Higher-Order Tensor-Based Method for Delayed Exponential Fitting

Rémy Boyer*, Lieven De Lathauwer and Karim Abed-Meraim

Abstract

We present subspace-based schemes for the estimation of the poles (angular frequencies and damping factors) of a sum of damped and delayed sinusoids. In our model each component is supported over a different time frame, depending on the delay parameter. Classical subspace based methods are not suited to handle signals with varying time supports. In this contribution, we propose solutions based on the approximation of a partially structured Hankel-type tensor on which the data are mapped. We show, by means of several examples, that the approach based on the best rank-(R_1, R_2, R_3) approximation of the data tensor outperforms the current tensor and matrix-based techniques in terms of the accuracy of the angular frequency and damping factor parameter estimates, especially in the context of difficult scenarios as in the low SNR regime and for closely spaced sinusoids.

Index Terms

Subspace-based parameter estimation, damped and delayed sinusoids, higher-order tensor, singular value decomposition, rank reduction, conditional Cramér-Rao bound.

EDICS: SSP-PARE Parameter estimation

Rémy Boyer is with Laboratoire des Signaux et Systemes (LSS-Supélec), CNRS, Université Paris XI, SUPELEC, Paris, France, Email: remy.boyer@lss.supelec.fr, Tel. : 33 1 69 85 17 12 - Fax : 33 1 69 85 17 69

Lieven De Lathauwer is with the Equipes Traitement des Images et du Signal (ETIS) (CNRS, ENSEA, UCP), Cergy-Pontoise, France; he also has a honorary research position with the K.U.Leuven, Leuven, Belgium. Part of this work is supported by (1) the Research Council K.U.Leuven: Concerted Research Action GOA-AMBioRICS, (2) the Flemish Government: the F.W.O. Research Communities ICCoS and ANMMM, F.W.O. Project G.0321.06, Tournesol 2005 - Project T20013, and (3) the Belgian Federal Government: Interuniversity Poles of Attraction Programme IUAP V-22., Email: delathau@ensea.fr, Tel: +33-(0)1-30.73.66.10

Karim Abed-Meraim is with the Ecole Nationale Supérieure des Télécommunications, Laboratoire Traitement du Signal et des Images (TSI), Paris, France, Email: abed@tsi.enst.fr, Tel: (33)(0) 1 45.81.73.27, Fax : (33)(0) 1 45.81.71.44

I. INTRODUCTION

Estimation of the poles of a sum of windowed sinusoidal components is a key problem in spectral analysis [22], transient audio signal modeling [2], biomedical signal processing [39], and in the analysis of power systems [15]. Among the numerous methods that have been proposed, the "subspace" methods [1], [22], [33], [38], based on the shift invariance property of the signal subspace, form an important class. Classically, these methods are used for the model-parameter estimation of a sum of Exponentially Damped Sinusoids (EDS) with the same time support. Each component has the same length, namely, the length of the analysis window. In this contribution, we use a more sophisticated model, called the *Partial* Damped and Delayed Sinusoidal model (PDDS), which is a generalization of the EDS model. In this model, we add time-delay parameters that allow time-shifting each burst of EDS components. This modification is useful for the compact modeling of fast time-varying signals. For instance, [2], [18] contain an application example in the context of audio modeling.

The link between multilinear algebra and exponential signals was first made in [24], [34]. By representing exponentials as higher-order rank-1 tensors, it was possible to relax previous bounds on the number of exponentials that can theoretically be recovered in multidimensional harmonic analysis. However, the data model on which [24], [34] are based, does not carry over to the PDDS model. Recently, multilinear algebra based variants of subspace methods for EDS modelling have been derived. In [30] the Tucker decomposition or Higher-Order Singular Value Decomposition (HOSVD), introduced in [36] and further discussed in [9], [11], and the best rank-(R_1, R_2, R_3) approximation, introduced in [28] and further discussed in [10], [11], [12], are used for the estimation of EDS from single-burst or multi-burst measurements. In this paper we will use these tools for PDDS modeling. The techniques proposed in Sections V-A and V-B have briefly been described in the conference papers [3] and [5], respectively. The algorithm of Section V-A is in fact equivalent to the matrix technique proposed in [2], as will be explained in Section V-A.

By means of various examples, we will show that our approach outperforms the current tensor and matrix methods for the estimation of the angular frequency and damping factor parameters.

The paper is organized as follows. In Section II we formally introduce the PDDS model. Section III introduces the prerequisite concepts of multilinear algebra. The PDDS modeling problem is cast in a multilinear algebraic framework in Section IV. Specific algorithms are discussed in Section V. Their performance is illustrated in Section VI in comparison to the Conditional Cramér-Rao Bound (CCRB) (derived in Appendix VIII-B). Section VII is the conclusion.

II. THE PDDS MODEL AND ITS MULTI-BURST STRUCTURE

A. Definition of the model

We define the complex M_k-PDDS model, for $n \in [0 : N-1]$ and given $k \in [0 : K-1]$, by

$$\hat{s}_k(n) = \sum_{m=1}^{M_k} \alpha_{m,k} \; . \; z_{m,k}^{n-t_k} \; . \; \psi(n-t_k) \tag{1}$$

where $\alpha_{m,k} = a_{m,k} e^{i\phi_{m,k}}$ is the m-th complex amplitude, with $a_{m,k}$ and $\phi_{m,k}$ the nonzero real amplitude and initial phase, respectively. $z_{m,k} = e^{d_{m,k}+i\omega_{m,k}}$ is the corresponding pole, with $d_{m,k}$ the (negative) damping factor and $\omega_{m,k}$ the angular frequency. t_k is a delay parameter, unique for the whole set of M_k EDS waveforms. The set of M_k undelayed EDS waveforms is briefly denoted as M_k-EDS. The Heaviside function $\psi(n)$ is equal to "1" for $n \geq 0$ and "0" otherwise. Now consider a set of delay parameters $\{t_k\}_{k\in[0:K-1]}$, with $t_0 = 0$, $t_{K-1} = N-1$, $0 \leq t_k < t_{k+1} \leq N-1$ and $B_k = t_{k+1} - t_k$. The noisy M-PDDS model, where $M = \sum_{k=0}^{K-1} M_k$, is then

$$s(n) = \hat{s}(n) + \sigma w(n) \quad \text{where} \quad \hat{s}(n) = \sum_{k=0}^{K-1} \hat{s}_k(n). \tag{2}$$

The final model can be seen as a sum of independently time-shift bursts corrupted by a zero-mean additive white Gaussian noise process with variance σ^2, denoted by $\sigma w(n)$.

B. Burst with interference

We assume that the set of time-delays $\{t_k\}_{k\in[0:K-1]}$ is known or has been previously estimated. We can for instance use a time-delay detector-estimator based on the variation of the energy of the time-envelope, which has been shown to provide satisfactory results for transient audio signal modeling [2]. Generally speaking, the choice of the method depends strongly on the target application [16], [19], [20], [23], [25]. The problem of time-delay estimation is not trivial and to the best of our knowledge, there is no method that jointly estimates the time-delays and the poles with an acceptable computational cost.

Our derivation of an algorithm for the estimation of the poles starts from the following observation. According to the noise-free PDDS model definition given in expressions (1) and (2), we define the k-th burst by

$$\hat{x}_k(n-t_k) = \hat{s}_k(n) + \sum_{k=0}^{k-1} \sum_{m=1}^{M_k} \alpha_{m,k} \; z_{m,k}^{n-t_k} \psi(n-t_k), \quad \text{for} \quad n \in [t_k, t_{k+1}-1].$$

As we can see the k-th burst is the sum of the k-th PDDS signal, $\hat{s}_k(n)$, and the sum of the tails associated to all the previous PDDS signals. Now introduce the following K signals

$$\{\hat{x}_k(n),\ n \in [0 : B_k - 1]\}_{k\in[0:K-1]}, \tag{3}$$

which can be obtained from signal $\hat{x}_k(n - t_k)$ by a t_k-sample shift as follows:

$$\begin{aligned}\hat{x}_k(n) &= \hat{s}_k(n + t_k) + \sum_{\ell=0}^{k-1}\sum_{m=1}^{M_\ell} \alpha_{m,\ell}\ z_{m,\ell}^{n+t_k-t_\ell}\psi(n + t_k - t_\ell) \\ &= \underbrace{\hat{s}_k(n + t_k)}_{M_k\text{-EDS}} + \underbrace{\sum_{\ell=0}^{k-1}\sum_{m=1}^{M_\ell} \alpha_{m,\ell}\ z_{m,\ell}^{\sum_{u=\ell}^{k-1} B_u}\ z_{m,\ell}^{n}}_{\text{“interference”}}.\end{aligned}$$

In the above expression, for $n \in [t_k, t_{k+1} - 1]$ and $t_k \geq t_{k-1} \geq \ldots \geq t_1 \geq 0$, we have $\psi(n+t_k-t_{k-1}) = \ldots = \psi(n+t_k-t_1) = \psi(n+t_k) = 1$. In addition, we have $\sum_{u=\ell}^{k-1} B_u = t_k - t_\ell$. Finally, each signal $\hat{x}_k(n)$, $n \in [0 : B_k - 1]$, will be considered as a separate burst of B_k samples consisting of an $\left(\sum_{\ell=0}^{k} M_\ell\right)$-EDS model with varying time supports that is the sum of

(1) an M_k-EDS signal $\hat{s}_k(n + t_k)$,
(2) an interfering attenuated $\left(\sum_{\ell=0}^{k-1} M_\ell\right)$-EDS model.

Depending on the application, B_k may be small. Because the Fourier resolution is $O(B_k^{-1})$, Fourier analysis may not allow for an accurate estimation of the model parameters. Instead, we propose a subspace approach for the estimation of the set of all the poles $\{z_{m,k}\}_{m\in[1:M_k],k\in[0:K-1]}$ from the set of bursts (3).

However, the fact that the bursts have a variable length poses a problem for the joint estimation of the poles. Indeed, subspace-based methods are not well suited to handle signals with varying time supports. In this paper, we present a solution.

C. Maximal and minimal burst length

We define the two following quantities associated with the PDDS model (see Fig. 1):

$$\begin{aligned}B^+ &= \max_k B_k \\ B^- &= \left\lfloor \gamma\ \min_k B_k \right\rfloor + 1\end{aligned}$$

where $\lfloor . \rfloor$ is the integer part of its argument. Parameter γ is strictly smaller than 1. B^+ is the maximal burst length and B^- is less than or equal to the minimal burst length: $B^- \leq B_k \leq B^+$, $\forall k$. In Section IV-B we will explain which constraints have to be satisfied when fixing the value of B^-.

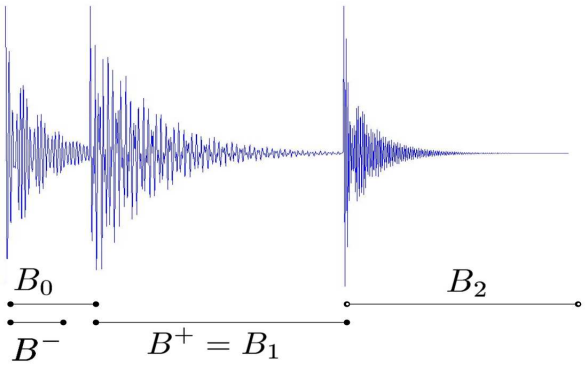


Fig. 1. Maximal and minimal burst length for an example where $K = 3$.

III. Some basic definitions in multilinear algebra

Consider an $(I_1 \times I_2 \times I_3)$ complex-valued tensor $\mathcal{T}$. We write $\mathcal{T} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$. The entries of $\mathcal{T}$ are characterized by three indices, say i_1, i_2, i_3, with $1 \leq i_1 \leq I_1$, $1 \leq i_2 \leq I_2$, $1 \leq i_3 \leq I_3$. A vector obtained by varying index i_1 while keeping i_2 and i_3 fixed, is called a column vector or 1-mode vector. A vector obtained by varying index i_2 while keeping i_1 and i_3 fixed, is called a row vector or 2-mode vector. A vector obtained by varying index i_3 while keeping i_1 and i_2 fixed, is called a 3-mode vector.

The dimensions of the vector spaces generated by the column vectors, row vectors and 3-mode vectors are called column rank (or 1-mode rank) R_1, row rank (or 2-mode rank) R_2 and 3-mode rank R_3, respectively. Contrary to matrices, where column rank and row rank are equal, the mode ranks of a higher-order tensor may be mutually different. However, we always have [37, p. 288]

$$R_1 \leq R_2 R_3 \quad R_2 \leq R_1 R_3 \quad R_3 \leq R_1 R_2. \tag{4}$$

Definition 1: A third-order tensor of which the n-mode rank is equal to R_n, $n \in [1:3]$, is called a rank-(R_1, R_2, R_3) tensor. A rank-$(1,1,1)$ tensor is briefly called a rank-1 tensor.

There are several ways to stack a tensor $\mathcal{T}$ in a matrix format.

Definition 2: We define "matrix representations" $\boldsymbol{T}_{(1)}$, $\boldsymbol{T}_{(2)}$, $\boldsymbol{T}_{(3)}$ of a tensor $\mathcal{T}$ as follows:

$$(\boldsymbol{T}_{(1)})_{i_1,(i_3-1)I_3+i_2} = (\mathcal{T})_{i_1 i_2 i_3}, \tag{5}$$

$$(\boldsymbol{T}_{(2)})_{i_2,(i_3-1)I_3+i_1} = (\mathcal{T})_{i_1 i_2 i_3}, \tag{6}$$

$$(\boldsymbol{T}_{(3)})_{i_3,(i_1-1)I_1+i_2} = (\mathcal{T})_{i_1 i_2 i_3}, \qquad \forall i_1, i_2, i_3. \tag{7}$$

These matrices are of dimension $(I_1 \times I_2 I_3)$, $(I_2 \times I_3 I_1)$, $(I_3 \times I_1 I_2)$, respectively. In $\boldsymbol{T}_{(1)}$ all the column vectors of $\boldsymbol{\mathcal{T}}$ are stacked one after the other. The row vectors and 3-mode vectors are stacked in

$\boldsymbol{T}_{(2)}$ and $\boldsymbol{T}_{(3)}$, respectively. Consequently, the n-mode rank R_n of a tensor $\mathcal{T}$ is equal to the rank of its matrix representation $\boldsymbol{T}_{(n)}$. Note that the definition of $\boldsymbol{T}_{(1)}$ differs from the one given in [11], [28], [36] by a permutation of the columns, which is irrelevant in our context. This variation facilitates the proof of Theorem 1.

Next, we consider the multiplication of a tensor and one or more matrices. For matrices $\boldsymbol{U}^{(1)} \in \mathbb{C}^{I_1 \times J_1}$, $\boldsymbol{U}^{(2)} \in \mathbb{C}^{I_2 \times J_2}$, $\boldsymbol{U}^{(3)} \in \mathbb{C}^{I_3 \times J_3}$ and tensors $\boldsymbol{\mathcal{T}} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$, $\boldsymbol{\mathcal{S}} \in \mathbb{C}^{J_1 \times J_2 \times J_3}$, the expression

$$\boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{S}} \times_1 \boldsymbol{U}^{(1)} \times_2 \boldsymbol{U}^{(2)} \times_3 \boldsymbol{U}^{(3)}, \tag{8}$$

in which $\times_n$ represents the Tucker n-mode product [36], means that

$$(\boldsymbol{\mathcal{T}})_{i_1 i_2 i_3} = \sum_{j_1 j_2 j_3} (\boldsymbol{\mathcal{S}})_{j_1 j_2 j_3} (\boldsymbol{U}^{(1)})_{i_1 j_1} (\boldsymbol{U}^{(2)})_{i_2 j_2} (\boldsymbol{U}^{(3)})_{i_3 j_3}, \quad \forall i_1, i_2, i_3.$$

Eq. (8) can be written in terms of the matrix representations of $\boldsymbol{\mathcal{T}}$ as follows:

$$\begin{aligned}
\boldsymbol{T}_{(1)} &= \boldsymbol{U}^{(1)} \cdot \boldsymbol{S}_{(1)} \cdot \left(\boldsymbol{U}^{(3)} \otimes \boldsymbol{U}^{(2)} \right)^T, \\
\boldsymbol{T}_{(2)} &= \boldsymbol{U}^{(2)} \cdot \boldsymbol{S}_{(2)} \cdot \left(\boldsymbol{U}^{(3)} \otimes \boldsymbol{U}^{(1)} \right)^T, \\
\boldsymbol{T}_{(3)} &= \boldsymbol{U}^{(3)} \cdot \boldsymbol{S}_{(3)} \cdot \left(\boldsymbol{U}^{(1)} \otimes \boldsymbol{U}^{(2)} \right)^T,
\end{aligned}$$

in which $\otimes$ denotes the Kronecker product. Note that, if (8) holds, the columns of $\boldsymbol{U}^{(n)}$ span the space generated by the n-mode vectors of $\mathcal{T}$.

Finally, the Frobenius norm of a tensor $\boldsymbol{\mathcal{T}} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ is defined as

$$\|\boldsymbol{\mathcal{T}}\| = \left(\sum_{i_1 i_2 i_3} |(\boldsymbol{\mathcal{T}})_{i_1 i_2 i_3}|^2 \right)^{\frac{1}{2}}.$$

IV. Tensor formulation of PDDS modeling

In this section, the PDDS modeling problem is cast in a multilinear algebraic framework. In Section IV-A we map the data on a tensor that has a specific structure. In Section IV-B we show that the 1-mode vector space of this tensor is generated by the Vandermonde vectors associated with the PDDS poles. In Section IV-C we explain how the poles can be derived from an arbitrary basis of the 1-mode vector space.

A. Tensor representation of the data

We map the noise-free data $\hat{s}(n)$, $n \in [0 : N-1]$ defined in (2), on the $(B^- \times (B^+ - B^- + 1) \times K)$ partially structured Hankel-type tensor $\boldsymbol{\mathcal{A}}$ of Fig. 2.

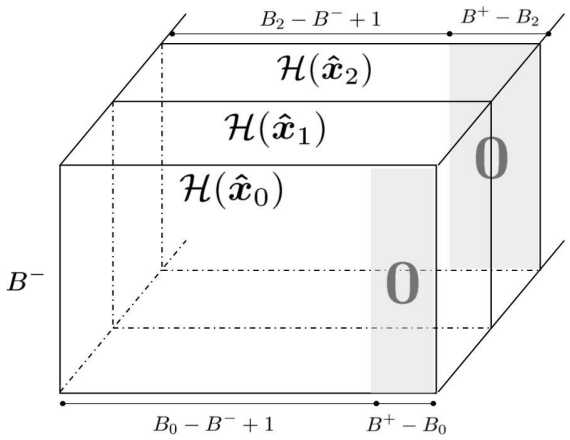


Fig. 2. Partially structured Hankel-type tensor $\boldsymbol{\mathcal{A}}$ on which the data are mapped where we assume that burst 1 has the maximal length (*i.e.*, $B^+ = B_1$).

This third-order tensor can be interpreted as a series of "slabs" indexed by the burst index. The k-th slab is given by

$$[\boldsymbol{\mathcal{A}}]_k = \mathcal{H}(\hat{\boldsymbol{x}}_k)\ \boldsymbol{W}_k. \tag{9}$$

In this equation $\mathcal{H}(\hat{\boldsymbol{x}}_k)$ represents the $(B^- \times (B_k - B^- + 1))$ Hankel matrix associated with the vector $\hat{\boldsymbol{x}}_k$, containing the B_k samples of the k-th data burst, namely

$$\mathcal{H}(\hat{\boldsymbol{x}}_k) = \begin{pmatrix} \hat{x}_k(0) & \hat{x}_k(1) & \dots & \hat{x}_k(B_k - B^-) \\ \hat{x}_k(1) & \hat{x}_k(2) & \dots & \hat{x}_k(B_k - B^- + 1) \\ \vdots & \vdots & & \vdots \\ \hat{x}_k(B^- - 1) & \hat{x}_k(B^-) & \dots & \hat{x}_k(B_k - 1) \end{pmatrix}. \tag{10}$$

Moreover, $\boldsymbol{W}_k$ is the $((B_k - B^- + 1) \times (B^+ - B^- + 1))$ matrix defined by

$$\boldsymbol{W}_k = \begin{pmatrix} 1 & & & 0 & \dots & 0 \\ & \ddots & & \vdots & & \vdots \\ & & 1 & 0 & \dots & 0 \end{pmatrix}. \tag{11}$$

The effect of multiplying $\mathcal{H}(\hat{\boldsymbol{x}}_k)$ by $\boldsymbol{W}_k$ consists of padding zeros to the right, such that the resulting tensor slab $[\boldsymbol{\mathcal{A}}]_k$ is of dimension $(B^- \times (B^+ - B^- + 1))$, $k \in [0 : K-1]$. Note that the column space and the rank of each slab $[\boldsymbol{\mathcal{A}}]_k$ are the same as those of the corresponding burst data matrix $\mathcal{H}(\hat{\boldsymbol{x}}_k)$.

B. A Vandermonde-type decomposition

The following theorem is key to our derivation. The proof is given in the Appendix.

Theorem 1: If all the poles in the PDDS model are distinct and if

$$M_k \leq B_k - B^-, \quad \forall k \in [0:K-1], \tag{12}$$

$$M < B^-, \tag{13}$$

then $\boldsymbol{\mathcal{A}}$ *is a rank-*(M, M', K) *tensor, where*

$$M' \leq \sum_{k=0}^{K-1} (K-k) M_k. \tag{14}$$

Tensor $\boldsymbol{\mathcal{A}}$ *admits a decomposition of the form*

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{C}} \times_1 \boldsymbol{\Theta}, \tag{15}$$

in which

$$\boldsymbol{\Theta} = \left[\boldsymbol{Z}_0^{(B^-)} \quad \boldsymbol{Z}_1^{(B^-)} \quad \dots \quad \boldsymbol{Z}_{K-1}^{(B^-)} \right], \tag{16}$$

where $\boldsymbol{Z}_k^{(P)}$, $k \in [0:K-1]$, *denotes the* $(P \times M_k)$ *Vandermonde matrix defined by*

$$\boldsymbol{Z}_k^{(P)} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ z_{1,k} & z_{2,k} & \dots & z_{M_k,k} \\ \vdots & \vdots & & \vdots \\ z_{1,k}^{P-1} & z_{2,k}^{P-1} & \dots & z_{M_k,k}^{P-1} \end{pmatrix}. \tag{17}$$

Decomposition (15) is visualized in Fig. 3.

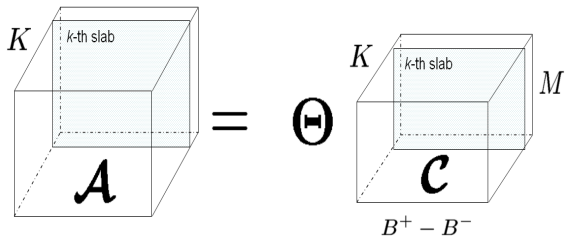


Fig. 3. Vandermonde-type decomposition of tensor $\boldsymbol{\mathcal{A}}$.

Note that (12) implies an upper bound on the value that can be chosen for B^-. Eq. (13), in turn, shows that B^- bounds the number of poles that can be recovered. This constraint is somewhat restrictive

but it is an obvious constraint that applies to all subspace based techniques: one cannot estimate more poles than the dimension of the subspace. If the number of bursts becomes excessively high (in the sense that the total number of poles is too high), then one should start a new analysis. If in some applications extremely short bursts occur (in the sense that (13) becomes too restrictive), then one could simply leave out the corresponding Hankel matrix from the analysis. We emphasize that it is standard in subspace based techniques to assume that the number of poles is less than the dimension of the subspace. This approach is different from the one in [24], [34], where the goal is the determination of the maximum number of poles for a given number of samples.

It is actually easy to see that $B^- - 1$ is in general not the maximum possible number of poles that can be extracted. Like before, we assume that the data are noise-free. Consider the first burst. From standard harmonic analysis, we have that, if $M_0 \leqslant \lfloor \frac{B_0}{2} \rfloor$, the poles in this burst are uniquely identifiable [22]. After computation of these poles, we reconstruct the tail that leaks into the following bursts and subtract it. In this way, we obtain "clean" remaining bursts. We then continue in the same fashion with the second burst, and so on. Hence, a sufficient condition for identifiability is that $M_k \leqslant \lfloor \frac{B_k}{2} \rfloor$, $k \in [0 : K-1]$.

C. Exploiting shift invariance

Matrix $\mathbf{\Theta}$, defined in (16), is Vandermonde. Its shift invariance allows one to estimate all the poles of the PDDS model. Let $\mathbf{\Theta}_\downarrow$ and $\mathbf{\Theta}_\uparrow$ be the two submatrices of $\mathbf{\Theta}$ obtained by deleting the last and the first row, respectively. We then have

$$\mathbf{\Theta}_\downarrow \ \mathbf{\Delta} - \mathbf{\Theta}_\uparrow = \mathbf{0} \tag{18}$$

where $\mathbf{\Delta} = \text{diag}\left\{z_{m,k},\ m \in [1 : M_k],\ k \in [0 : K-1]\right\}$. The matrix $\mathbf{\Delta}$ can be computed from this set of linear equations, provided $\mathbf{\Theta}$ is known.

Actually it is not necessary to know $\mathbf{\Theta}$ itself; knowledge of its column space is sufficient. One can use the standard algorithms that are available from the literature, like ESPRIT [33], HSVD [1], HTLS [38] or the Matrix-Pencil (MP) algorithm [22]. MP works as follows. Let $\boldsymbol{U}$ be a column-wise orthonormal matrix that has the same column space as $\mathbf{\Theta}$. Then we have

$$\mathbf{\Theta} = \boldsymbol{U}\ \boldsymbol{P}, \tag{19}$$

for some nonsingular $(M \times M)$ matrix $\boldsymbol{P}$. Hence,

$$(\boldsymbol{U}\ \boldsymbol{P})_\downarrow \ = \ \boldsymbol{U}_\downarrow\ \boldsymbol{P} \tag{20}$$

$$(\boldsymbol{U}\ \boldsymbol{P})_\uparrow \ = \ \boldsymbol{U}_\uparrow\ \boldsymbol{P}. \tag{21}$$

Combining (18)–(21), we obtain

$$\boldsymbol{P}\ \boldsymbol{\Delta}\ \boldsymbol{P}^{-1} = \boldsymbol{U}_{\downarrow}^{(1)^{\dagger}} \boldsymbol{U}_{\uparrow}^{(1)}, \tag{22}$$

in which $\dagger$ denotes the Moore-Penrose pseudo-inverse. Eq. (22) shows that the poles $\{z_{m,k}\}$ can be found as the eigenvalues of $\boldsymbol{U}_{\downarrow}^{(1)^{\dagger}}\ \boldsymbol{U}_{\uparrow}^{(1)}$.

In the following, we will explain how an orthonormal basis for the column space of $\boldsymbol{\Theta}$ can be calculated. We start from Eq. (15).

V. Computation of the signal subspace

A. Approach based on the HOSVD

Theorem 2 (Third-Order SVD [9], [11], [36]): Every complex $(I_1 \times I_2 \times I_3)$-tensor $\boldsymbol{\mathcal{A}}$ can be written as the product

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{S}} \times_1 \boldsymbol{U}^{(1)} \times_2 \boldsymbol{U}^{(2)} \times_3 \boldsymbol{U}^{(3)}, \tag{23}$$

in which $\boldsymbol{U}^{(j)}$, $j = 1, 2, 3$ is a unitary $(I_j \times I_j)$-matrix and $\boldsymbol{\mathcal{S}}$ is an all-orthogonal and ordered complex $(I_1 \times I_2 \times I_3)$-tensor. All-orthogonality means that the matrices $\boldsymbol{S}_{i_j=\alpha}$, obtained by fixing the j-th index to α, are mutually orthogonal w.r.t. the standard inner product. Ordering means that $\left\|\boldsymbol{S}_{i_j=1}\right\| \geqslant \left\|\boldsymbol{S}_{i_j=2}\right\| \geqslant \left\|\boldsymbol{S}_{i_j=I_j}\right\| \geqslant 0$ for all possible values of j. The Frobenius norms $\left\|\boldsymbol{S}_{i_j=i}\right\|$, symbolized by $\sigma_i^{(j)}$, are the j-mode singular values of $\boldsymbol{\mathcal{A}}$ and the columns of $\boldsymbol{U}^{(j)}$ are j-mode singular vectors.

This decomposition is a generalization of the matrix SVD because diagonality of the matrix containing the singular values, in the matrix case, is a special case of all-orthogonality. Also, the HOSVD of a second-order tensor (matrix) yields the matrix SVD, up to trivial indeterminacies. The matrix of j-mode singular vectors, $\boldsymbol{U}^{(j)}$, can be found as the matrix of left singular vectors of the matrix representation $\boldsymbol{A}_{(j)}$, defined in (5)–(7). The j-mode singular values correspond to the singular values of this matrix representation. Note that the j-mode singular vectors of a tensor, corresponding to the nonzero j-mode singular values, form an orthonormal basis for its j-mode vector subspace as in the matrix case (cf. Section III).

From Eq. (15) it follows that the column space of the data tensor $\boldsymbol{\mathcal{A}}$ is spanned by the columns of the Vandermonde matrix $\boldsymbol{\Theta}$ (cf. Section III). On the other hand, the 1-mode singular vectors of $\boldsymbol{\mathcal{A}}$, corresponding to the nonzero 1-mode singular values, form an orthonormal basis for the same subspace. Hence the signal poles can be determined from the $(B^{-} \times M)$ matrix $\boldsymbol{U}$ in which these 1-mode singular vectors are stacked, as explained in Section IV-C. In the presence of noise, the smallest 1-mode singular values are only approximately equal to zero. The number of poles contributing to the signal is then estimated as the number of significant 1-mode singular values, and the matrix $\boldsymbol{U}$ is obtained by truncating

$\boldsymbol{U}^{(1)}$ after the M-th column. This algorithm is summarized in Table I. From the preceding discussion it is clear that the core of the algorithm (step 3) requires the computation of the dominant M-dimensional subspace of the column space of the $(B^- \times (B^+ - B^- + 1)K)$ matrix $\boldsymbol{A}_{(1)}$; the complexity of this operation is $O(mB^-(B^+ - B^- + 1)K)$ flops with m proportional to M.

1) Stack the data in the $(B^- \times (B^+ - B^- + 1) \times K)$ Hankel-type tensor $\boldsymbol{\mathcal{A}}$ defined by (9), with B^-, B^+ satisfying (12)–(13).
2) Estimate the column rank M of $\boldsymbol{\mathcal{A}}$.
3) Compute the M dominant 1-mode singular vectors of $\boldsymbol{\mathcal{A}}$, i.e., compute the M dominant left singular vectors of $\boldsymbol{A}_{(1)}$, and stack them in a $(B^- \times M)$ matrix $\boldsymbol{U}$.
4) Compute the PDDS signal poles as the eigenvalues of $\boldsymbol{U}_{\downarrow}^{(1)\dagger}\, \boldsymbol{U}_{\uparrow}^{(1)}$ (Section IV-C).

TABLE I

ALGORITHM BASED ON THE TRUNCATED HOSVD.

We emphasize that we first compute (an orthonormal basis for) the column space of $\boldsymbol{\Theta}$ and subsequently compute the signal poles from this subspace. That is, we do not compute decomposition (15) directly. Our approach is thus fundamentally different from fitting to the data tensor a minimal sum of rank-1 tensors (the latter approach is known as "fitting a Canonical Decomposition" or "Parallel Factor Analysis" [6], [14], [21], [35]).

In [2] the signal poles were computed from the left dominant subspace of a matrix, say $\boldsymbol{H}$, in which all the Hankel matrices $\mathcal{H}(\hat{\boldsymbol{x}}_k)$, $k \in [0 : K-1]$, were stacked one after the other. This approach, although derived without using multilinear algebra, is in fact equivalent to the algorithm presented in Table I. The only difference is the presence of zero columns in $\boldsymbol{A}_{(1)}$ (due to the zero padding in (9)), which does not affect the dominant subspace.

Note that, as the HOSVD is computed by means of several matrix SVD's, we can decrease the computational cost of the HOSVD by using fast schemes for the computation of the matrix SVD. We refer to [8], [17] and the references therein.

B. Approach based on the best rank-(R_1, R_2, R_3) approximation

In this section we consider a multilinear generalization of the best rank-R approximation of a given matrix. More precisely, given a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$, we want to find a rank-(R_1, R_2, R_3) tensor

$\hat{\boldsymbol{\mathcal{A}}} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ that minimizes the least-squares cost function

$$f(\hat{\boldsymbol{\mathcal{A}}}) = \|\boldsymbol{\mathcal{A}} - \hat{\boldsymbol{\mathcal{A}}}\|^2. \tag{24}$$

The n-mode rank conditions imply that $\hat{\boldsymbol{\mathcal{A}}}$ can be decomposed as

$$\hat{\boldsymbol{\mathcal{A}}} = \boldsymbol{\mathcal{B}} \times_1 \boldsymbol{U}^{(1)} \times_2 \boldsymbol{U}^{(2)} \times_3 \boldsymbol{U}^{(3)}, \tag{25}$$

in which $\boldsymbol{U}^{(1)} \in \mathbb{C}^{I_1 \times R_1}$, $\boldsymbol{U}^{(2)} \in \mathbb{C}^{I_2 \times R_2}$, $\boldsymbol{U}^{(3)} \in \mathbb{C}^{I_3 \times R_3}$ each have orthonormal columns and $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{R_1 \times R_2 \times R_3}$.

Similarly to the second-order case, where the best approximation of a given matrix $\boldsymbol{A} \in \mathbb{C}^{I_1 \times I_2}$ by a matrix $\hat{\boldsymbol{A}} = \boldsymbol{U}^{(1)} \cdot \boldsymbol{B} \cdot \boldsymbol{U}^{(2)H}$, with $\boldsymbol{U}^{(1)} \in \mathbb{C}^{I_1 \times R}$ and $\boldsymbol{U}^{(2)} \in \mathbb{C}^{I_2 \times R}$ column-wise orthonormal, is equivalent to the maximization of $\|\boldsymbol{U}^{(1)H} \cdot \boldsymbol{A} \cdot \boldsymbol{U}^{(2)}\|$, we have that the minimization of $f(\hat{A})$ is equivalent to the maximization of

$$g(\boldsymbol{U}^{(1)}, \boldsymbol{U}^{(2)}, \boldsymbol{U}^{(3)}) = \|\boldsymbol{\mathcal{A}} \times_1 \boldsymbol{U}^{(1)H} \times_2 \boldsymbol{U}^{(2)H} \times_3 \boldsymbol{U}^{(3)H}\|^2. \tag{26}$$

As explained in [10], [28], the tensor $\boldsymbol{\mathcal{B}}$ in (25) that minimizes (24) for given $\boldsymbol{U}^{(1)}$, $\boldsymbol{U}^{(2)}$, $\boldsymbol{U}^{(3)}$, is given by

$$\boldsymbol{\mathcal{B}} = \boldsymbol{\mathcal{A}} \times_1 \boldsymbol{U}^{(1)H} \times_2 \boldsymbol{U}^{(2)H} \times_3 \boldsymbol{U}^{(3)H}. \tag{27}$$

It is natural to question whether the best rank-(R_1, R_2, R_3) approximation of a third-order tensor can be obtained by truncation of the HOSVD, in analogy with the matrix case. The situation turns out to be quite different for tensors [10], [28]. Discarding the smallest n-mode singular values generally yields a tensor $\hat{\boldsymbol{\mathcal{A}}}$ that is a good but not the best possible approximation under the given n-mode rank constraints. The truncated HOSVD and the best rank-(R_1, R_2, R_3) approximation are generally only equal in the absence of noise. In this section we will estimate the column space of $\boldsymbol{\Theta}$ as the column space of $\boldsymbol{U}^{(1)}$ in (26).

The fact that the truncated HOSVD usually yields a *good* tensor approximation, stems from the ordering constraint in Theorem 2. Namely, this constraint implies that the "energy" of $\boldsymbol{\mathcal{A}}$ is mainly concentrated in the part corresponding to low values of i_1, i_2, i_3. First-order perturbation properties of the HOSVD, describing the possible effect of a noise term, are explained in [9].

On the other hand, in this section we explicitly look for the *optimal* approximating tensor that is rank-(R_1, R_2, R_3). Forcing this structure may indeed improve the signal subspace estimation, as will be confirmed by the simulations in Section VI. First-order perturbation properties of the best rank-(R_1, R_2, R_3) approximation are discussed in [12].

The best rank-(R_1, R_2, R_3) approximation may be obtained by means of tensor generalizations of algorithms for the computation of the dominant subspace of a matrix.

In [10], [28] the following approach was used. Imagine that the matrices $\boldsymbol{U}^{(2)}$ and $\boldsymbol{U}^{(3)}$ are fixed and that the only unknown is the column-wise orthonormal matrix $\boldsymbol{U}^{(1)}$. We have

$$g = \|\tilde{\boldsymbol{\mathcal{A}}}^{(1)} \times_1 \boldsymbol{U}^{(1)H}\|^2 = \|\boldsymbol{U}^{(1)H} \cdot \tilde{\boldsymbol{A}}^{(1)}_{(1)}\|^2, \tag{28}$$

in which $\tilde{\boldsymbol{\mathcal{A}}}^{(1)} = \boldsymbol{\mathcal{A}} \times_2 \boldsymbol{U}^{(2)H} \times_3 \boldsymbol{U}^{(3)H}$. The function g is maximized by a matrix $\boldsymbol{U}^{(1)}$ of which the columns form an orthonormal basis for the left dominant subspace of $\tilde{\boldsymbol{A}}^{(1)}_{(1)}$. An Alternating Least Squares (ALS) algorithm for the (local) minimization of $f(\hat{\boldsymbol{\mathcal{A}}})$ is obtained by iterating over such conditional updates. In each step the estimate of one of the matrices $\boldsymbol{U}^{(1)}, \boldsymbol{U}^{(2)}, \boldsymbol{U}^{(3)}$ is optimized, while the other matrix estimates are kept constant. This algorithm is a tensor generalization of the well-known Orthogonal Iteration method for the computation of the dominant subspace of a matrix [17] and was therefore called the Higher-Order Orthogonal Iteration (HOOI). In our application $\boldsymbol{\mathcal{A}}$ is full mode-3 rank (cf. Theorem 1). Hence, we can take $\boldsymbol{U}^{(3)} = \mathbf{I}$ and alternate only between updates of $\boldsymbol{U}^{(1)}$ and $\boldsymbol{U}^{(2)}$.

We also mention that a Grassmann-Rayleigh quotient based method for the computation of the best rank-(R_1, R_2, R_3) approximation has been derived in [13].

It makes sense to initialize the HOOI (or the Grassmann-Rayleigh quotient based method) with the truncated HOSVD. The HOSVD-estimate usually belongs to the attraction region of the best rank-(R_1, R_2, R_3) approximation, although there is no absolute guarantee of convergence to the global optimum [10]. Other ways to initialize the algorithm may be derived in analogy with [26], [32].

We conclude that it is possible to estimate the column space of the Vandermonde matrix $\boldsymbol{\Theta}$ as the column space of the best rank-(M, M', K) approximation of $\boldsymbol{\mathcal{A}}$; M and M' are the 1-mode and 2-mode rank of $\boldsymbol{\mathcal{A}}$, estimated from the 1-mode and 2-mode singular value spectra. Actually, one can also use the best rank-$(M, \tilde{M}, K)$ approximation, with $\tilde{M} \leq M'$. This approach is suggested by the following theorem.

Theorem 3 ([31]): Let the HOSVD of an $(I_1 \times I_2 \times I_3)$ rank-(R_1, R_2, R_3) tensor $\boldsymbol{\mathcal{A}}$ be given as in (23). Consider a value $\tilde{R}_2 < R_2$ that satisfies $R_1 \leq \tilde{R}_2 R_3$ and $R_3 \leq R_1 \tilde{R}_2$. Then the best rank-$(R_1, \tilde{R}_2, R_3)$ approximation $\tilde{\boldsymbol{\mathcal{A}}}$ of $\boldsymbol{\mathcal{A}}$ is obtained by truncation of the HOSVD components. Formally, let the $(I_1 \times R_1)$ matrix $\tilde{\mathbf{U}}^{(1)}$, the $(I_2 \times \tilde{R}_2)$ matrix $\tilde{\mathbf{U}}^{(2)}$, the $(I_3 \times R_3)$ matrix $\tilde{\mathbf{U}}^{(3)}$ and the $(R_1 \times \tilde{R}_2 \times R_3)$ tensor $\tilde{\boldsymbol{\mathcal{S}}}$ be

defined by

$$\begin{aligned}
(\tilde{\mathbf{U}}^{(1)})_{i_1 r_1} &= (\mathbf{U}^{(1)})_{i_1 r_1} \qquad r_1 \leq R_1 \\
(\tilde{\mathbf{U}}^{(2)})_{i_2 r_2} &= (\mathbf{U}^{(2)})_{i_2 r_2} \qquad r_2 \leq \tilde{R}_2 \\
(\tilde{\mathbf{U}}^{(3)})_{i_3 r_3} &= (\mathbf{U}^{(3)})_{i_3 r_3} \qquad r_3 \leq R_3 \\
\tilde{s}_{i_1 i_2 i_3} &= s_{i_1 i_2 i_3} \quad i_1 \leq R_1, i_2 \leq \tilde{R}_2, i_3 \leq R_3
\end{aligned}$$

Then we have

$$\tilde{\boldsymbol{\mathcal{A}}} = \tilde{\boldsymbol{\mathcal{S}}} \times_1 \tilde{\boldsymbol{U}}^{(1)} \times_2 \tilde{\boldsymbol{U}}^{(2)} \times_3 \tilde{\boldsymbol{U}}^{(3)}.$$

It may actually be advantageous to choose a value for $\tilde{M}$ that is strictly smaller than M' (and satisfies $M \leq \tilde{M}K$). (The other constraint $K \leq \tilde{M}M$ is automatically satisfied since $K \leq M$.) The reason is the following. The 2-mode vectors of $\boldsymbol{\mathcal{A}}$ consist of linear combinations of the columns of the matrix $\tilde{\boldsymbol{\Theta}}$ in (41). The submatrix $\boldsymbol{W}_k^T \boldsymbol{Z}_l^{(B_k - B^- + 1)}$ takes into account the contribution of the poles, introduced in burst l, to the signal in burst $k \geq l$. However, for burst $k > l$, the poles in burst l may already have damped out, so that leaving $\boldsymbol{W}_k^T \boldsymbol{Z}_l^{(B_k - B^- + 1)}$ out of $\tilde{\boldsymbol{\Theta}}$ will only yield a small error. This means that, from a practical point of view, the 2-mode vectors of $\boldsymbol{\mathcal{A}}$ rather consist of linear combinations of $M' - M_l$ columns of $\tilde{\boldsymbol{\Theta}}$. In other words, the 2-mode vector space of $\boldsymbol{\mathcal{A}}$ is ill-conditioned. The same happens when B_k and B_l are close or equal. If they are equal, then the submatrices $\boldsymbol{W}_k^T \boldsymbol{Z}_l^{(B_k - B^- + 1)}$ and $\boldsymbol{W}_l^T \boldsymbol{Z}_l^{(B_k - B^- + 1)}$ are the same, which decreases again the effective 2-mode rank of $\boldsymbol{\mathcal{A}}$. Ill-conditioning is visible in the n-mode singular value spectrum of a tensor. In our application, some of the first M' 2-mode singular values of $\boldsymbol{\mathcal{A}}$ may be close or equal to zero. In that case, the estimation of the signal poles will be more robust if one starts from the best rank-$(M, \tilde{M}, K)$ approximation of $\boldsymbol{\mathcal{A}}$, assuming that $\tilde{M} < M'$ 2-mode singular values of $\boldsymbol{\mathcal{A}}$ are significantly different from zero.

The overall algorithm for the estimation of the signal poles via the HOOI is summarized in Table II. The core of the algorithm (step 4) is an iteration that alternates between (i) the computation of the dominant M-dimensional subspace of the column space of the $(B^- \times \tilde{M}K)$ matrix $\boldsymbol{A}_{(1)} \cdot (\mathbf{I} \otimes \boldsymbol{U}_k^{(2)})$ (the complexity of this operation is $O(mB^-\tilde{M}K)$ flops, with m proportional to M, or less, since we can start from the previous estimate), (ii) the computation of the dominant $\tilde{M}$-dimensional subspace of the column space of the $((B^+ - B^- + 1) \times KM)$ matrix $\boldsymbol{A}_{(2)} \cdot (\mathbf{I} \otimes \boldsymbol{U}_{k+1}^{(1)})$ (the complexity of this operation is $O(m(B^+ - B^- + 1)KM$ flops, with m proportional to M, or less, since we can start from the previous estimate).

1) Stack the data in the $(B^- \times (B^+ - B^- + 1) \times K)$ Hankel-type tensor $\boldsymbol{\mathcal{A}}$ defined by (9), with B^-, B^+ satisfying (12)–(13).
2) Estimate the column rank M and the row rank M' of $\boldsymbol{\mathcal{A}}$.
3) Truncated HOSVD: Compute the M dominant 1-mode singular vectors of $\boldsymbol{\mathcal{A}}$, i.e., compute the M dominant left singular vectors of $\boldsymbol{A}_{(1)}$, and stack them in a $(B^- \times M)$ matrix $\boldsymbol{U}_0^{(1)}$. Compute the $\tilde{M} \leq M'$, with $M \leq \tilde{M}K$, dominant 2-mode singular vectors of $\boldsymbol{\mathcal{A}}$, i.e., compute the $\tilde{M}$ dominant left singular vectors of $\boldsymbol{A}_{(2)}$, and stack them in a $((B^+ - B^- + 1) \times \tilde{M})$ matrix $\boldsymbol{U}_0^{(2)}$.
4) HOOI: Iterate until convergence:
   - Stack the M dominant left singular vectors of $\boldsymbol{A}_{(1)} \cdot (\mathbf{I} \otimes \boldsymbol{U}_k^{(2)})$ in $\boldsymbol{U}_{k+1}^{(1)}$.
   - Stack the $\tilde{M}$ dominant left singular vectors of $\boldsymbol{A}_{(2)} \cdot (\mathbf{I} \otimes \boldsymbol{U}_{k+1}^{(1)})$ in $\boldsymbol{U}_{k+1}^{(2)}$.

   $\boldsymbol{U}^{(1)} = \boldsymbol{U}_{k_{\max}}^{(1)}$.
5) Compute the PDDS signal poles as the eigenvalues of $\boldsymbol{U}_{\downarrow}^{(1)\dagger}\, \boldsymbol{U}_{\uparrow}^{(1)}$ (Section IV-C).

TABLE II

ALGORITHM BASED ON THE COMPUTATION OF THE BEST RANK-$(M, \tilde{M}, K)$ APPROXIMATION VIA THE HOOI.

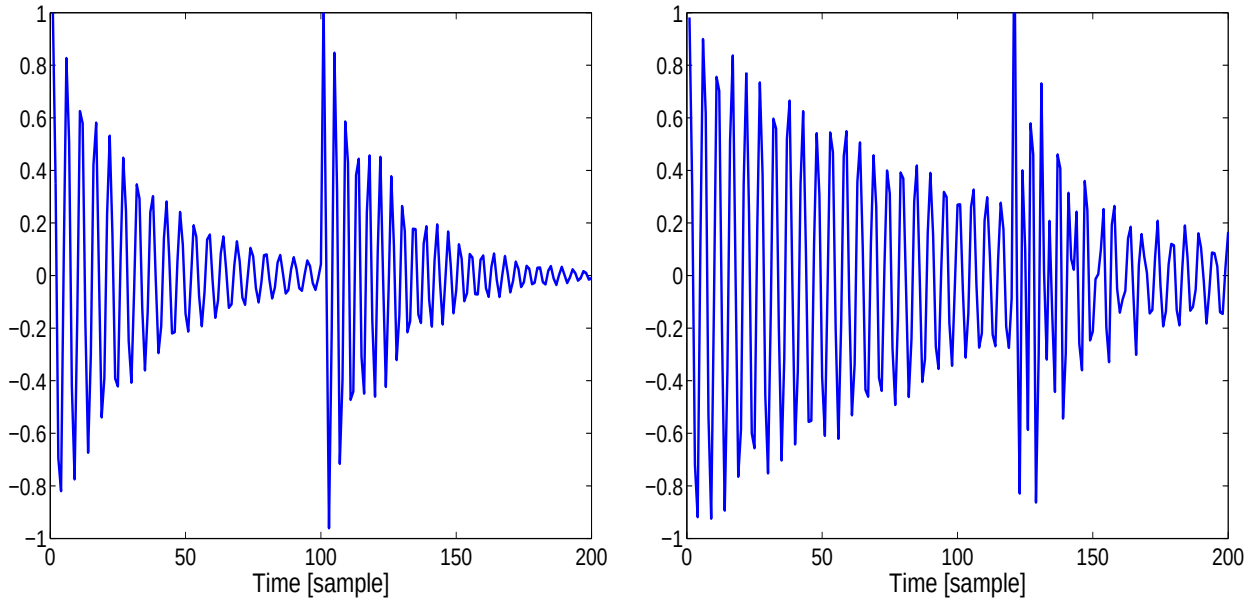
VI. SIMULATIONS

In this section, we test the proposed methods on noisy data. We consider two scenarios. In Fig. 4 (a) there is almost no overlap between the two bursts (quasi-orthogonal time sequences). In Fig. 4 (b) there is a considerable overlap between the two bursts (non-orthogonal time sequences).

The quasi-orthogonal case is simple since we can assume that the influence on burst k of the previous bursts $k-1, k-2, \ldots$ is negligible, *i.e.*, the problem can be decoupled into smaller parts. As a consequence, we focus in this simulation section on the non-orthogonal case. So, for all the simulations, the first exponential in the first burst has a very small damping factor so that the total signal is non-orthogonal.

We compare the following four methods:

- MA-Seq is a Sequential Matrix-based Approach inspired by references [2], [7]. (Contrary to the sequential method proposed in [2], the implementation of the proposed method avoids the estimation of the complex amplitudes.) In this method the poles are not jointly estimated but burst after burst. We work as follows:
  1) The poles in the first burst are estimated by applying a shift-invariant method to the $(B_0/2 \times B_0/2)$ Hankel matrix $\tilde{\mathcal{H}}(\hat{\boldsymbol{x}}_0)$ [7]. Next, projector $\boldsymbol{P}_1^{\perp} = \boldsymbol{I} - \boldsymbol{Z}_0^{(B_1/2)} \boldsymbol{Z}_0^{(B_1/2)\dagger}$ is computed from the Vandermonde matrix $\boldsymbol{Z}_0^{(B_1/2)}$ that contains the estimates of the poles for the first burst.



(a) (b)

Fig. 4. Test signals: (a) quasi-orthogonal case, (b) non-orthogonal case.

2) The $(B_1/2 \times B_1/2)$ Hankel matrix $\tilde{\mathcal{H}}(\hat{\boldsymbol{x}}_1)$ corresponding to the second burst is calculated. The poles belonging to the first burst are cancelled by the multiplication $\boldsymbol{P}_1^{\perp}\tilde{\mathcal{H}}(\hat{\boldsymbol{x}}_1)$. Since left-multiplication of the Hankel matrix by the projector destroys the shift invariance of the basis of left singular vectors, we estimate the new poles in the second burst by applying a shift-invariant technique to the basis of right singular vectors of $\boldsymbol{P}_1^{\perp}\tilde{\mathcal{H}}(\hat{\boldsymbol{x}}_1)$ [7].

- MA-SVD is a Matrix-based Approach based on the SVD of matrix $\sum_k [\boldsymbol{\mathcal{A}}]_k$. Summation of the matrices corresponding to the different bursts, and computing the SVD of the resulting matrix, is a classical approach for harmonic analysis of multi-burst data [29]. We showed in [3] that this approach also applies to PDDS data. The complexity of this technique is low and depends only on B^- and B^+ (and not directly on the number of samples N).
- TA-HOSVD is the Tensor-based Approach using the truncated HOSVD, presented in Table I.
- TA-HOOI, is the Tensor Approach based on the best rank-(M, M, K) approximation (Section V-B), computed by means of the HOOI algorithm as in Table II.

Our performance criterion is the Mean Square Error (MSE), on a logarithmic scale, evaluated for several Signal-to-Noise Ratios (SNRs) and averaged over 500 trials. The MSE is defined by the mean ratio of the squared difference between the true parameter value and its estimate. In appendix VIII-B,

we have derived a Conditional Cramér-Rao Bound, which is a fundamental limit on the MSE of the estimated parameters. As explained in the appendix, we call this bound *conditional* CRB because we assume that we have the exact knowledge of the time-delay parameters.

A. Two bursts, two sinusoids

Consider the noisy synthetic signal given by $\boldsymbol{s} = \hat{\boldsymbol{s}} + \sigma \boldsymbol{w}$ where $\sigma \boldsymbol{w}$ denotes a zero-mean white Gaussian noise with variance σ^2. Let us begin by a 80-sample 2-PDDS model signal with relatively well separated sinusoids given by

$$\begin{cases} e^{(0.2i-0.001)n} & \text{in the first burst,} \\ e^{(i-0.02)n} & \text{in the second burst with } t_1 = 29. \end{cases} \tag{29}$$

The MSE of the angular frequency and damping factor is reported in Fig. 5. In this simulation, we can say that the three methods that simultaneously estimate the two poles (TA-HOOI, TA-HOSVD and MA-SVD) are equivalent, with a small advantage to the TA-HOOI in the context of the damping factor estimation at low SNR. The sequential MA-Seq method has the lowest accuracy for the first burst.

In Fig. 6, we have plotted the MSE vs. an error of $\pm\delta$ samples on the estimate of time-delay t_1. We consider $\delta \in [0:6]$. The SNR is equal to 10 dB. The TA-HOOI and the TA-HOSVD methods are the most robust to inaccuracies in the estimation of the time-delay parameter, since their MSE curves are quite flat. The matrix-based methods, MA-SVD and MA-Seq, are more sensitive to a small error on the time-delay parameter. We conclude that, evidently, the accuracy of the tensor methods decreases but that this decrease is not dramatic.

Next, we consider closely spaced sinusoids given by

$$\begin{cases} e^{(0.55i-0.01)n} & \text{in the first burst,} \\ e^{(0.7i-0.01)n} & \text{in the second burst with } t_1 = 29. \end{cases} \tag{30}$$

The MSE is plotted in Fig. 7. TA-HOOI shows the best accuracy at low SNRs in this situation. TA-HOOI, TA-HOSVD and MA-Seq are equivalent at high SNR. Finally, MA-SVD has the lowest accuracy in this scenario.

B. Two bursts, three sinusoids

In this part, $\hat{\boldsymbol{s}}$ is a 80-sample 3-PDDS signal with:

$$\begin{cases} e^{(0.1i-0.001)n} & \text{in the first burst,} \\ e^{(0.5i-0.01)n} + e^{(i-0.02)n} & \text{in the second burst with } t_1 = 19. \end{cases} \tag{31}$$

Next, we consider two closely spaced sinusoids given by:

$$\begin{cases} e^{(1.2i-0.005)n} & \text{in the first burst,} \\ e^{(1.3i-0.05)n} + e^{(1.8i-0.05)n} & \text{in the second burst with } t_1 = 39. \end{cases} \tag{32}$$

The results are plotted in Fig. 8 and 9, respectively. For well separated sinusoids, the precision of the methods that simultaneously estimate all poles, is comparable. MA-Seq is less reliable at high SNR. If an angular frequency in the second burst is close to an angular frequency in the first burst, the tensor-based methods perform significantly better than the matrix-based methods. Especially the performance of TA-HOOI at low SNR ($\leq$ 10 dB) is remarkable. MA-Seq yields lower MSE values than MA-SVD, but at a higher computational cost.

C. Three bursts, three sinusoids

Now, $\hat{\boldsymbol{s}}$ is a 100-sample 3-PDDS signal with:

$$\begin{cases} e^{(0.1i-0.001)n} & \text{in the first burst,} \\ e^{(0.8i-0.05)n} & \text{in the second burst with } t_1 = 39, \\ e^{(1.4i-0.05)n} & \text{in the third burst with } t_2 = 79. \end{cases} \tag{33}$$

The MSE is plotted in Fig. 10. TA-HOOI, TA-HOSVD and MA-Seq are equivalent as far as the estimation of the angular frequencies is concerned. However, TA-HOOI is the most accurate method for the estimation of the damping factors at low SNR. For higher SNR, TA-HOOI and TA-HOSVD yield a comparable precision. MA-SVD is again the least reliable method.

D. Conclusion of the simulations

We observe that:

1) MA-SVD has the lowest computational cost but is the least accurate method.
2) MA-Seq is in general less accurate than the tensor-based methods. On the other hand, this method generally allows one to estimate more poles. The computational complexity of MA-Seq is comparable to that of TA-HOSVD.
3) TA-HOSVD is more accurate than MA-SVD at low SNR and is generally comparable to TA-HOOI at high SNR.
4) TA-HOOI has the highest computational cost but is the most reliable method in difficult scenarios (e.g. low SNR, closely spaced sinusoids).

5) The tensor-based methods (TA-HOSVD and TA-HOOI) are relatively robust to a small error on the time-delay parameter.
6) The performance of the methods, in particular of the matrix-based methods, is relatively far from the CRB. A Total Least-Squares (TLS) [39] approach can be considered to enhance the accuracy of the proposed algorithms.

VII. Conclusion

In this paper, we have presented subspace-based methods for the estimation of the poles (angular frequencies and damping factors) of damped and delayed sinusoids that have different time supports. The algorithms use multilinear algebraic tools, applied to a structured data tensor. Fitting a synthetic transient signal showed that the best rank-(R_1, R_2, R_3) approach outperforms the current tensor and matrix methods, especially in difficult scenarios such as low SNR and closely spaced sinusoids.

VIII. Appendix

A. Proof of Theorem 1

We first show that (15) holds and that the 1-mode rank of $\mathcal{A}$ is equal to M. The first matrix representation of $\mathcal{A}$ is given by:

$$\boldsymbol{A}_{(1)} = \begin{bmatrix} \mathcal{H}(\hat{\boldsymbol{x}}_0)\boldsymbol{W}_0 & \mathcal{H}(\hat{\boldsymbol{x}}_1)\boldsymbol{W}_1 & \cdots & \mathcal{H}(\hat{\boldsymbol{x}}_{K-1})\boldsymbol{W}_{K-1} \end{bmatrix}. \tag{34}$$

From (4) we have the following expression for $\mathcal{H}(\hat{\boldsymbol{x}}_k)$:

$$\mathcal{H}(\hat{\boldsymbol{x}}_k) = \sum_{\ell=0}^{k} \boldsymbol{Z}_\ell^{(B^-)} \boldsymbol{\Psi}_\ell \boldsymbol{\Delta}_\ell^{\sum_{u=\ell}^{k-1} B_u} \boldsymbol{Z}_\ell^{(B_k - B^- + 1)^T}, \tag{35}$$

in which $\boldsymbol{\Delta}_k = diag\{z_{1,k}, ..., z_{M_k,k}\}$ and $\boldsymbol{\Psi}_k = diag\{\alpha_{1,k}, ..., \alpha_{M_k,k}\}$.

Substituting (35) in (34), we obtain:

$$\begin{aligned} \boldsymbol{A}_{(1)} &= \begin{bmatrix} \boldsymbol{Z}_0^{(B^-)} & \boldsymbol{Z}_1^{(B^-)} & \cdots & \boldsymbol{Z}_{K-1}^{(B^-)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Psi}_0 & & & \\ & \boldsymbol{\Psi}_1 & & \\ & & \ddots & \\ & & & \boldsymbol{\Psi}_{K-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{E}_{00} & \boldsymbol{E}_{01} & \cdots & \boldsymbol{E}_{0,K-1} \\ \boldsymbol{0} & \boldsymbol{E}_{11} & \cdots & \boldsymbol{E}_{1,K-1} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{E}_{K-1,K-1} \end{bmatrix} && (36) \\ &= \boldsymbol{\Theta}\boldsymbol{\Psi}\boldsymbol{E}_{(1)} && (37) \end{aligned}$$

in which:

$$\boldsymbol{E}_{kk} = \boldsymbol{Z}_k^{(B_k-B^-+1)^T}\boldsymbol{W}_k \qquad 0 \leq k \leq K-1 \tag{38}$$

$$\boldsymbol{E}_{lk} = \boldsymbol{\Delta}_l^{\sum_{u=l}^{k-1} B_u}\boldsymbol{Z}_l^{(B_k-B^-+1)T}\boldsymbol{W}_k \qquad 0 \leq k < l \leq K-1. \tag{39}$$

Matrix $\boldsymbol{E}_{(1)}$ is a block upper triangular matrix, of which the diagonal blocks are full row rank, because of condition (12) and the fact that all signal poles are different [27]. Hence $\boldsymbol{E}_{(1)}$ is full row rank, i.e., rank$(\boldsymbol{E}_{(1)}) = M$. On the other hand, rank$(\boldsymbol{\Theta}) = M$ as well, because of condition (13) and the fact that all poles are different. We conclude that the rank of $\boldsymbol{A}_{(1)}$, and hence the 1-mode rank of $\boldsymbol{\mathcal{A}}$, is equal to M. Let us further interpret $\boldsymbol{\Psi}\boldsymbol{E}_{(1)}$ as the first matrix representation of a $(B^- \times (B^+ - B^- + 1) \times K)$ tensor $\boldsymbol{\mathcal{C}}$. Then (15) is just a tensor representation of (37).

Now we prove that the 2-mode rank of $\boldsymbol{\mathcal{A}}$ is bounded as in (14). The second matrix representation of $\mathcal{A}$ is given by:

$$\boldsymbol{A}_{(2)} = \left[(\mathcal{H}(\hat{\boldsymbol{x}}_0)\boldsymbol{W}_0)^T \quad (\mathcal{H}(\hat{\boldsymbol{x}}_1)\boldsymbol{W}_1)^T \cdots (\mathcal{H}(\hat{\boldsymbol{x}}_{K-1})\boldsymbol{W}_{K-1})^T\right]. \tag{40}$$

From (35) we have that the columns of $\boldsymbol{A}_{(2)}$ are linear combinations of the columns of the $(B^+ - B^- + 1) \times (\sum_{k=0}^{K-1}(K-k)M_k)$ matrix

$$\begin{aligned}\tilde{\boldsymbol{\Theta}} = \Big[& \boldsymbol{W}_0^T\boldsymbol{Z}_0^{(B_0-B^-+1)} \quad \boldsymbol{W}_1^T\boldsymbol{Z}_0^{(B_1-B^-+1)} \quad \boldsymbol{W}_1^T\boldsymbol{Z}_1^{(B_1-B^-+1)} \quad \cdots \\ & \boldsymbol{W}_{K-1}^T\boldsymbol{Z}_0^{(B_{K-1}-B^-+1)} \quad \cdots \quad \boldsymbol{W}_{K-1}^T\boldsymbol{Z}_{K-1}^{(B_{K-1}-B^-+1)} \Big].\end{aligned} \tag{41}$$

We conclude that the rank of $\boldsymbol{A}_{(2)}$, and hence the 2-mode rank of $\mathcal{A}$, is bounded by M'.

Finally, we prove that $\boldsymbol{\mathcal{A}}$ is full 3-mode rank. From (35) we deduce that the column space of $[\boldsymbol{\mathcal{A}}]_k = \mathcal{H}(\hat{\boldsymbol{x}}_k)\ \boldsymbol{W}_k$ has non-vanishing components in the directions of the columns of $\boldsymbol{Z}_k^{(B^-)}$. Since all columns of $\boldsymbol{\Theta}$ are linearly independent, all matrices $[\boldsymbol{\mathcal{A}}]_k$ are linearly independent. These matrices are stacked as rows in the third matrix representation of $\boldsymbol{\mathcal{A}}$. We conclude that the rank of $\boldsymbol{A}_{(3)}$, and hence the 3-mode rank of $\mathcal{A}$, is equal to K. ■

B. Derivation of the Conditional Cramér-Rao Bound (CCRB)

The Cramér-Rao Bound (CRB) is useful as a touchstone against which the efficiency of the considered estimators can be tested. Consider a M-PDDS process corrupted by zero-mean white gaussian noise $\boldsymbol{w}$ according to:

$$\boldsymbol{s} = \sum_{k=0}^{K-1}\sum_{m=1}^{M_k} \boldsymbol{c}_{m,k} + \sigma\boldsymbol{w} \tag{42}$$

where $\boldsymbol{c}_{m,k}$ is a N-sample 1-PDDS waveform defined by expression (1). Let $\boldsymbol{\Upsilon} = [\boldsymbol{\omega}^T \ \boldsymbol{d}^T \ \boldsymbol{a}^T \ \boldsymbol{\phi}^T \ \boldsymbol{t}^T \ \sigma^2]^T$ be the vector of desired parameters composed by:

$$\boldsymbol{\omega} = \left[\boldsymbol{\omega}_0^T \ \ldots \ \boldsymbol{\omega}_{K-1}^T\right]^T \quad \text{where} \quad \boldsymbol{\omega}_k = \left(\omega_{1,k} \ \ldots \ \omega_{M_k,k}\right)^T, \tag{43}$$

$$\boldsymbol{d} = \left[\boldsymbol{d}_0^T \ \ldots \ \boldsymbol{d}_{K-1}^T\right]^T \quad \text{where} \quad \boldsymbol{d}_k = \left(d_{1,k} \ \ldots \ d_{M_k,k}\right)^T, \tag{44}$$

$$\boldsymbol{a} = \left[\boldsymbol{a}_0^T \ \ldots \ \boldsymbol{a}_{K-1}^T\right]^T \quad \text{where} \quad \boldsymbol{a}_k = \left(a_{1,k} \ \ldots \ a_{M_k,k}\right)^T, \tag{45}$$

$$\boldsymbol{\phi} = \left[\boldsymbol{\phi}_0^T \ \ldots \ \boldsymbol{\phi}_{K-1}^T\right]^T \quad \text{where} \quad \boldsymbol{\phi}_k = \left(\phi_{1,k} \ \ldots \ \phi_{M_k,k}\right)^T, \tag{46}$$

$$\boldsymbol{t} = \left(t_0 \ \ldots \ t_{K-1}\right)^T. \tag{47}$$

The Cramér-Rao bound, which is given by the diagonal terms of the Fisher Information Matrix (FIM) inverse [40], is a lower bound on the variance of the model parameters, *i.e.*,

$$MSE(\boldsymbol{\Upsilon}) \geq CRB(\boldsymbol{\Upsilon}) = \boldsymbol{F}_{\boldsymbol{\Upsilon}}^{-1} \tag{48}$$

where $\boldsymbol{F}_{\boldsymbol{\Upsilon}}$ denotes the FIM for parameter $\boldsymbol{\Upsilon}$.

We follow the methodology introduced in [4] where the authors define the Conditional CRB (CCRB). As the time-delay has discrete value and is considered as perfectly known, this parameter will be omitted in the CCRB. In addition, it has been shown in [4] that the CCRB for $[\boldsymbol{\omega}^T \ \boldsymbol{d}^T \ \boldsymbol{a}^T \ \boldsymbol{\phi}^T]^T$ is decoupled from the CCRB for σ^2, so we can also omit the noise variance in the computation of the CCRB. Consequently, we retain only vector $\boldsymbol{\Theta} = [\boldsymbol{\omega}^T \ \boldsymbol{d}^T \ \boldsymbol{a}^T \ \boldsymbol{\phi}^T]^T$ to derive the CCRB. Its definition is given according to:

$$CCRB(\boldsymbol{\Theta}) = \boldsymbol{F}_{\boldsymbol{\Theta}}^{-1} \tag{49}$$

where:

$$[\boldsymbol{F}_{\boldsymbol{\Theta}}]_{i,j} = E\left[\frac{\partial \mathcal{L}_t(\boldsymbol{s}|\boldsymbol{\Theta})}{\partial \Theta_i} \cdot \frac{\partial \mathcal{L}_t(\boldsymbol{s}|\boldsymbol{\Theta})}{\partial \Theta_j}\right] \tag{50}$$

where $\mathcal{L}_t(\boldsymbol{s}|\boldsymbol{\Theta})$ is the logarithmic conditional likelihood function.

Theorem 4: The CCRB for the variance of any unbiased estimate of $\boldsymbol{\Theta}$ (conditionally to the perfect knowledge of the time-delay parameter vector $\boldsymbol{t}$) is given by:

$$CCRB(\boldsymbol{\Theta}) = \frac{\sigma^2}{2}\Re e\left\{\boldsymbol{P}_{(\boldsymbol{\Theta})}^H \boldsymbol{P}_{(\boldsymbol{\Theta})}\right\}^{-1} \tag{51}$$

where:

$$\boldsymbol{P}_{(\boldsymbol{\Theta})} = \begin{bmatrix} \boldsymbol{D}^{(\omega)} & \boldsymbol{D}^{(d)} & \boldsymbol{D}^{(a)} & \boldsymbol{D}^{(\phi)} \end{bmatrix}_{N \times 4M} \tag{52}$$

with:

$$\boldsymbol{D}^{(\omega)} = i\boldsymbol{N} \bullet \boldsymbol{S}, \tag{53}$$

$$\boldsymbol{D}^{(d)} = \boldsymbol{N} \bullet \boldsymbol{S}, \tag{54}$$

$$\boldsymbol{D}^{(a)} = \boldsymbol{S}\boldsymbol{A}^{-1}, \tag{55}$$

$$\boldsymbol{D}^{(\phi)} = i\boldsymbol{S} \tag{56}$$

where $\bullet$ denotes the Hadamard product and:

$$\boldsymbol{N} = \begin{bmatrix} \boldsymbol{n}_0 & \dots & \boldsymbol{n}_{K-1} \end{bmatrix} \text{ where } \boldsymbol{n}_k = \begin{pmatrix} -t_k & \dots & N - t_k - 1 \end{pmatrix}^T, \tag{57}$$

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_0 & \dots & \boldsymbol{S}_{K-1} \end{bmatrix} \text{ where } \boldsymbol{S}_k = \begin{bmatrix} \boldsymbol{c}_{1,k} & \dots & \boldsymbol{c}_{M_k,k} \end{bmatrix}, \tag{58}$$

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_0 & & \\ & \ddots & \\ & & \boldsymbol{A}_{K-1} \end{bmatrix} \text{ where } \boldsymbol{A}_k = \text{diag}\{a_{1,k}, \dots, a_{M_k,k}\}. \tag{59}$$

REFERENCES

[1] H. Barkhuijsen, R. de Beer and D. van Ormondt, "Improved Algorithm for Noniterative Time-Domain Model Fitting to Exponentially Damped Magnetic Resonance Signals", *Journal of Magnetic Resonance*, Vol. 73, 1987, pp.553–557.

[2] R. Boyer and K. Abed-Meraim, "Audio Modeling based on Delayed Sinusoids", *IEEE Trans. on Speech and Audio Processing*, Vol. 12, Issue 2, March 2004, pp.110–120.

[3] R. Boyer and K. Abed-Meraim, "Structured tensor based-algorithm for delayed exponential fitting", *Proc. of Asilomar Conference on Signals, Systems and Computers*, Nov. 2004.

[4] R. Boyer and K. Abed-Meraim, "Damped and Delayed Sinuosidal Model for Transient Modeling", *IEEE Trans. on Signal Processing*, Vol. 53, Issue 5, May 2005, pp.1720–1730.

[5] R. Boyer, L. De Lathauwer and K. Abed-Meraim, "Delayed Exponential Fitting by Best Tensor Rank-(R_1, R_2, R_3) Approximation", *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2005)*, Philadelphia, PA, USA, March 19–23, 2005.

[6] J. Carroll and J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition", *Psychometrika*, VOl. 9, 1970, pp.267–283.

[7] H. Chen, S. Van Huffel, D. van Ormondt and R. de Beer, "Parameter Estimation with Prior Knowledge of Known Signal Poles for the Quantification of NMR Spectroscopy Data in the Time Domain", *Journal of Magnetic Resonance*, Series A, Vol. 119, Issue 2, April 1996, pp.225-234.

[8] P. Comon and G.H. Golub, "Tracking a few extreme singular values and vectors in signal processing", *Proceedings of the IEEE*, Vol. 78, Issue 8, Aug. 1990, pp.1327–1343.

[9] L. De Lathauwer, B. De Moor and J. Vandewalle, "A Multilinear Singular Value Decomposition", *SIAM J. Matrix Anal. Appl.*, Vol. 21, Issue 4, April 2000, pp.1253–1278.

[10] L. De Lathauwer, B. De Moor and J. Vandewalle, "On the Best Rank-1 and Rank-$(R_1, R_2, \ldots, R_N)$ Approximation of Higher-Order Tensors", *SIAM J. Matrix Anal. Appl.*, Vol. 21, Issue 4, April 2000, pp.1324–1342.

[11] L. De Lathauwer and J. Vandewalle, "Dimensionality Reduction in Higher-Order Signal Processing and Rank-$(R_1, R_2, \ldots, R_N)$ Reduction in Multilinear Algebra", *Lin. Alg. Appl.*, Special Issue *Linear Algebra in Signal and Image Processing*, Vol. 391, Nov. 2004, pp.31–55.

[12] L. De Lathauwer, "First-Order Perturbation Analysis of the Best Rank-(R_1, R_2, R_3) Approximation in Multilinear Algebra", *J. Chemometrics*, B. Rayens (Ed.), Special Issue *TRICAP*, Vol. 18, Issue 1, 2004, pp.2–11.

[13] L. De Lathauwer, L. Hoegaerts and J. Vandewalle, "A Grassmann-Rayleigh Quotient Iteration for Dimensionality Reduction in ICA", in: C.G. Puntonet, A. Prieto (Eds.), *Proc. Fifth Int. Conf. on Independent Component Analysis and Blind Signal Separation (ICA 2004)*, Granada, Spain, Sept. 2004, *Lecture Notes in Computer Science* 3195, Springer, Berlin Heidelberg, pp.335–342.

[14] L. De Lathauwer, B. De Moor, and J. Vandewalle, *Computation of the canonical decomposition by means of simultaneous generalized Schur decomposition*, SIAM J. Matrix Anal. Appl., Vol. 26, 2004, pp.295–327.

[15] R. Doraiswami and W. Liu, "Real-time estimation of the parameters of power system small signal oscillations", *IEEE Trans. on Power Systems*, Vol. 8, Issue 1, February 1993, pp.74–83.

[16] D. Etter and S. Stearns, "Adaptive estimation of time delays in sampled data systems", *IEEE Trans. on Signal Processing*, Vol. 29, Issue 3, Jun 1981, pp.582–587.

[17] G.H. Golub, C.F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, Maryland: Johns Hopkins University Press, 1996.

[18] M. Goodwin, *Adaptive Signal Models: Theory, Algorithms, and Audio Applications*, PhD thesis, University of California, Berkeley, 1997.

[19] M.M. Goodwin, "Multiresolution sinusoidal modeling using adaptive segmentation", *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '98)*, May 1998.

[20] M.M. Goodwin and J. Laroche, "A dynamic programming approach to audio segmentation and speech/music discrimination", *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, May 2004.

[21] R.A. Harshman, "Foundations of the PARAFAC procedure: Model and conditions for an "explanatory" multi-mode factor analysis", *UCLA Working Papers in Phonetics*, 16 (1970), pp.1–84.

[22] Y. Hua and T.K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise", *IEEE Trans. on Acoustic, Speech and Signal Pocessing*, Vol. 38, Issue 5, May 1990, pp.814–824.

[23] A. Jakobsson, A.L. Swindlehurst and P. Stoica, "Subspace-based estimation of time delays and Doppler shifts", *IEEE Trans. on Signal Processing*, Vol. 46, Issue 9, Sept. 1998, pp.2472–2483.

[24] T. Jiang, N.D. Sidiropoulos and J. M.F. ten Berge, "Almost Sure Identifiability of Multidimensional Harmonic Retrieval", *IEEE Trans. on Signal Processing*, Vol. 49, Issue 9, Sep. 2001, pp.1849–1859.

[25] J. Kliewer and A. Mertins, "Audio Subband Coding With Improved Representation Of Transient Signal Segments", *Proc. of Europ. Signal Processing Conf.*, September 1998.

[26] E. Kofidis and P.A. Regalia, "On the best rank-1 approximation of higher-order supersymmetric tensors", *SIAM J. Matrix Anal. Appl.*, Vol. 23, 2001, pp.863–884.

[27] P. Koiran, "A Rank Theorem for Vandermonde Matrices", *Lin. Alg. Appl.*, Vol. 378, 2004, pp.99–107.

[28] P.M Kroonenberg, *Three-Mode Principal Component Analysis*, Leiden: DSWO Press, 1983.

[29] G. Morren, P. Lemmerling and S. Van Huffel, "Decimative Subspace-Based Parameter Estimation Techniques", *Signal Processing*, Vol. 83, 2003, pp.1025–1033.

[30] J.-M. Papy, L. De Lathauwer and S. Van Huffel, "Exponential Data Fitting Using Multilinear Algebra. The Single-Channel and Multichannel Case", *Num. Lin. Alg. Appl.*, Vol. 12, Issue 8, Oct. 2005, pp.809–826.

[31] J.-M. Papy, L. De Lathauwer and S. Van Huffel, "Exponential Data Fitting Using Multilinear Algebra: the Decimative Case." K.U.Leuven, Electrical Eng. Dept. (ESAT)-SCD, Tech. Report No. 04-70, 2006.

[32] P.A. Regalia and E. Kofidis, "The higher-order power method revisited: convergence proofs and effective initialization", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 5, June 2000, pp.2709–2712.

[33] R. Roy and T. Kailath, "Esprit — Estimation of Signal Parameters via Rotational Invariance Techniques", *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. 37, Issue 7, 1989, pp.984–995.

[34] N.D. Sidiropoulos, "Generalizing Caratheodory's Uniqueness of Harmonic Parameterization to N Dimensions", *IEEE Trans. Information Theory*, Vol. 47, Issue 4, May 2001, pp.1687–1690.

[35] A. Smilde, R. Bro, and P. Geladi, *Multi-way Analysis. Applications in the Chemical Sciences*, John Wiley and Sons, Chichester, U.K., 2004.

[36] L.R. Tucker, "The extension of factor analysis to three-dimensional matrices", in: Gulliksen, H. and Frederiksen, N. (Eds.) *Contributions to mathematical psychology*, NY, Holt, Rinehart & Winston, 1964, pp.109–127.

[37] L.R. Tucker, "Some Mathematical Notes on Three-mode Factor Analysis", *Psychometrika*, Vol. 31, 1966, pp.279–311.

[38] S. Van Huffel, "Enhanced Resolution Based on Minimum Variance Estimation and Exponential Data Modeling", *Signal Processing*, Vol. 33, Issue 3, 1993, pp.333–355.

[39] S. Van Huffel, C. Decanniere, H. Chen and P. Van Hecke, "Algorithm for time-domain NMR data fitting based on total least squares", *Journal of Magnetic Resonance A*, Vol. 110, 1994, pp.228–237.

[40] T. Wigren and A. Nehorai, "Asymptotic Cramér-Rao Bounds for Estimation of the Parameters of Damped Sine Waves in Noise", *IEEE Trans. on Signal Processing*, Vol. 39, Issue 4, April 1991, pp.1017–1020.

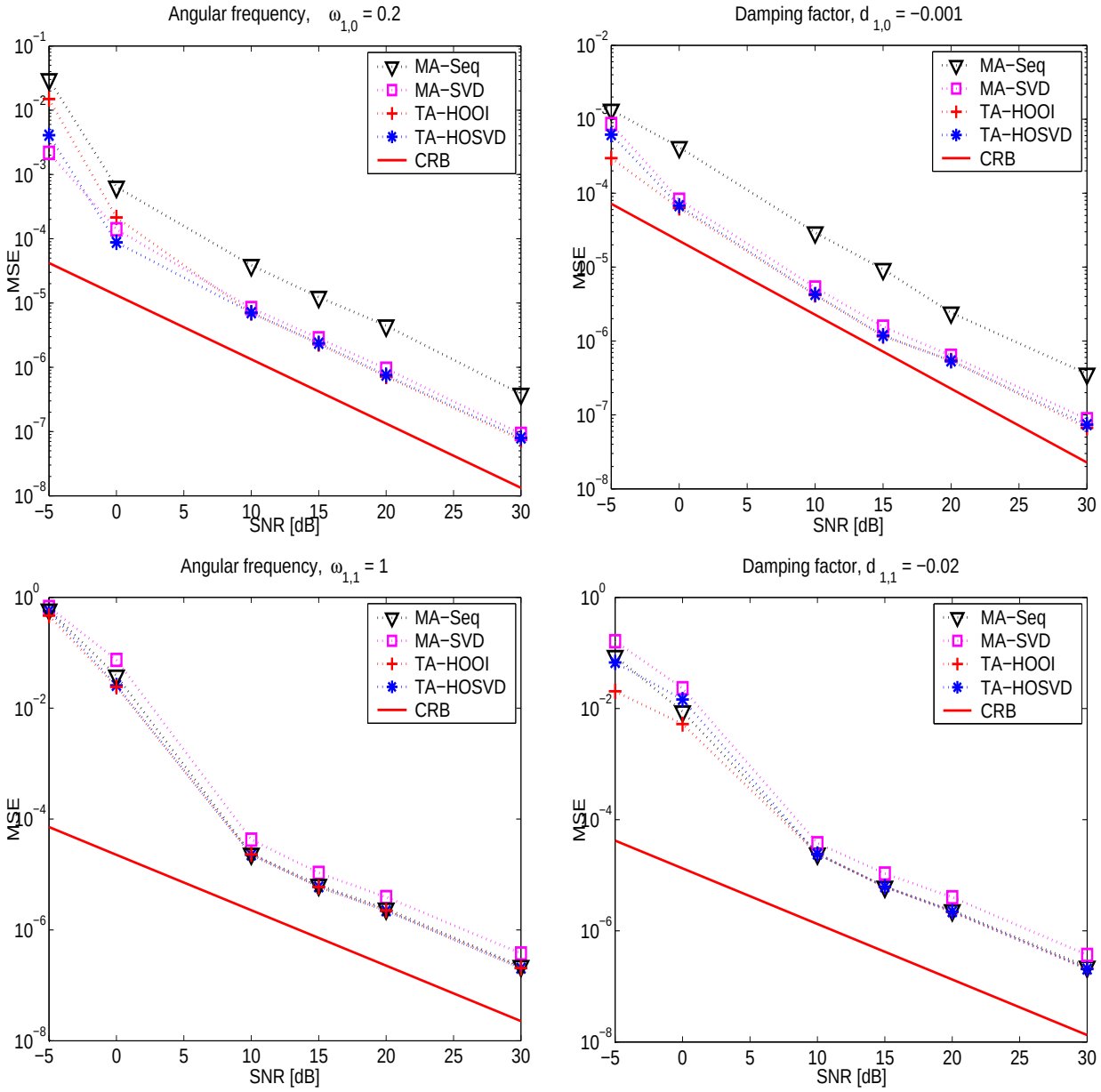


Fig. 5. MSE vs. SNR for two bursts and two sinusoids that are not closely spaced.

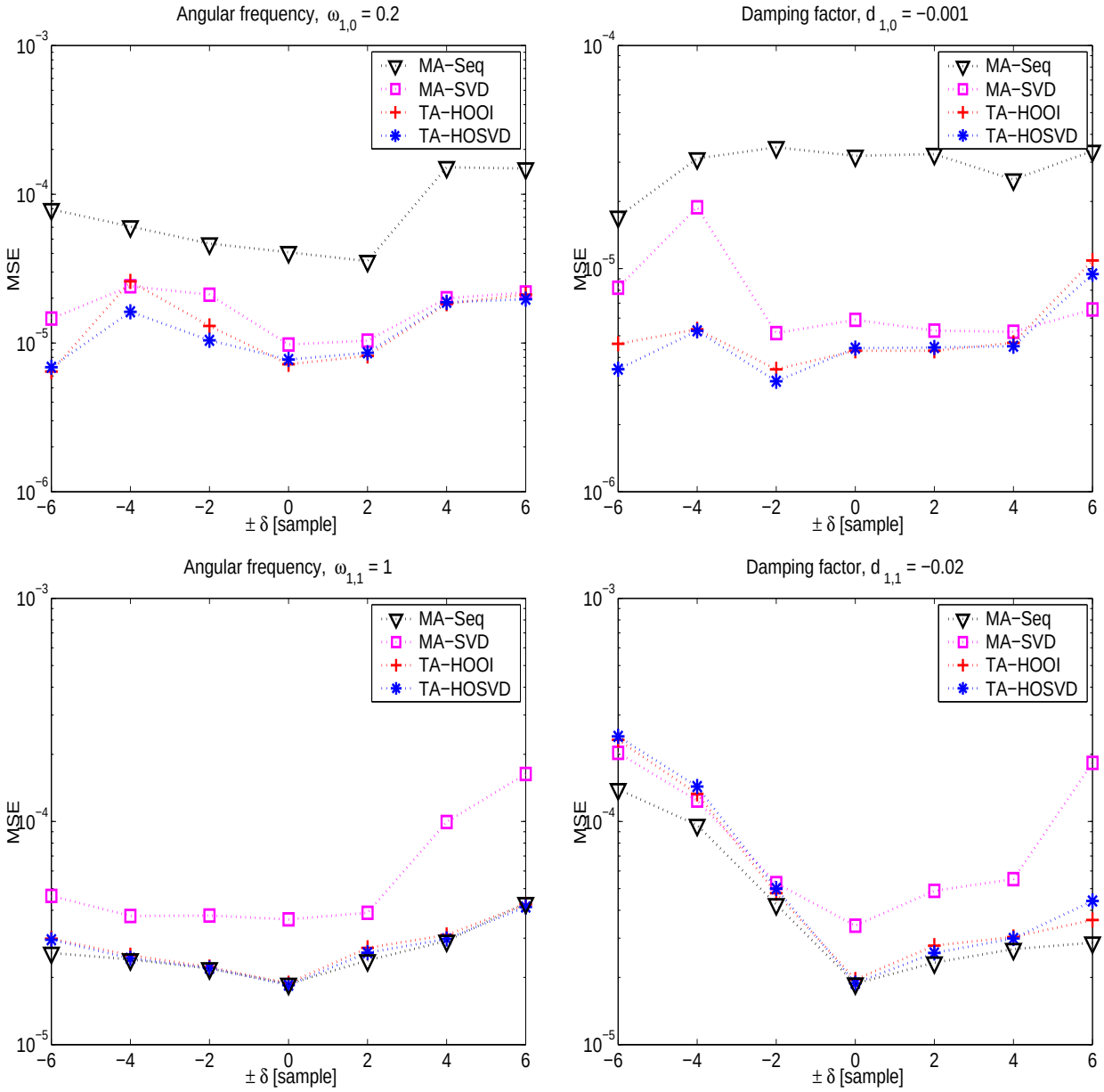


Fig. 6. MSE vs. an error of δ samples in the estimate of t_1. The SNR is 10 dB.

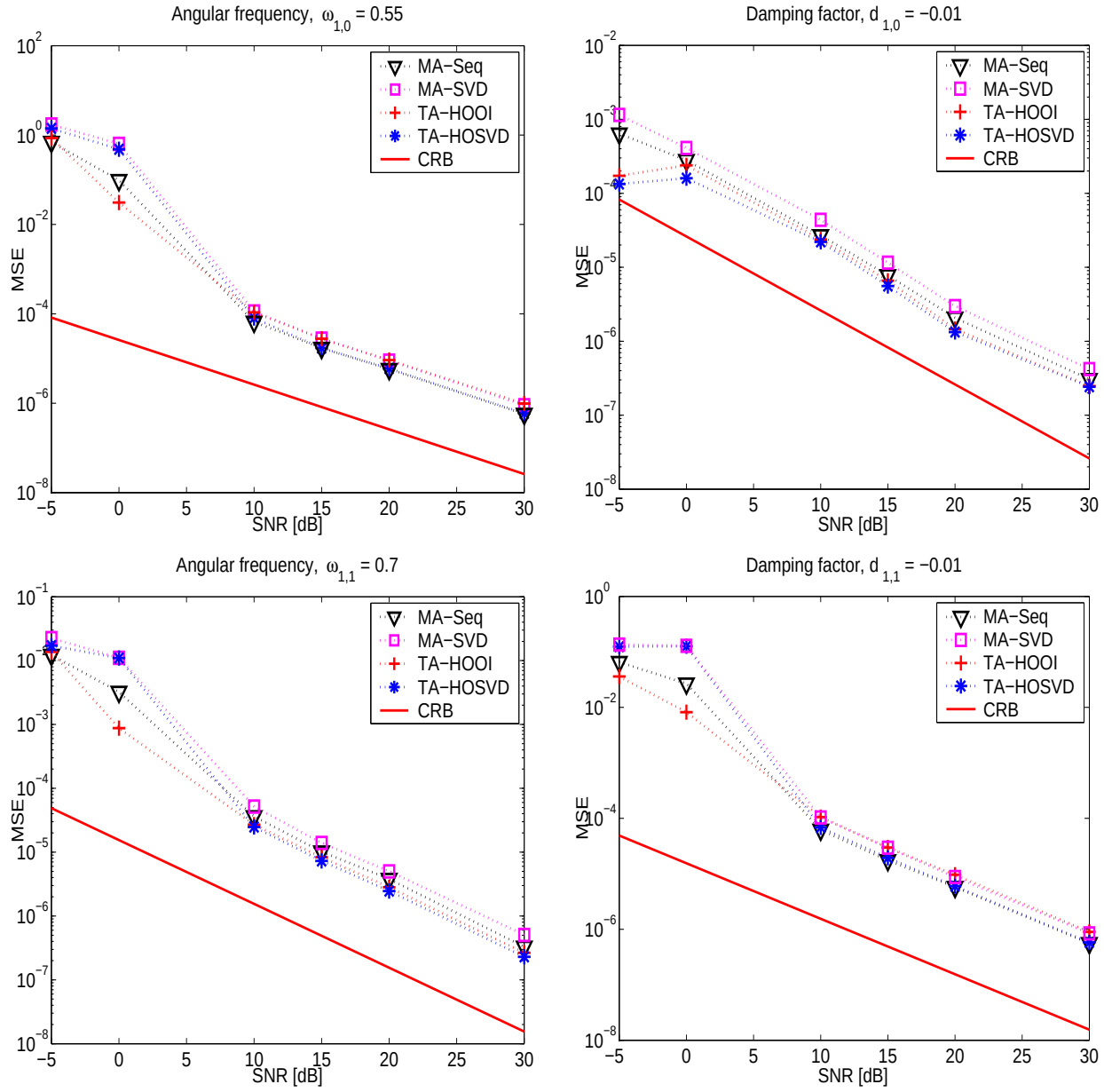


Fig. 7. MSE vs. SNR for two bursts and two closely spaced sinusoids.

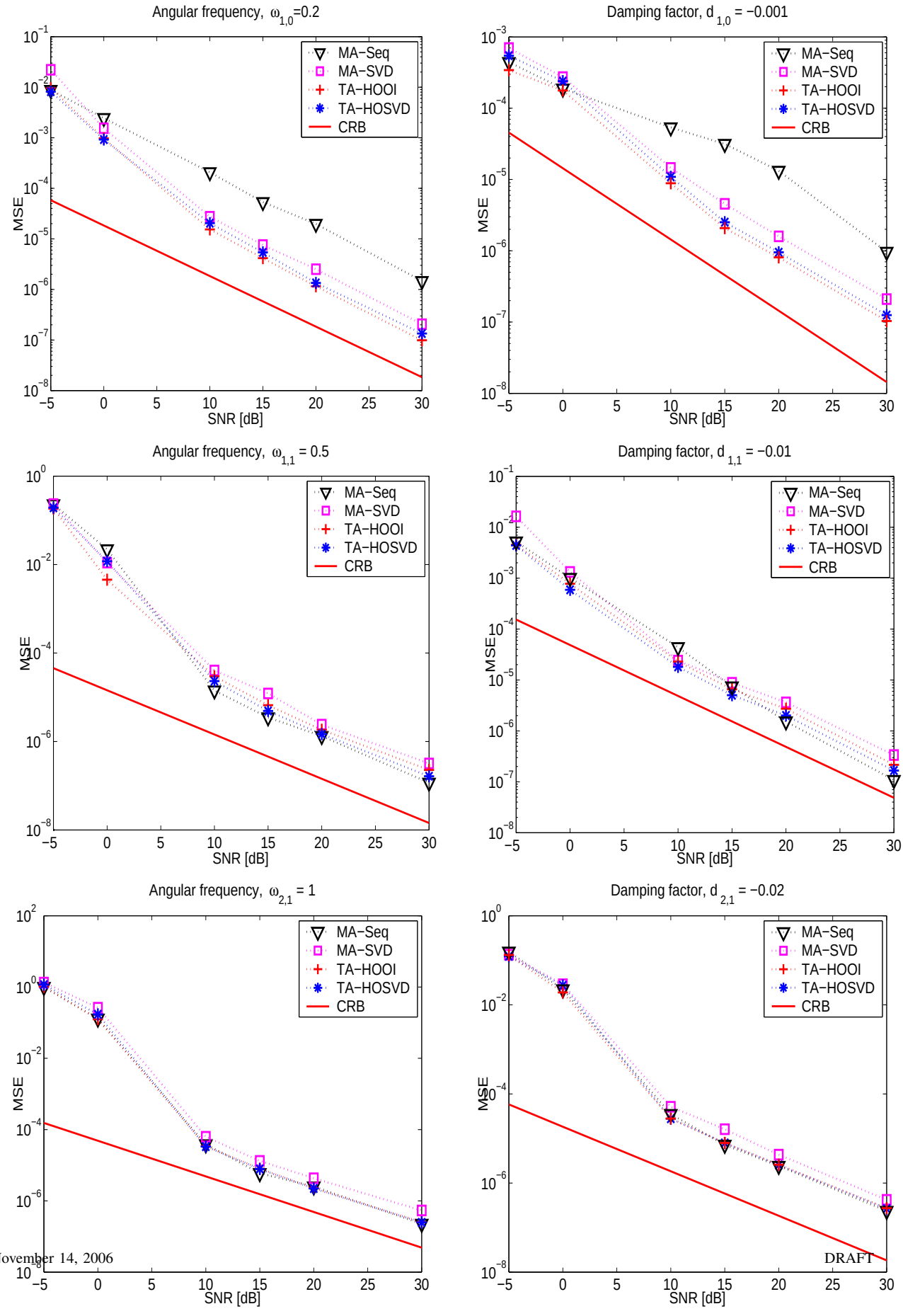


Fig. 8. MSE vs. SNR for two bursts and three well separated sinusoids.

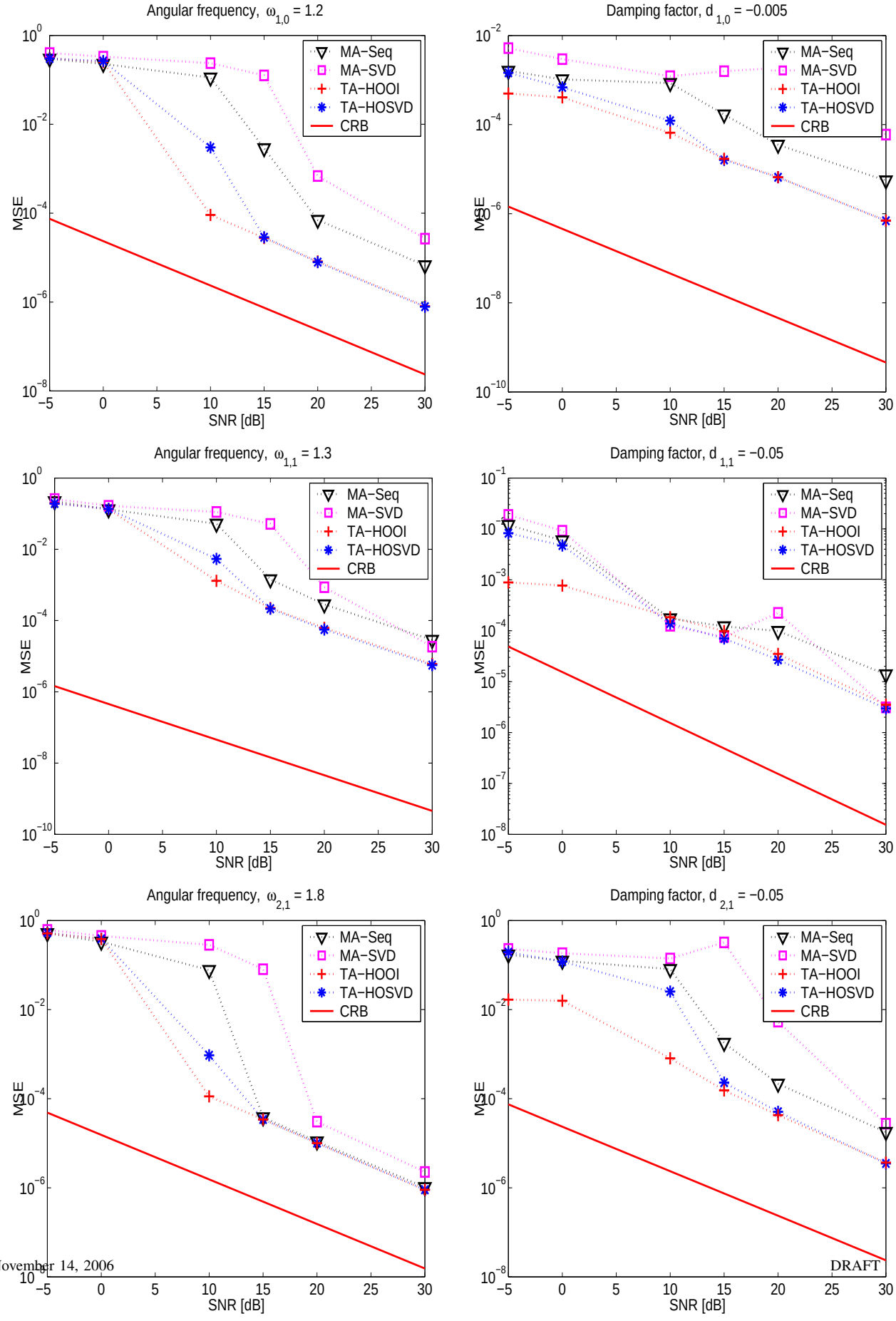


Fig. 9. MSE vs. SNR for two bursts and three sinusoids, where two sinusoids are closely spaced.

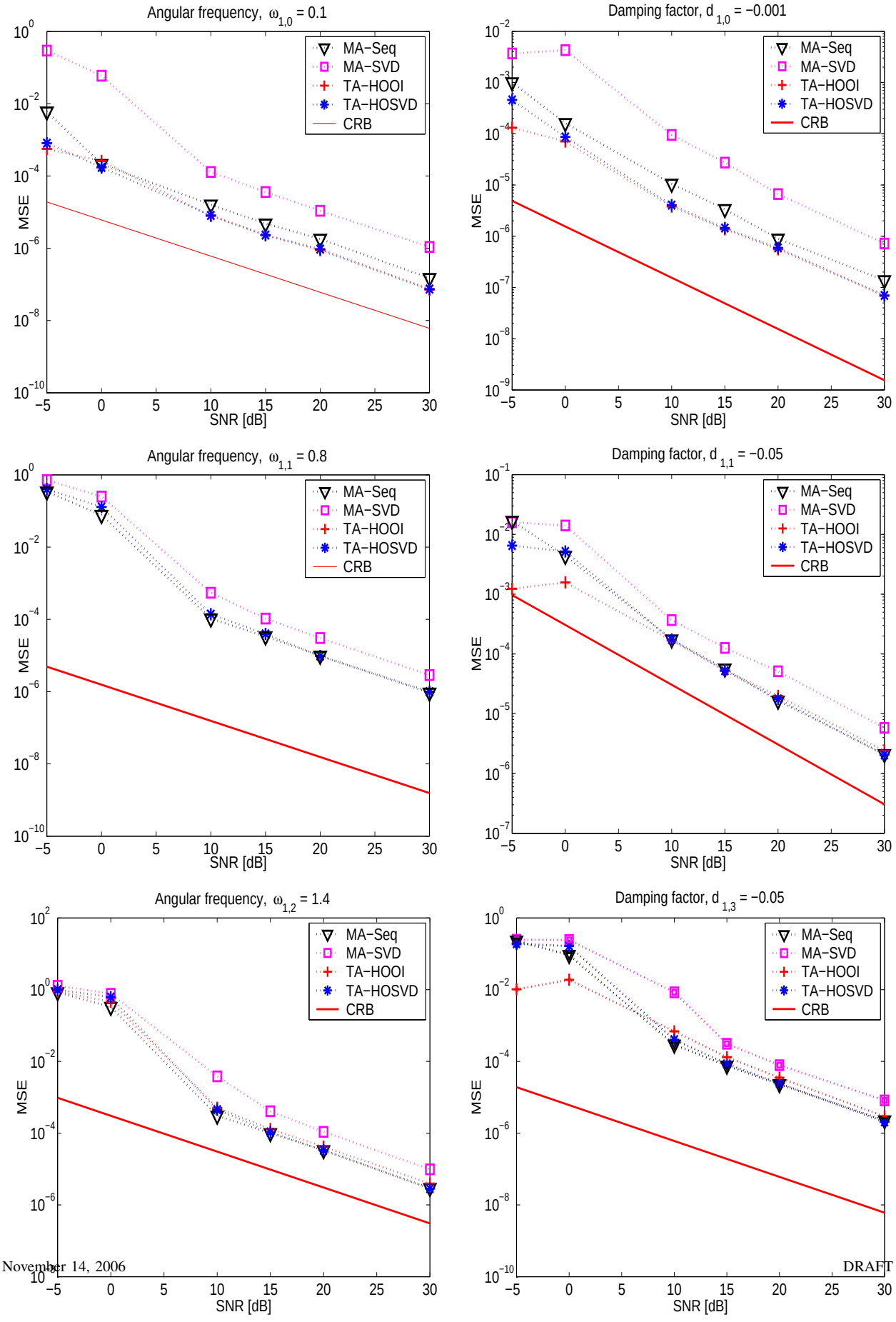


Fig. 10. MSE vs. SNR for three bursts and three sinusoids with three relatively closely spaced sinusoids.