$
(b) a posteriori error bound *vs.* maximum eigenvalue error as a function of p

In figure 1-b, the solid line represents the maximum eigenvalue error obtained for all modeling orders, *i.e.*

$$\max_{m \in \{1 \ldots p\}} \min_{k \in \{1 \ldots r\}} |\widehat{z}_{(p,m)} - z_k|$$

as a function of p. The dotted line represents the corresponding a posteriori error bounds, *i.e.* $\kappa_2 \, \|\boldsymbol{E}(p)\|_2$ as a function of p. As for the a priori error bound, it can be noticed that the dotted line is above the solid line (as expected), and that the variations of the dotted line follow those of the solid line. This suggests that the a posteriori error bound is a relevant criterion for minimizing the bias of the estimated eigenvalues.
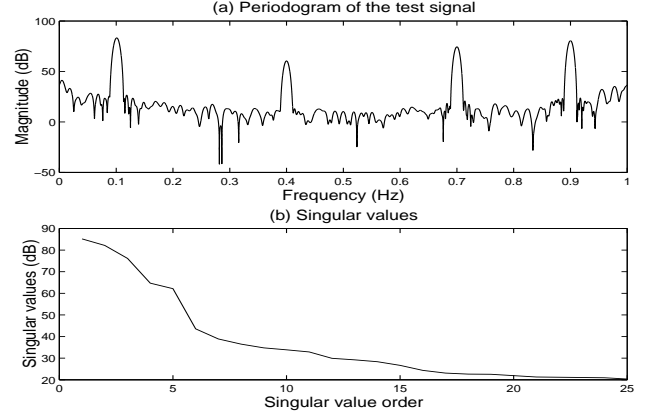

Fig. 2. Periodogram and singular values of a synthetic signal
(a) periodogram of the synthetic signal in dB
(b) singular values of the synthetic signal sorted in decreasing order

B. Model order selection for a synthetic signal

The test signal is a sum of $r = 5$ undamped complex exponentials plus a complex colored noise. For each $k \in \{1, \ldots, r\}$, the kth exponential is characterized by its amplitude α_k and its pole $z_k = e^{i2\pi f_k}$, where $f_k \in \mathbb{R}$ is its frequency. The values of the parameters are given in table II. The additive noise has been obtained by applying the filter

TABLE II
PARAMETERS OF THE SYNTHETIC SIGNAL

Frequency	0.1 Hz	0.102 Hz	0.4 Hz	0.7 Hz	0.9 Hz
Amplitude	100	100	10	50	100

$H(z) = \frac{1}{1 - 0.95\,z^{-1}}$ to a complex white gaussian noise, whose variance has been chosen so that the resulting Signal Noise Ratio (SNR) is 40 dB. The periodogram of the resulting test signal is represented in figure 2-a. It was computed from a signal of length 255, multiplied by a Blackman window, chosen for its high leakage rejection (-57 dB), and zero-padded to obtain 65536 points in the frequency domain. In particular, it can be noticed that the two complex exponentials of lowest frequency are not resolved by the periodogram.

The ESTER method is compared to several other signal enumeration techniques, among which three Information Theoretic Criteria (ITC), known as the *Akaike Information Criterion* (AIC) [5], the *Minimum Description Length* (MDL) [5], and the *Efficient Detection Criterion* (EDC) [9] which is known to be a robust generalization of AIC and MDL. These methods consist in minimizing a cost function which involves the singular values $\{\sigma_1, \ldots, \sigma_n\}$:

$$\text{ITC}(p) = -(n-p)\,l \ln \left(\frac{\left(\prod\limits_{q=p+1}^{n} \sigma_q^2 \right)^{\frac{1}{n-p}}}{\frac{1}{n-p} \sum\limits_{q=p+1}^{n} \sigma_q^2} \right) + p\,(2n-p)\,C(l) \quad (23)$$

where $C(l)$ is a function of l. The AIC criterion is defined by choosing $C(l) = 1$ and the MDL criterion is defined by choosing $C(l) = \frac{1}{2}\ln(l)$. The EDC criteria are obtained for all

functions of l such that $\lim\limits_{l \to +\infty} \frac{C(l)}{l} = 0$ and $\lim\limits_{l \to +\infty} \frac{C(l)}{\ln(\ln(l))} = +\infty$. We chose $C(l) = \sqrt{l \ln(\ln(l))}$, for which we obtained the best results. The singular values have been obtained by computing the SVD of a Hankel data matrix containing $n = 128$ rows and $l = 128$ columns, involving the $n+l-1 = 255$ samples of the whole signal. Figure 2-b displays the highest $p_{\max} = 25$ singular values. Note that the singular values do not present a significant decrease beyond $p = 5$.
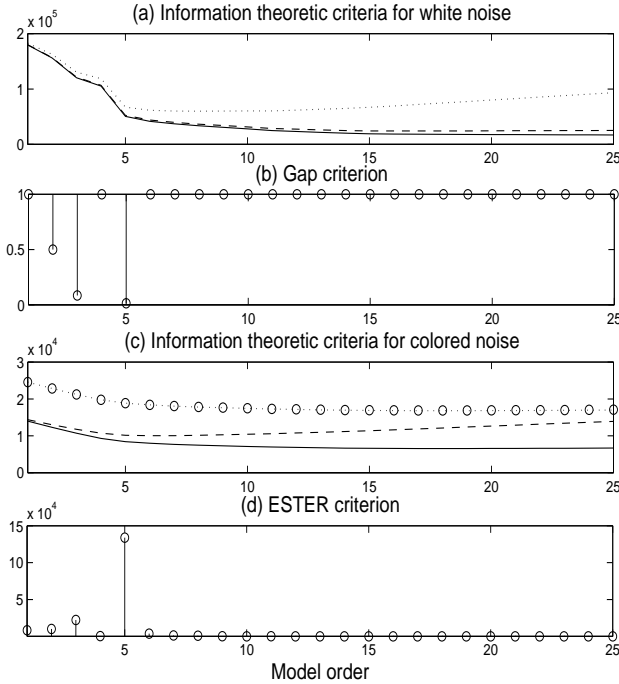

Fig. 3.   Model order selection for the synthetic signal
(a) AIC, MDL, and EDC criteria
(b) Gap criterion
(c) C_1, C_2, C_{m1}, and C_{m2} criteria
(d) ESTER criterion

Figure 3-a displays the values of the AIC (solid line), MDL (dashed line) and EDC (dotted line) criteria, such as formulated in the above equation, for $p \in \{1 \ldots p_{\max}\}$. None of them reaches a minimum at $p = r = 5$. This failure might be explained by the presence of the surrounding noise, whose power spectral density is not uniform, contrary to the additive white noise hypothesis on which these estimators basically rely. As expected, the EDC criterion is more robust than AIC and MDL, but its minimum is obtained for $p = 8$. Figure 3-b represents the criterion proposed in [12] for detecting the gap in the singular values decrease. It can be noticed that this criterion selects the right value $p = r = 5$, but the value $p = 3$ is almost as much emphasized. Figure 3-c displays new ITC criteria proposed in [16] to address the colored noise case[5]. The best results were obtained with C_2, which reaches a minimum at $p = 6 \simeq r$.

Finally, figure 3-d displays the ESTER criterion $J(p)$ for $p \in \{1, \ldots, p_{\max}\}$. It can be noticed that the global maximum

---

[5]These new criteria are referred to as C_1 (solid line), C_2 (dashed line), C_{m1} (dotted line) and C_{m2} (circles line). The common value of the parameters M_1 and M_2 defined in [16] was set to $\frac{n}{2} - 1$.

TABLE III
RATES OF SUCCESSFUL SIGNAL ENUMERATION FOR VARIOUS N

N	125	250	500
AIC	8 %	2 %	0 %
MDL	48 %	61 %	70 %
EDC	38 %	58 %	77 %
gap criterion	47 %	63 %	76 %
C2	35 %	44 %	44 %
ESTER	48 %	63 %	76 %

TABLE IV
RATES OF SUCCESSFUL SIGNAL ENUMERATION FOR VARIOUS SNR

SNR	10 dB	20 dB	30 dB
AIC	3 %	2 %	2 %
MDL	45 %	61 %	65 %
EDC	18 %	58 %	77 %
gap criterion	28 %	63 %	75 %
C2	13 %	44 %	61 %
ESTER	36 %	63 %	76 %

is reached at $p = r = 5$, despite the surrounding noise, which was not included in the model.

C. Statistical performance comparison

Below, the ESTER method and the above-mentioned signal enumeration techniques are applied to various synthetic signals. These signals consist of a sum of real-valued and undamped sinusoids, plus a colored noise. The number of sinusoids is uniformly distributed between 1 and 10, so that the model order r belongs to $\{2 \ldots 20\}$. Their amplitudes, phases and frequencies are randomly distributed in the intervals $[1, 10]$, $]-\pi, \pi]$ and $]-\frac{1}{2}, \frac{1}{2}]$. The additive noise is obtained by filtering a white gaussian noise by the high-pass filter $1 - 0.5 \, z^{-1}$ (whose rejection is lower than 10 dB).

As proposed in section IV, the robustness of the ESTER method is improved here by detecting the greatest value of p for which the ESTER criterion $J(p)$ reaches a local maximum which is greater than one tenth of the global maximum. Tables III and IV show the rates of successful signal enumeration, averaged over 10000 independent runs, for various values of the window length N and the SNR (other analysis parameters are $n = \lfloor N/2 \rfloor$ and $p_{\max} = 22$). In table III, the SNR is fixed to 20 dB, and the experiment is run for $N = 125$, $N = 250$ and $N = 500$. In table IV, N is fixed to 250, and the experiment is run for a low SNR (10 dB), a moderate SNR (20 dB) and a high SNR (30 dB). The obtained percentages have been rounded towards the nearest integer, since the number of independent runs (10000) guarantees that the confidence interval lies between $\pm 1\%$ around the estimated rate, for a 95% confidence level.

It can be noticed that the AIC is ineffective for processing these synthetic data. Besides, the successful rate of the C_2 criterion [16] is always lower than that of all the other methods. The MDL criterion seems to be more robust than ESTER to low SNRs. However, ESTER outperforms MDL both for high values of N and high SNRs. Compared to the EDC criterion, ESTER presents similar performance for high N and high SNRs, but it is more robust to low values of N and low SNRs. Finally, ESTER behaves similarly to the gap

criterion [12] for all values of N. However, it is more robust to low SNRs. It can be noticed that although the ESTER method relies on a noise-free signal model, its performance does not collapse at low SNRs.

D. Model order selection for a musical signal
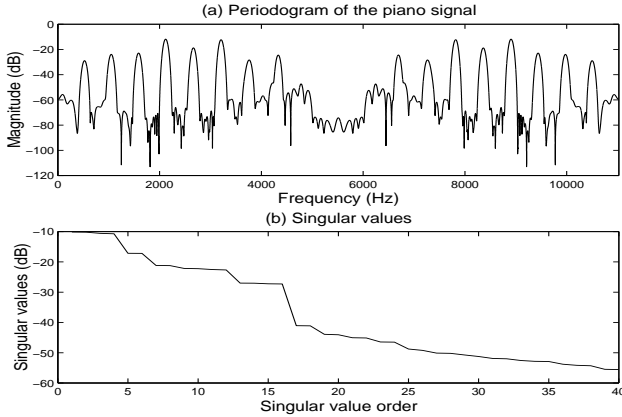

Fig. 4.   Periodogram and singular values of a piano signal
   (a) periodogram of the piano signal in dB
   (b) singular values of the piano signal sorted in decreasing order

This section illustrates the application of the ESTER method to a musical signal. The study deals with a piano tone, C5, sampled at 11025 Hz, from which a segment of 255 samples (23 ms) has been extracted. Note that audio signals often require some pre-processing before applying the ESPRIT algorithm. For example, signals with a high number of sinusoids (typically low-pitched sounds) may first be decomposed into several sub-band signals (via filtering/decimating, as proposed in [33]). In this example, this pre-processing is not used, since the chosen piano tone has few sinusoidal components. In other respects, it is well known that the energy of audio signals is not evenly distributed over frequencies. Therefore we used a pre-emphasis filter obtained by linear prediction at order 7 to compensate the energy decrease.

The periodogram of the filtered piano signal is displayed in figure 4-a. In this figure, sixteen sharp spectral peaks clearly rise above the surrounding noise level. The $p_{\max} = 40$ highest singular values of the data matrix[6] are represented in figure 4-b. Clearly, these singular values collapse beyond $p = 16$, which suggests a modeling order equal to 16.

Figure 5-a displays the AIC (solid line), MDL (dashed line) and EDC (dotted line) criteria. Only EDC reaches a minimum at $p = 16$. However this minimum is not substantially lower than the neighboring values. Figure 5-b represents the gap criterion proposed in [12]. Contrary to the above-mentioned ITC, this criterion here selects the right value $p = 16$. Figure 5-c displays the ITC criteria proposed in [16], with the same parameters as in section V-B. None of them reaches a minimum at $p = 16$. Finally, figure 5-d displays the ESTER criterion $J(p)$ for all $p \in \{1, \dots, p_{\max}\}$. The global

---

[6]The singular values have been obtained by computing the SVD of a Hankel data matrix containing $n = 128$ rows and $l = 128$ columns, as in section V-B.

maximum is reached at $p = 16$. It can also be noticed that the error bounds obtained for lower values of p are relevant. Indeed, high values are reached at $p = 4, 6, 12$, which in fact correspond to small jumps in the decrease of the singular values (represented in figure 4-b). Therefore, the ESTER method gives the expected model order, and moreover the error bounds can be used to quantify the adequacy of a possible lower modeling order. In particular, it can be noticed that odd model orders do not fit the signal. Indeed, since this signal is real-valued and centered, its spectrum is hermitian symmetric with no constant component, which underlies an even model order.
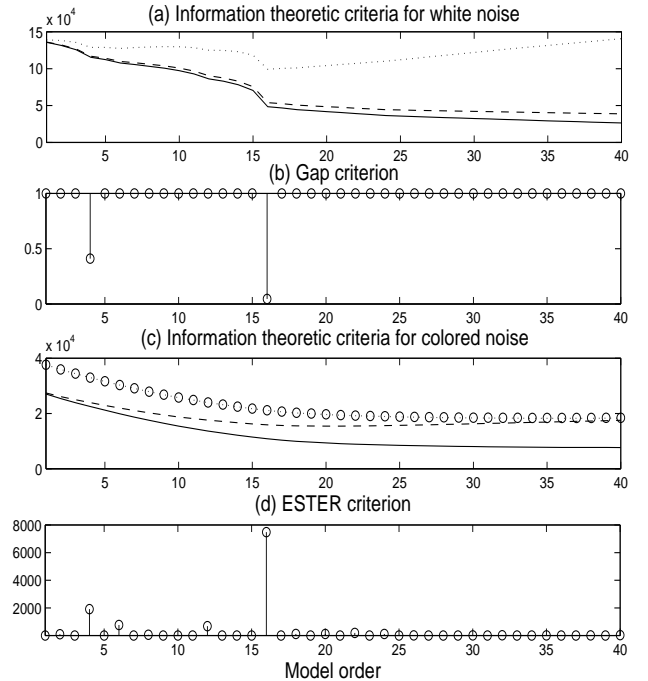

Fig. 5.   Model order selection for the piano signal
   (a) AIC, MDL, and EDC criteria
   (b) Gap criterion
   (c) C_1, C_2, C_{m1}, and C_{m2} criteria
   (d) ESTER criterion

VI. Conclusions

In this paper, we described how the estimation of a noiseless ESM model is affected by applying the ESPRIT algorithm with an erroneous model order. If the model order is overestimated, the true poles are among those estimated. On the contrary, if the order is underestimated, the estimated poles can be seen as approximations of some of the true ones. In this last case, an *a posteriori error bound* was given, which can be computed without knowing the exact model order. Following from this observation, we introduced the ESTER criterion for selecting an appropriate model order. Since the initial method was computationally expensive, we proposed a fast algorithm for recursively computing the a posteriori error bounds. Then, we showed the relevance of our criterion as an error bound, and we illustrated the performance of the ESTER criterion on synthetic signals and on a piano signal. In addition, we noted

that the error bounds could be used to quantify the adequacy of a possible lower modeling order. Finally, although it was designed for the ESPRIT algorithm, the ESTER criterion can be used with any High Resolution method. It can also be adapted to estimate the order of the more general *Polynomial Amplitude Complex Exponentials* (PACE) model [34].

In presence of noise, we mentioned in section IV that a robust way of selecting the modeling order consists in detecting the greatest value of p for which the ESTER criterion $J(p)$ reaches a local maximum greater than a fraction of its global maximum. Future work will be dedicated to an analysis of the effect of noise, in order to better exploit the information provided by the ESTER criterion.

Appendix

Proof of theorem III.2. If $\widehat{z} = z_k$ for some $k \in \{1 \dots r\}$, the assertion is trivial, so we may assume that $\forall k \in \{1 \dots r\}$, $\widehat{z} \neq z_k$. Let $\widehat{v}$ be a given unitary vector, and define the residual

$$e(p) \triangleq (\boldsymbol{W}_\uparrow(p) - \widehat{z}\,\boldsymbol{W}_\downarrow(p))\,\widehat{v}. \qquad (24)$$

Since $\boldsymbol{W}(p)\,\widehat{v} = \boldsymbol{W}(r)\begin{bmatrix}\widehat{v}\\0\end{bmatrix}$, equation (24) yields

$$e(p) = (\boldsymbol{W}_\uparrow(r) - \widehat{z}\,\boldsymbol{W}_\downarrow(r))\begin{bmatrix}\widehat{v}\\0\end{bmatrix}. \qquad (25)$$

Substituting equations (3) and (4) into equation (25) yields

$$e(p) = \boldsymbol{V}_\downarrow^n\,(\boldsymbol{D} - \widehat{z}\,\boldsymbol{I}_r)\,\boldsymbol{G}^{-1}\begin{bmatrix}\widehat{v}\\0\end{bmatrix}. \qquad (26)$$

Since $\forall k \in \{1 \dots r\}$, $\widehat{z} \neq z_k$, $\boldsymbol{D} - \widehat{z}\,\boldsymbol{I}_r$ is non-singular. Therefore, equation (26) yields

$$\begin{bmatrix}\widehat{v}\\0\end{bmatrix} = \boldsymbol{G}\,(\boldsymbol{D} - \widehat{z}\,\boldsymbol{I}_r)^{-1}\,\boldsymbol{V}_\downarrow^{n\dagger}\,e(p). \qquad (27)$$

Consider a given diagonal matrix $\boldsymbol{\Lambda}$ whose diagonal coefficients are positive. Then equation (27) is equivalent to

$$\begin{bmatrix}\widehat{v}\\0\end{bmatrix} = \boldsymbol{G}\,\boldsymbol{\Lambda}\,(\boldsymbol{D} - \widehat{z}\,\boldsymbol{I}_r)^{-1}\,\boldsymbol{\Lambda}^{-1}\,\boldsymbol{V}_\downarrow^{n\dagger}\,e(p). \qquad (28)$$

Applying the 2-norm into equation (28) yields

$$\|\widehat{v}\|_2 \leq \|\boldsymbol{G}\,\boldsymbol{\Lambda}\|_2\,\left\|(\boldsymbol{D} - \widehat{z}\,\boldsymbol{I}_r)^{-1}\right\|_2\,\left\|\boldsymbol{\Lambda}^{-1}\boldsymbol{V}_\downarrow^{n\dagger}\right\|_2\,\|e(p)\|_2. \qquad (29)$$

Since $\boldsymbol{W}(r)$ is orthonormal and $\boldsymbol{V}^n\,\boldsymbol{\Lambda} = \boldsymbol{W}(r)\,\boldsymbol{G}\,\boldsymbol{\Lambda}$, $\|\boldsymbol{G}\,\boldsymbol{\Lambda}\|_2 = \|\boldsymbol{V}^n\,\boldsymbol{\Lambda}\|_2 = \sigma_{\max}(\boldsymbol{V}^n\,\boldsymbol{\Lambda})$. Moreover, $(\boldsymbol{D} - \widehat{z}\,\boldsymbol{I}_r)^{-1}$ is diagonal with diagonal entries $\frac{1}{z_k - \widehat{z}}$, thus $\left\|(\boldsymbol{D} - \widehat{z}\,\boldsymbol{I}_r)^{-1}\right\|_2 = \frac{1}{\min\limits_{k \in \{1 \dots r\}} |\widehat{z} - z_k|}$. Since $\boldsymbol{V}_\downarrow^n$ is full-rank, the singular values of $\boldsymbol{\Lambda}^{-1}\boldsymbol{V}_\downarrow^{n\dagger}$ are the inverses of those of $\boldsymbol{V}_\downarrow^n\,\boldsymbol{\Lambda}$, so that $\left\|\boldsymbol{\Lambda}^{-1}\boldsymbol{V}_\downarrow^{n\dagger}\right\|_2 = \frac{1}{\sigma_{\min}(\boldsymbol{V}_\downarrow^n\,\boldsymbol{\Lambda})}$. Finally, since $\widehat{v}$ is unitary, equation (29) yields

$$\min_{k \in \{1 \dots r\}} |\widehat{z} - z_k| \leq \frac{\sigma_{\max}(\boldsymbol{V}^n\,\boldsymbol{\Lambda})}{\sigma_{\min}(\boldsymbol{V}_\downarrow^n\,\boldsymbol{\Lambda})}\,\|e(p)\|_2. \qquad (30)$$

Note that equation (30) is satisfied for all unitary vector $\widehat{v}$ and all matrix $\boldsymbol{\Lambda} \in \mathrm{diag}(\mathbb{R}_+^{*\,r})$. Consequently, equation (7) follows from equation (30). $\qquad \square$

Proof of proposition IV.1. It can be noticed that

$$\boldsymbol{E}(p) = \left(\boldsymbol{W}_\uparrow(p)\,\boldsymbol{W}_\uparrow(p)^\dagger - \boldsymbol{W}_\downarrow(p)\,\boldsymbol{W}_\downarrow(p)^\dagger\right)\boldsymbol{W}_\uparrow(p).$$

Applying the 2-norm yields

$$\|\boldsymbol{E}(p)\|_2 \leq \mathrm{dist}\,(\boldsymbol{W}_\uparrow(p), \boldsymbol{W}_\downarrow(p))\,\|\boldsymbol{W}_\uparrow(p)\|_2 \qquad (31)$$

where

$$\mathrm{dist}\,(\boldsymbol{W}_\uparrow(p), \boldsymbol{W}_\downarrow(p)) \triangleq \left\|\boldsymbol{W}_\uparrow(p)\,\boldsymbol{W}_\uparrow(p)^\dagger - \boldsymbol{W}_\downarrow(p)\,\boldsymbol{W}_\downarrow(p)^\dagger\right\|_2$$

is the distance between the subspaces $\mathrm{Span}(\boldsymbol{W}_\uparrow(p))$ and $\mathrm{Span}(\boldsymbol{W}_\downarrow(p))$, which satisfies $\mathrm{dist}\,(\boldsymbol{W}_\uparrow(p), \boldsymbol{W}_\downarrow(p)) \leq 1$, as shown in [32, pp. 76-77]. Since $\|\boldsymbol{W}_\uparrow(p)\|_2 \leq \|\boldsymbol{W}(p)\|_2 = 1$, the result follows from equation (31). $\qquad \square$

References

[1] R. J. McAulay and T. F. Quatieri, "Speech analysis and synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 4, Aug. 1986.

[2] B. David, G. Richard, and R. Badeau, "An EDS modeling tool for tracking and modifying musical signals," in *Proc. of SMAC 03*, vol. 2, Stockholm, Sweden, Aug. 2003, pp. 715–718.

[3] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT–A subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 5, pp. 1340–1342, Oct. 1986.

[4] G. Bienvenu and L. Kopp, "Optimality of high-resolution array processing using the eigensystem method," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, no. 5, pp. 1235–1245, Oct. 1983.

[5] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, no. 2, pp. 387–392, Apr. 1985.

[6] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. of the 2nd International Symposium on Information Theory*, B. N. Petrov and F. Csaki, Eds. Budapest: Akademia Kiado, 1973, pp. 267–281.

[7] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[8] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[9] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in presence of white noise," *Journal of Multivariate Analysis*, vol. 20, no. 1, pp. 1–25, 1986.

[10] F. Gini and F. Bordoni, "On the behavior of information theoretic criteria for model order selection of InSAR signals corrupted by multiplicative noise," *Signal Processing*, vol. 83, pp. 1047–1063, 2003.

[11] A. P. Liavas and P. A. Regalia, "On the behavior of Information Theoretic Criteria for model order selection," *IEEE Trans. Signal Processing*, vol. 49, no. 8, pp. 1689–1695, Aug. 2001.

[12] A. P. Liavas, P. A. Regalia, and J.-P. Delmas, "Blind channel approximation: effective channel order determination," *IEEE Trans. Signal Processing*, vol. 47, no. 12, pp. 3336–3344, Dec. 1999.

[13] J. Grouffaud, P. Larzabal, and H. Clergeot, "Some properties of ordered eigenvalues of a Wishart matrix: application in detection test and model order selection," in *Proc. of ICASSP'96*, vol. 5. IEEE, 1996, pp. 2465–2468.

[14] D. Kundu and A. Mitra, "Detecting the number of signals for an undamped exponential model using cross-validation approach," *Signal Processing*, vol. 80, no. 3, pp. 525–534, 2000.

[15] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals when the noise covariance matrix is arbitrary," *Journal of Multivariate Analysis*, vol. 20, no. 1, pp. 26–49, 1986.

[16] Q. T. Zhang and K. M. Wong, "Information theoretic criteria for the determination of the number of signals in spatially correlated noise," *IEEE Trans. Signal Processing*, vol. 41, no. 4, pp. 1652–1663, Apr. 1993.

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

[18] W. B. Bishop and P. M. Djuric, "Model order selection of damped sinusoids in noise by predictive densities," *IEEE Trans. Signal Processing*, vol. 44, no. 3, pp. 611–619, Mar. 1996.

[19] B. Rao and K. Hari, "Performance analysis of ESPRIT and TAM in determining the direction of arrival of plane waves in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 12, pp. 1990–1995, Dec. 1989.

[20] P. Stoica and T. Söderström, "Statistical Analysis of MUSIC and Subspace Rotation Estimates of Sinusoidal Frequencies," *IEEE Trans. Signal Processing*, vol. 39, pp. 1836–1847, Aug. 1991.

[21] A. Eriksson, P. Stoica, and T. Soderstrom, "Second-order properties of MUSIC and ESPRIT estimates of sinusoidal frequencies in high SNR scenarios," in *IEE Proceedings on Radar, Sonar and Navigation*, vol. 140, no. 4, Aug. 1993, pp. 266–272.

[22] A. Kangas, P. Stoica, and T. Soderstrom, "Finite sample and modelling error effects on ESPRIT and MUSIC direction estimators," in *IEE Proceedings on Radar, Sonar and Navigation*, vol. 141, no. 5, Oct. 1994, pp. 249–255.

[23] 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 *Proc. of the Military Communications Conference MILCOM'99*, Atlantic City, USA, 1999.

[24] F. Li, R. J. Vaccaro, and D. W. Tufts, "Performance analysis of the state-space realization (TAM) and ESPRIT algorithms for DOA estimation," *IEEE Trans. Antennas Propagat.*, vol. 39, no. 3, pp. 418–423, Mar. 1991.

[25] R. Badeau, B. David, and G. Richard, "Selecting the modeling order for the ESPRIT high resolution method: an alternative approach," in *Proc. of ICASSP'04*, vol. 2. Montreal, Quebec: IEEE, May 2004, pp. 1025–1028.

[26] R. Badeau, R. Boyer, and B. David, "EDS parametric modeling and tracking of audio signals," in *Proc. of Int. Conf. on Digital Audio Effects DAFx-02*, Hamburg, Sept. 2002, pp. 139–144.

[27] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 5, pp. 814–824, May 1990.

[28] R. Badeau, G. Richard, and B. David, "Sliding window adaptive SVD algorithms," *IEEE Trans. Signal Processing*, vol. 52, no. 1, pp. 1–10, Jan. 2004.

[29] R. Badeau, B. David, and G. Richard, "Fast Approximated Power Iteration Subspace Tracking," *IEEE Trans. Signal Processing*, Aug. 2005.

[30] ——, "Yet Another Subspace Tracker," in *Proc. of ICASSP'05*, vol. 4. Philadelphia, PA, USA: IEEE, Mar. 2005, pp. 329–332.

[31] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge: Cambridge University Press, 1985.

[32] G. H. Golub and C. F. V. Loan, *Matrix computations*, 3rd ed. Baltimore and London: The Johns Hopkins University Press, 1996.

[33] J. Laroche, "The use of the Matrix Pencil method for the spectrum analysis of musical signals," *Journal of the Acoustical Society of America*, vol. 94, no. 4, pp. 1958–1965, Oct. 1993.

[34] R. Badeau, B. David, and G. Richard, "High resolution spectral analysis of mixtures of complex exponentials modulated by polynomials," *IEEE Trans. Signal Processing*, (accepted for publication).