NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2005; **12**:809–826
Published online 13 June 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/nla.453

Exponential data fitting using multilinear algebra: The single-channel and multi-channel case

J. M. Papy[1,*,†], L. De Lathauwer[2,1] and S. Van Huffel[1]

[1]*Katholieke Universiteit Leuven, ESAT, division SCD-SISTA, Kasteelpark Arenberg 10, 3001 Leuven-Heverlee, Belgium*
[2]*Lab. ETIS, UMR 8051 (ENSEA, UCP, CNRS), 6, avenue du Ponceau, BP 44, F 95014 Cergy-Pontoise Cedex, France*

SUMMARY

There is a wide variety of signal processing applications in which the data are assumed to be modelled as a sum of exponentially damped sinusoids. Many subspace-based approaches (such as ESPRIT, matrix pencil, Prony, etc.) aim to estimate the parameters of this model. Typically, the data are arranged in Toeplitz or Hankel matrices and suitable parameter estimates are obtained via a truncated singular value decomposition (SVD) of the data matrix. It is shown that the parameter accuracy may be improved by arranging single-channel or multi-channel data in a higher-order tensor and estimating the model parameters via a multilinear generalization of the SVD. The algorithm is presented and its performance is illustrated by means of simulations. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: signal processing; exponentially damped sinusoids; Toeplitz; Hankel; SVD; higher-order tensor; multilinear algebra

1. INTRODUCTION

Many real-world signals are naturally modelled as a sum of exponentially damped sinusoids. This type of signal lends itself to subspace-based data processing. This is, for instance, the case in nuclear magnetic resonance (NMR) spectroscopy where the quantification of the time-domain signals [1–3] is of great importance for brain tumour detection, or in material health

*Correspondence to: Jean-Michel Papy, Katholieke Universiteit Leuven, ESAT, division SCD-SISTA, Kasteelpark Arenberg 10, 3001 Leuven-Heverlee, Belgium.
† E-mail: michel.papy@esat.kuleuven.ac.be

Contract/grant sponsor: Research Council K.U.Leuven; Contract/grant numbers: GOA-MEFISTO-666, IDO/99/003, IDO/02/009
Contract/grant sponsor: Flemish Government: FWO; Contract/grant numbers: G.0240.99, G.0200.00, G.0078.01, G.0269.02, G.0407.02, G.0270.02
Contract/grant sponsor: Belgian Federal Government; Contract/grant numbers: IUAP IV-02, IUAP V-22
Contract/grant sponsor: EU

Received 6 December 2003
Revised 23 May 2004

monitoring where the model parameters help to characterize the damage occurring in a structure [4].

In this paper we introduce a generalization of the matrix approach for harmonic retrieval. We use data arrays of which the entries are characterized by more than two indices. These are called N-way arrays or higher-order tensors. We first outline the matrix method whose aim is to estimate the poles of a signal that can be modelled by a sum of complex exponentials. The matrix approach is described for single-channel signals in Section 1.1 and for multi-channel signals in Section 1.2. We use the Hankel total least squares (HTLS) method without preprocessing the raw data. This algorithm is a special case of the ESPRIT algorithm [5] and is a TLS variant of Kung *et al.*'s algorithm [6]. Section 2 introduces some tensor-based techniques. First some mathematical background is given in Section 2.1. Section 2.2.1 subsequently describes a tensor variant of the matrix technique for single-channel data. A tensor approach to multi-channel data is discussed in Section 2.2.2. Eventually in Section 3 we compare the matrix and the tensor approach by applying them to a theoretical NMR signal.

Unless stated otherwise, calligraphic letters represent tensors (e.g. $\mathscr{A}$), bold upper-case letters represent matrices (e.g. $\boldsymbol{U}$), upper-case letters represent vectors (e.g. U), and lower-case letters represent elements of tensors, matrices and vectors (e.g. x_n) and other scalars.

1.1. Single-channel signals.

Assume a one-dimensional noiseless time domain signal containing N complex samples x_n, $n=0,1,\ldots,N-1$. Let this time series be modelled as a finite sum of K exponentially damped complex sinusoids

$$x_n = \sum_{k=1}^{K} a_k \exp\{\mathrm{j}\varphi_k\} \exp\{(-\alpha_k + \mathrm{j}\omega_k)t_n\} \tag{1}$$

where $j=\sqrt[2]{-1}$, $t_n = n\Delta t$ is the time lapse between the time origin and the sample x_n and Δt is the sampling time interval. The parameters of this model are the amplitudes a_k, the phases φ_k, the damping factors α_k and the pulsations ω_k. The frequencies ν_k can be obtained from the pulsations by means of the equality : $\omega_k = 2\pi\nu_k$. One can rewrite Equation (1) in a more compact form

$$x_n = \sum_{k=1}^{K} c_k z_k^n \tag{2}$$

where c_k denotes the kth complex amplitude including the phase. $z_k = \exp\{(-\alpha_k + \mathrm{j}\omega_k)\Delta t\}$ is called the kth pole of the signal. We will now briefly repeat a basic matrix technique for the estimation of the poles $\{z_k\}$ from the data $\{x_n\}$ [1–3]. First, the data samples are arranged into an $L \times M$ Hankel matrix as follows:

$$\boldsymbol{H} = \begin{pmatrix} x_0 & x_1 & x_2 & \cdots & x_{M-1} \\ x_1 & x_2 & \iddots & \cdots & \vdots \\ x_2 & \iddots & \iddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & x_{N-2} \\ x_{L-1} & \cdots & \cdots & x_{N-2} & x_{N-1} \end{pmatrix} \tag{3}$$

with $\{L,M\}$ chosen such that $N=L+M-1$ and $\{L>K, M\geqslant K\}$. This Hankel matrix can be decomposed as the following product of three matrices:

$$\boldsymbol{H} = \begin{pmatrix} 1 & \cdots & 1 \\ z_1^1 & \cdots & z_K^1 \\ z_1^2 & \cdots & z_K^2 \\ \vdots & \vdots & \vdots \\ z_1^{L-1} & \cdots & z_K^{L-1} \end{pmatrix} \begin{pmatrix} c_1 & & 0 \\ & \ddots & \\ 0 & & c_K \end{pmatrix} \begin{pmatrix} 1 & z_1^1 & z_1^2 & \cdots & z_1^{M-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & z_K^1 & z_K^2 & \cdots & z_K^{M-1} \end{pmatrix} = \boldsymbol{S}\boldsymbol{C}\boldsymbol{T}^{\mathrm{T}} \quad (4)$$

Equation (4) is called a *Vandermonde decomposition* (VDMD) in which the poles $\{z_k\}$ are also called the *generators*. The superscript T denotes the transpose. The rank deficiency of $\boldsymbol{H}$ is also reflected by the *singular value decomposition* (SVD)

$$\boldsymbol{H} = \begin{pmatrix} \widehat{\boldsymbol{U}} & \boldsymbol{U}_0 \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\Sigma}} & 0 \\ 0 & \boldsymbol{\Sigma}_0 \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{V}}^{\mathrm{H}} \\ \boldsymbol{V}_0^{\mathrm{H}} \end{pmatrix} \quad (5)$$

where $\widehat{\boldsymbol{U}} \in \mathbb{C}^{L\times K}$, $\boldsymbol{U}_0 \in \mathbb{C}^{L\times(L-K)}$, $\widehat{\boldsymbol{\Sigma}} \in \mathbb{R}^{K\times K}$, $\boldsymbol{\Sigma}_0 \in \mathbb{R}^{(L-K)\times(M-K)}$, $\widehat{\boldsymbol{V}} \in \mathbb{C}^{M\times K}$, $\boldsymbol{V}_0 \in \mathbb{C}^{K\times(M-K)}$. In the absence of noise, $\boldsymbol{\Sigma}_0$ is a null matrix and the SVD of $\boldsymbol{H}$ reduces to the product $\widehat{\boldsymbol{U}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{V}}^{\mathrm{H}}$. The columns of $\boldsymbol{S}$ (resp. $\boldsymbol{T}$) span the same subspace as the columns of $\widehat{\boldsymbol{U}}$ (resp. $\widehat{\boldsymbol{V}}^*$). If the signal is corrupted by noise, $\boldsymbol{\Sigma}_0$ is full rank. In this case,

$$\widehat{\boldsymbol{H}}_{L\times M} = \widehat{\boldsymbol{U}}_{L\times K}\ \widehat{\boldsymbol{\Sigma}}_{K\times K}\ \widehat{\boldsymbol{V}}^{\mathrm{H}}_{K\times M} \quad (6)$$

is the best rank-K approximation of $\boldsymbol{H}$. The matrices $\boldsymbol{S}$ and $\boldsymbol{T}$ possess a shift-invariance property that can be expressed as

$$\begin{aligned} \boldsymbol{S}_{\downarrow}\boldsymbol{Z} &= \boldsymbol{S}^{\uparrow} \\ \boldsymbol{T}_{\downarrow}\boldsymbol{Z} &= \boldsymbol{T}^{\uparrow} \end{aligned} \quad (7)$$

In this equation, the up (down) arrow placed behind a matrix stands for deleting the top (bottom) row of the considered matrix and $\boldsymbol{Z} = \mathrm{diag}(z_1, z_2, \ldots, z_K) \in \mathbb{C}^{K\times K}$. In the case of white noise, $\widehat{\boldsymbol{U}}$ equals $\boldsymbol{S}$ up to a multiplication by a square non-singular matrix $\boldsymbol{Q} \in \mathbb{C}^{K\times K}$

$$\widehat{\boldsymbol{U}} = \boldsymbol{S}\boldsymbol{Q} \quad (8)$$

Here, $\widehat{\boldsymbol{U}}^{\uparrow}$ (resp. $\widehat{\boldsymbol{U}}_{\downarrow}$) are related to $\boldsymbol{S}^{\uparrow}$ (resp. $\boldsymbol{S}_{\downarrow}$) in the following way:

$$\begin{aligned} \widehat{\boldsymbol{U}}^{\uparrow} &= \boldsymbol{S}^{\uparrow}\boldsymbol{Q} \\ \widehat{\boldsymbol{U}}_{\downarrow} &= \boldsymbol{S}_{\downarrow}\boldsymbol{Q} \end{aligned} \quad (9)$$

Combining Equations (7) and (9) results in the shift-invariance property of $\widehat{\boldsymbol{U}}$

$$\widehat{\boldsymbol{U}}^{\uparrow} = \widehat{\boldsymbol{U}}_{\downarrow}\bar{\boldsymbol{Z}} \quad (10)$$

with $\tilde{\boldsymbol{Z}} = \boldsymbol{Q}^{-1}\boldsymbol{Z}\boldsymbol{Q}$. The matrices $\boldsymbol{Z}$ and $\tilde{\boldsymbol{Z}}$ have the same eigenvalues. Although the same reasoning applies to the matrix $\widehat{\boldsymbol{V}}^{*}$ [1], we only use $\boldsymbol{U}$ for the parameter estimation. The TLS solution of the overdetermined set of linear equations $\widehat{\boldsymbol{U}}^{\uparrow} \approx \widehat{\boldsymbol{U}}_{\downarrow}\tilde{\boldsymbol{Z}}$ is given by

$$\hat{\tilde{\boldsymbol{Z}}} = -\boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1} \tag{11}$$

and

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{W}_{11} & \boldsymbol{W}_{12} \\ \boldsymbol{W}_{21} & \boldsymbol{W}_{22} \end{bmatrix} \tag{12}$$

is obtained from the SVD of the matrix $[\,\widehat{\boldsymbol{U}}_{\downarrow}\ \ \widehat{\boldsymbol{U}}^{\uparrow}]$

$$[\,\widehat{\boldsymbol{U}}_{\downarrow}\ \ \widehat{\boldsymbol{U}}^{\uparrow}] \overset{\text{SVD}}{=} \boldsymbol{Y}_{(L-1)\times(L-1)}\boldsymbol{\Gamma}\boldsymbol{W}^{\mathrm{H}}_{2K\times 2K} \tag{13}$$

The eigenvalues λ_k of the matrix $\hat{\tilde{\boldsymbol{Z}}}$ give an estimate of the signal poles

$$\lambda_k = \hat{z}_k = \exp\{(-\hat{\alpha}_k + 2\pi j\hat{v}_k)\Delta t\} \tag{14}$$

where $\hat{z}_k$, $\hat{\alpha}_k$ and $\hat{v}_k$ denote the estimates of z_k, α_k and v_k respectively.

1.2. Multi-channel signals

Assume Q simultaneous one-dimensional noiseless time domain signals each consisting of N complex samples. Let this multi-channel signal be modelled as follows:

$$x_n^{(q)} = \sum_{k=1}^{K} a_k^{(q)} \exp\{\mathrm{j}\varphi_k^{(q)}\} \exp\{(-\alpha_k + \mathrm{j}\omega_k)t_n\} = \sum_{k=1}^{K} c_k^{(q)} z_k^n \tag{15}$$

in which $q(1 \leqslant q \leqslant Q)$ denotes the channel number. First, the data samples of each channel are arranged into an $L \times M$ Hankel matrix as follows:

$$\boldsymbol{H}_q = \begin{pmatrix} x_0^{(q)} & x_1^{(q)} & x_2^{(q)} & \cdots & x_{M-1}^{(q)} \\ x_1^{(q)} & x_2^{(q)} & \iddots & \cdots & \vdots \\ x_2^{(q)} & \iddots & \iddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & x_{N-2}^{(q)} \\ x_{L-1}^{(q)} & \cdots & \cdots & x_{N-2}^{(q)} & x_{N-1}^{(q)} \end{pmatrix} \tag{16}$$

The Q Hankel matrices are stacked in one block Hankel matrix

$$\boldsymbol{H} = [\ \boldsymbol{H}_1 \mid \boldsymbol{H}_2 \mid \ \ldots\ \mid \boldsymbol{H}_Q\ ] \tag{17}$$

This matrix can be decomposed as follows:

$$\boldsymbol{H} = \boldsymbol{S}\ [\ \boldsymbol{C}_1\boldsymbol{T}^{\mathrm{T}} \mid \boldsymbol{C}_2\boldsymbol{T}^{\mathrm{T}} \mid \ \ldots\ \mid \boldsymbol{C}_Q\boldsymbol{T}^{\mathrm{T}}\ ] \tag{18}$$

in which $\boldsymbol{S} \in \mathbb{C}^{L\times K}$, $\boldsymbol{T} \in \mathbb{C}^{K\times M}$ are the Vandermonde matrices defined in (4) and $\boldsymbol{C}_q = \mathrm{diag}\{c_1^{(q)}, c_2^{(q)}, \ldots, c_K^{(q)}\}$. Let $\widehat{\boldsymbol{U}}$ be defined by the best rank-K approximation of $\boldsymbol{H}$, in analogy with

(6). Since the columns of $\widehat{\boldsymbol{U}}$ and $\boldsymbol{S}$ span the same subspace, one can repeat the procedure described by Equations (8)–(14) to retrieve the parameters-of-interest.

2. THE TENSOR APPROACH

2.1. Background material

First, we briefly describe the tensor-algebraic material that will be needed in the further derivations. Interested readers are referred to References [7–10] for more details. In analogy with the matrix case, a third-order tensor $\mathscr{A}$ is called rank-1 when it equals the *outer product* of three vectors $U^{(1)}$, $U^{(2)}$, $U^{(3)}$

$$a_{i_1 i_2 i_3} = u^{(1)}_{i_1} u^{(2)}_{i_2} u^{(3)}_{i_3} \tag{19}$$

for all index values, which we will denote as

$$\mathscr{A} = U^{(1)} \circ U^{(2)} \circ U^{(3)} \tag{20}$$

A tensor that can be written as the sum of R, but not less than R, rank-1 terms is called rank-R [11]. An n-mode vector of a tensor $\mathscr{A}$ is a vector obtained from $\mathscr{A}$ by varying only the nth index. As such, the n-mode vectors are the generalization of column and row vectors in the matrix case. The subspace spanned by all the n-mode vectors, for a given value of n, is called the n-mode vector space. The dimension of this subspace is called the n-mode rank. A difference between matrices and tensors is that, for tensors, the different n-mode ranks are not necessarily the same. A third-order tensor whose n-mode ranks are equal to $R_n (1 \leqslant n \leqslant 3)$ is called a rank-(R_1, R_2, R_3) tensor. A second difference between matrices and tensors is that the rank of a tensor is not necessarily equal to its n-mode ranks. From the definition we have that always $R_n \leqslant R$. The inner product of two tensors $\mathscr{A}, \mathscr{B}$ of same dimensionality is given by

$$\langle \mathscr{A}, \mathscr{B} \rangle = \sum_{i_1 i_2 i_3} a_{i_1 i_2 i_3} b^*_{i_1 i_2 i_3} \tag{21}$$

in which the summation is over all the indices. The Frobenius-norm of a tensor is then $\|\mathscr{A}\| = (\langle \mathscr{A}, \mathscr{A} \rangle)^{1/2}$. A tensor–matrix product can be defined as follows [12]. We have

$$\mathscr{A} = \mathscr{S} \times_1 \boldsymbol{V}^{(1)} \times_2 \boldsymbol{V}^{(2)} \times_3 \boldsymbol{V}^{(3)} \tag{22}$$

if and only if

$$a_{i_1 i_2 i_3} = \sum_{j_1 j_2 j_3} s_{j_1 j_2 j_3} v^{(1)}_{j_1 i_1} v^{(2)}_{j_1 i_1} v^{(3)}_{j_1 i_1} \tag{23}$$

for all values of the indices. In References [8, 12], an Nth-order natural extension of the matrix SVD is discussed. For the third-order case, we have

Theorem 2.1 (*Third-order singular value decomposition*)
Every complex $(I_1 \times I_2 \times I_3)$-tensor $\mathscr{A}$ can be written as the product

$$\mathscr{A} = \mathscr{S} \times_1 \boldsymbol{V}^{(1)} \times_2 \boldsymbol{V}^{(2)} \times_3 \boldsymbol{V}^{(3)} \tag{24}$$

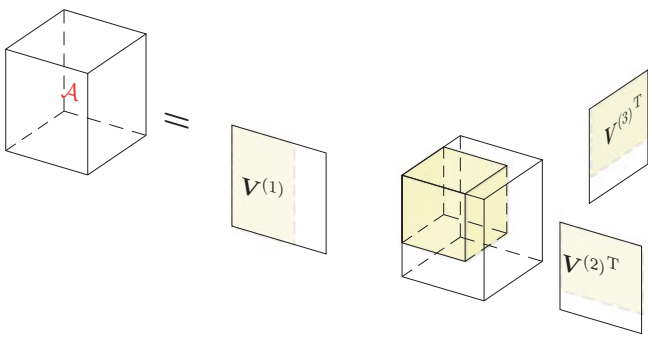

Figure 1. Visualization of the HOSVD of a third-order tensor $\mathcal{A}$. If $\mathcal{A}$ is n-mode rank deficient, only the shaded part of the core-tensor contains entries different from zero.

in which

- $\boldsymbol{V}^{(n)} = \left[V_1^{(n)} V_2^{(n)} \ldots V_{I_n}^{(n)}\right]$ is a unitary $(I_n \times I_n)$-matrix,
- $\mathcal{S}$ is a complex all-orthogonal and ordered $(I_1 \times I_2 \times I_3)$-tensor. This means that the subtensors $\mathcal{S}_{i_n=\alpha}$, obtained by fixing the nth index to α, have the following properties:
  — *All-orthogonality*: Two subtensors $\mathcal{S}_{i_n=\alpha}$ and $\mathcal{S}_{i_n=\beta}$ are orthogonal for all possible values of n, α and β subject to $\alpha \neq \beta$:

$$\langle \mathcal{S}_{i_n=\alpha}, \mathcal{S}_{i_n=\beta} \rangle = 0 \quad \text{when} \quad \alpha \neq \beta \tag{25}$$

  — *Ordering*:

$$\|\mathcal{S}_{i_n=1}\| \geqslant \|\mathcal{S}_{i_n=2}\| \geqslant \cdots \|\mathcal{S}_{i_n=I_n}\| \geqslant 0 \tag{26}$$

  for all possible values of n.

The Frobenius norm $\|\mathcal{S}_{i_n=i}\|$, symbolized by $\sigma_i^{(n)}$, is the ith n-mode singular value of $\mathcal{A}$ and the vector $V_i^{(n)}$ is the n-mode singular vector. The decomposition is visualized for third-order tensors in Figure 1.

It can be proved that it is generally impossible to diagonalize a higher-order tensor by means of unitary transformations: the shaded part of the core tensor in Figure 1 is generally a full $(R_1 \times R_2 \times R_3)$-tensor. Instead, it satisfies the condition of all-orthogonality. All-orthogonality is a suitable multilinear generalization of diagonality because (1) diagonal tensors are also all-orthogonal, (2) all-orthogonality ensures that the HOSVD is always defined and has similar uniqueness properties as the matrix SVD, (3) the HOSVD of a matrix (second-order tensor) reduces, up to some trivial indeterminacies, to its matrix SVD and (4) all-orthogonality is a sufficiently strong condition for many key properties of the matrix SVD to carry over to the HOSVD (namely, properties expressing a link between the distribution of the column/row (n-mode) vectors on one hand and the (HO)SVD-components on the other hand). In particular, the n-mode vectors corresponding to non-zero n-mode singular values form an orthonormal basis for the n-mode vector space of $\mathcal{A}$. The n-mode vectors corresponding to the vanishing n-mode singular values form an orthonormal basis for the orthogonal complement of the

n-mode vector space. The number of (significant) n-mode singular values equals the (numerical) n-mode rank of $\mathscr{A}$. Matrix $\boldsymbol{V}^{(n)}$ $(1 \leqslant n \leqslant 3)$ can easily be computed as the matrix of left singular vectors of a matrix in which all n-mode vectors of $\mathscr{A}$ are stacked one after the other. The n-mode singular values are the singular values of this matrix. The core tensor $\mathscr{S}$ follows from bringing $\boldsymbol{V}^{(1)}$, $\boldsymbol{V}^{(2)}$, $\boldsymbol{V}^{(3)}$ to the left side of Equation (24).

The best rank-R approximation problem in matrix algebra, can be generalized in the following way: *Given a complex third-order tensor* $\mathscr{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$, *find a rank-*$(R_1, R_2, R_3)$ *tensor* $\widehat{\mathscr{A}}$ *that minimizes the least-squares cost function*

$$f(\widehat{\mathscr{A}}) = \|\mathscr{A} - \widehat{\mathscr{A}}\|^2 \tag{27}$$

Due to the n-rank constraints, $\widehat{\mathscr{A}}$ can be decomposed as

$$\widehat{\mathscr{A}} = \mathscr{B} \times_1 \boldsymbol{V}^{(1)} \times_2 \boldsymbol{V}^{(2)} \times_3 \boldsymbol{V}^{(3)} \tag{28}$$

in which $\boldsymbol{V}^{(1)} \in \mathbb{C}^{I_1 \times R_1}$, $\boldsymbol{V}^{(2)} \in \mathbb{C}^{I_2 \times R_2}$, $\boldsymbol{V}^{(3)} \in \mathbb{C}^{I_3 \times R_3}$ each have orthonormal columns and $\mathscr{B} \in \mathbb{C}^{R_1 \times R_2 \times R_3}$ is an all-orthogonal tensor. As is well-known, in the matrix case, the best rank-R approximation can simply be obtained by truncation of the matrix SVD. An important difference between matrices and higher-order tensors however, is that the best rank-(R_1, R_2, R_3) approximation cannot in general be obtained by mere truncation of the HOSVD. This is due to the fact that the core tensor is not diagonal. Nevertheless, due to the ordering constraint (26), the truncated HOSVD is usually a pretty good approximation, albeit not the optimal one. The best rank-(R_1, R_2, R_3) approximation can be calculated by means of tensor generalizations of algorithms for the calculation of a dominant subspace. In Reference [9] we have studied a higher-order orthogonal iteration (HOOI) method. The HOOI algorithm is of the alternating least squares (ALS) type and its convergence rate is linear. Furthermore, we have recently derived a Rayleigh quotient iteration (RQI) algorithm exhibiting quadratic convergence [13].

2.2. Tensor-based algorithm for harmonic retrieval

The concepts introduced in the previous subsection can be used to derive tensor-based algorithms for harmonic retrieval as follows.

2.2.1. Single-channel signals. First, we stack the data in a third-order tensor as follows. A first $(I_1 \times I_2)$-Hankel matrix is built using the $P (P < N)$ first samples of this signal $x_0, x_1, \ldots, x_{P-1}$. A second matrix is built using the segment $x_1, x_2, \ldots, x_P$, obtained by shifting the previous segment over one sample, and so on until the end of the signal. Then these Hankel matrices are stacked one behind the other in a $(I_1 \times I_2 \times I_3)$-Hankel tensor $\mathscr{H}$ as visualized in Figure 2. An element $h_{i_1 i_2 i_3}$ of this tensor is given by

$$h_{i_1 i_2 i_3} = x_{(i_1-1)+(i_2-1)+(i_3-1)} = x_{i_1+i_2+i_3-3} \tag{29}$$

where $1 \leqslant i_1 \leqslant I_1$, $1 \leqslant i_2 \leqslant I_2$, $1 \leqslant i_3 \leqslant I_3$ and $I_1 + I_2 + I_3 = N + 2$. As long as the latter constraint is verified, the dimensions of the tensor may be chosen by the user. In what follows, we assume, $I_1 > K$. Let the one-dimensional complex signal x_n be a K-pole time domain signal modelled by Equation (2). Then we have

$$h_{i_1 i_2 i_3} = \sum_{k=1}^{K} c_k (z_k^{i_1-1} z_k^{i_2-1} z_k^{i_3-1}) \tag{30}$$

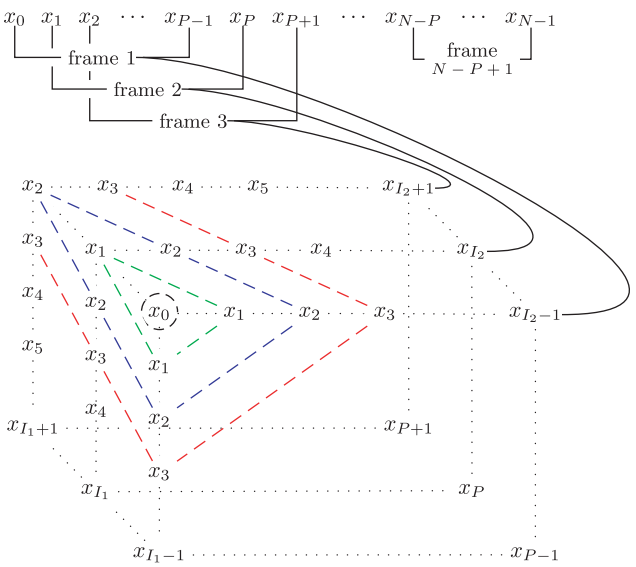

Figure 2. Segmentation of the signal and construction of a tensor with Hankel matrices. The dotted lines delimit the tensor while the dashed lines show his Hankel structure in the three directions. A dashed line of one colour generate a diagonal tensor slice on which all entries are equal.

In other words, the tensor $\mathscr{H}$ is a weighted sum of third-order rank-1 tensors, consisting of the outer product of three Vandermonde vectors

$$\mathscr{H} = \sum_{k=1}^{K} c_k \begin{pmatrix} 1 \\ z_k^1 \\ z_k^2 \\ \vdots \\ z_k^{I_1'} \end{pmatrix} \circ \begin{pmatrix} 1 \\ z_k^1 \\ z_k^2 \\ \vdots \\ z_k^{I_2'} \end{pmatrix} \circ \begin{pmatrix} 1 \\ z_k^1 \\ z_k^2 \\ \vdots \\ z_k^{I_3'} \end{pmatrix} \tag{31}$$

where $I_n' = I_n - 1$. This can also be written as

$$\mathscr{H} = \mathscr{C} \times_1 \boldsymbol{S}^{(1)} \times_2 \boldsymbol{S}^{(2)} \times_3 \boldsymbol{S}^{(3)} \tag{32}$$

in which $\mathscr{C}$ is the pseudo-diagonal $(K \times K \times K)$-core-tensor containing the K complex amplitudes c_k; $\boldsymbol{S}^{(1)} \in \mathbb{C}^{K \times I_1}$, $\boldsymbol{S}^{(2)} \in \mathbb{C}^{K \times I_2}$ and $\boldsymbol{S}^{(3)} \in \mathbb{C}^{K \times I_3}$ are Vandermonde matrices. In analogy with the VDMD (4), the decomposition (32) is called a higher-order vandermonde decomposition (HOVDMD). It is visualized in Figure 3. By way of comparison, note that in the

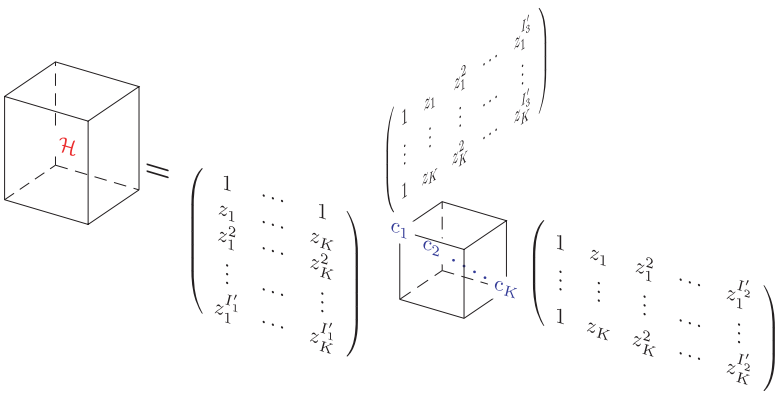

Figure 3. Visualization of the HOVDMD for a third-order Hankel tensor.

matrix case we can write

$$\boldsymbol{H} = \sum_{k=1}^{K} c_k \begin{pmatrix} 1 \\ z_k^1 \\ z_k^2 \\ \vdots \\ z_k^{L'} \end{pmatrix} \circ \begin{pmatrix} 1 \\ z_k^1 \\ z_k^2 \\ \vdots \\ z_k^{M'} \end{pmatrix}$$

$$= \boldsymbol{C} \times_1 \boldsymbol{S} \times_2 \boldsymbol{T} \tag{33}$$

where $M' = M - 1$ and $L' = L - 1$. The three matrices $\boldsymbol{C}$, $\boldsymbol{S}$ and $\boldsymbol{T}$ are defined in (4). From the structure of the HOVDMD (32) it follows that the n-ranks of $\mathscr{H}$ are all equal to the number of signal poles K. A decomposition reflecting the column/row rank deficiency of $\mathscr{H}$ can be used to retrieve the parameters of interest. Note that, the n-mode vector space of $\mathscr{H}$ equals the column space of $\boldsymbol{S}^{(n)}$ $(1 \leqslant n \leqslant 3)$.

The structure induced by (32) implies that, in the absence of noise, the HOSVD of $\mathscr{H}$ takes the following form:

$$\mathscr{H} = \widehat{\mathscr{D}} \times_1 \widehat{\boldsymbol{U}}^{(1)} \times_2 \widehat{\boldsymbol{U}}^{(2)} \times_3 \widehat{\boldsymbol{U}}^{(3)} \tag{34}$$

in which $\widehat{\mathscr{D}}$ is an all-orthogonal, ordered, complex $(K \times K \times K)$-tensor and $\widehat{\boldsymbol{U}}^{(n)} = [U_1^{(n)}, \ldots, U_K^{(n)}]$ is a complex $(I_n \times K)$-matrix whose orthonormal columns span the column space of $\boldsymbol{S}^{(n)}$. In the presence of noise, $\mathscr{H}$ will be a full n-rank tensor. Just like in the matrix case, it makes sense to proceed with the HOSVD-components of the best rank-(R_1, R_2, R_3) approximation of $\mathscr{H}$. Like in the matrix case, we claim that $\widehat{\boldsymbol{U}}^{(n)}$ equals $\boldsymbol{S}^{(n)}$ $(n = 1, 2, 3)$ up to a multiplication by a square non-singular matrix $\boldsymbol{Q} \in \mathbb{C}^{K \times K}$

$$\widehat{\boldsymbol{U}}^{(n)} = \boldsymbol{S}^{(n)} \boldsymbol{Q} \tag{35}$$

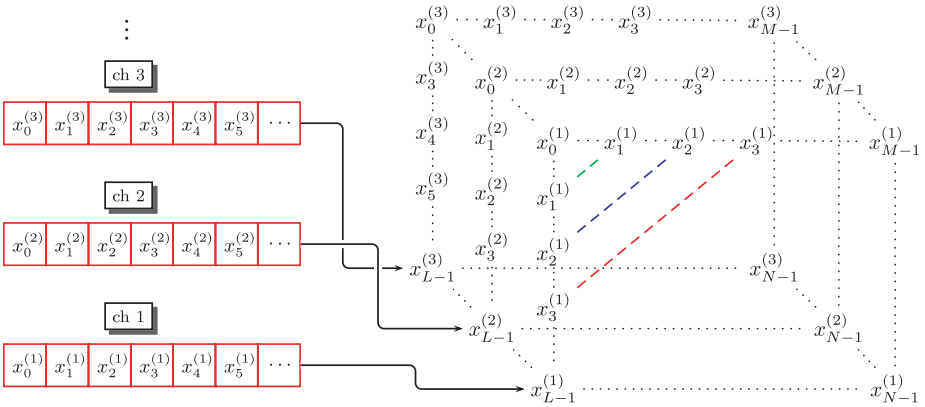

Figure 4. Construction of a tensor from Hankel matrices representing a multi-channel signal. The dotted lines delimit the tensor while the dashed lines show its partial Hankel structure.

On the other hand, the shift-invariance property still holds in the tensor case

$$\boldsymbol{S}_{\downarrow}^{(n)}\boldsymbol{Z}=\boldsymbol{S}^{(n)\uparrow} \tag{36}$$

with $\boldsymbol{Z}=\mathrm{diag}\{z_1,z_2,\ldots,z_K\}$. Combining (35) and (36) yields the following matrix equation:

$$\widehat{\boldsymbol{U}}^{(n)\uparrow}=\widehat{\boldsymbol{U}}_{\downarrow}^{(n)}\ \bar{\boldsymbol{Z}} \tag{37}$$

with $\bar{\boldsymbol{Z}}=\boldsymbol{Q}^{-1}\boldsymbol{Z}\boldsymbol{Q}$. This equation is similar to (10) and can be processed in the same way.

To conclude, we outline a tensor-based algorithm for the estimate of the K signal poles of x_n (1):

- map x_n on a third-order tensor $\mathscr{H}$ as in Figure 2,
- find the best rank-(K,K,K) approximation of $\mathscr{H}$ (for instance, by applying the HOOI algorithm [9, Section 4.1]), and get the matrices $\widehat{\boldsymbol{U}}^{(1)}$, $\widehat{\boldsymbol{U}}^{(2)}$ and $\widehat{\boldsymbol{U}}^{(3)}$,
- compute the TLS solution $\bar{\boldsymbol{Z}}$ of the overdetermined set of linear equations (37),
- compute the eigenvalues of $\bar{\boldsymbol{Z}}$.

2.2.2. Multi-channel signals. We now consider the multi-channel signal (15). In this case, a structure similar to the one in (31) can be obtained in a simple way. We now form a third-order tensor $\mathscr{H}$ by stacking the Q Hankel matrices $\boldsymbol{H}_q$ in (16) one behind the other. This is visualized in Figure 4. An element of $\mathscr{H}$ is given by

$$h_{i_1i_2i_3}=x^{(i_3)}_{(i_1-1)+(i_2-1)}=x^{(i_3)}_{i_1+i_2-2} \tag{38}$$

where $1\leqslant i_1\leqslant L$, $1\leqslant i_2\leqslant M$, $1\leqslant i_3\leqslant Q$, $\min\{L,M\}>K$, and $L+M=N+1$. Expressing each element of $\mathscr{H}$ with respect to the model (15) yield

$$h_{i_1i_2i_3}=\sum_{k=1}^{K}c_k^{(i_3)}\left(z_k^{i_1-1}\,z_k^{i_2-1}\right) \tag{39}$$

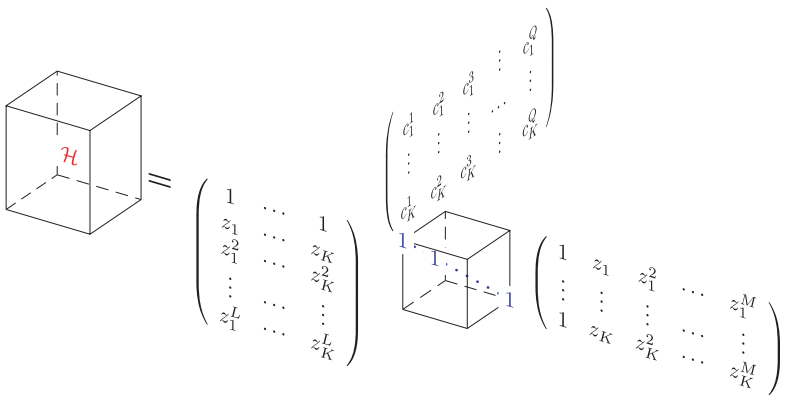

Figure 5. Decomposition of the tensor $\mathscr{H}$ in the multi-channel case.

In the most general case, the coefficients $c_k^{(q)}$ are all different. One can decompose $\mathscr{H}$ as follows:

$$\mathscr{H} = \sum_k \begin{pmatrix} 1 \\ z_k^1 \\ z_k^2 \\ \vdots \\ z_k^L \end{pmatrix} \circ \begin{pmatrix} 1 \\ z_k^1 \\ z_k^2 \\ \vdots \\ z_k^M \end{pmatrix} \circ \begin{pmatrix} c_k^{(1)} \\ c_k^{(2)} \\ c_k^{(3)} \\ \vdots \\ c_k^{(Q)} \end{pmatrix} \tag{40}$$

This might compactly be written as

$$\mathscr{H} = \mathscr{I} \times_1 \boldsymbol{S}^{(1)} \times_2 \boldsymbol{S}^{(2)} \times_3 \boldsymbol{C} \tag{41}$$

in which $\mathscr{I}$ is a pseudo-diagonal $(K \times K \times K)$-tensor with ones on its diagonal (or unit tensor), $\boldsymbol{S}^{(1)} \in \mathbb{C}^{L \times K}$ and $\boldsymbol{S}^{(2)} \in \mathbb{C}^{M \times K}$ are Vandermonde matrices, and $\boldsymbol{C} \in \mathbb{C}^{K \times Q}$ is a matrix containing the complex amplitudes. This decomposition is visualized in Figure 5. The structure induced by (41) implies that, in the absence of noise, the HOSVD of $\mathscr{H}$ takes the following form:

$$\mathscr{H} = \widehat{\mathscr{F}} \times_1 \widehat{\boldsymbol{U}}^{(1)} \times_2 \widehat{\boldsymbol{U}}^{(2)} \times_3 \widehat{\boldsymbol{W}} \tag{42}$$

in which $\widehat{\mathscr{F}}$ is an all-orthogonal, ordered, complex $(K \times K \times K)$-tensor, $\widehat{\boldsymbol{U}}^{(1)} = [U_1^{(1)}, \ldots, U_K^{(1)}]$ is a complex $(L \times K)$-matrix whose orthonormal columns span the column space of $\boldsymbol{S}^{(1)}$, $\widehat{\boldsymbol{U}}^{(2)} = [U_1^{(2)}, \ldots, U_K^{(2)}]$ is a complex $(M \times K)$-matrix whose orthonormal columns span the column space of $\boldsymbol{S}^{(2)}$, and $\widehat{\boldsymbol{W}}$ is a complex $(Q \times K)$-matrix whose orthonormal columns span the column space of $\boldsymbol{C}$. In the presence of noise, as $\mathscr{H}$ becomes n-mode full rank, we can proceed with the HOSVD components of the best rank-(K,K,K) approximation of $\mathscr{H}$. Like in Equation (35), we can claim that $\widehat{\boldsymbol{U}}^{(n)}$ equals $\boldsymbol{S}^{(n)}$ $(n = 1, 2)$ up to a multiplication by a

square non-singular matrix $\boldsymbol{Q}$. The shift-invariance property of $\boldsymbol{S}^{(n)}$ can be exploited in the same way as in (36), (37) and (10).

Note that the algebraic structure reflected by (40) arises very naturally from the structure of the data. If the number of exponentials K is smaller than the number of channels Q, the tensor is mode-3 rank deficient. This property, which is not exploited in the matrix approach, makes the tensor approach more accurate.

3. RESULTS

For one-channel data, the accuracy of the matrix approach is compared in this section to the accuracy of the tensor approach by means of the two peaks example from [1]

$$\begin{aligned} y_n &= x_n + e_n \\ &= \underbrace{\exp\{(-0.01 + 2j\pi 0.2)\, n\}}_{\text{pole 1}} + \underbrace{\exp\{(-0.02 + 2j\pi 0.22)\, n\}}_{\text{pole 2}} + \underbrace{e_n}_{\text{noise}},\ (0 \leqslant n \leqslant N = 24) \end{aligned} \tag{43}$$

in which e_n is a complex circular symmetric white Gaussian noise (WGN). The reason for the choice of this problem is related to the difficulty of its resolution: first, the signal is composed with a few number of points and then the two peaks are very closely spaced. For the matrix approach, we work with the matrix $\widehat{\boldsymbol{U}}$ containing the 2 dominant left singular vectors of the Hankel matrix $\boldsymbol{H}$. For the tensor approach, we work with the matrix $\widehat{\boldsymbol{U}}^{(3)}$ in the HOSVD (28) of the best rank-$(2,2,2)$ approximation of $\mathscr{H}$. We conduct Monte Carlo experiments consisting of $N_{\text{runs}} = 1000$ runs. The performance is expressed in terms of the relative root mean square error (RRMSE)

$$\text{RRMSE} = \left[\frac{1}{N_{\text{runs}}} \sum_{i=1}^{N_{\text{runs}}} |\gamma - \hat{\gamma}_i|^2\right]^{1/2} \frac{100}{\gamma}\ (\%) \tag{44}$$

where $\hat{\gamma}_i$ is the estimate of the parameter γ in the ith run. It is a fact that, when dealing with subspace-based processing, the dimension of the initial data matrix is very important. But to our knowledge, no strong theoretical foundation exists which enable to find easily the optimal matrix size, and this is unfortunately the same problem for the tensor approach. However, we wish to compare the matrix and tensor approach in a consistent way, which means that we must start with the optimal size for both the data matrix $\boldsymbol{H}$ and tensor $\mathscr{H}$. As far as the matrix approach is concerned, a method consists in exploring all the possible matrix sizes and determine for which one the performance of the algorithm is optimal. This idea is used in Reference [2] where it is explained that the HTLS algorithm applied to a $i \times (N + 1 - i)$ Hankel matrix yields very good results when $i \in [N/3, 2N/3]$. This is confirmed by Figure 6. To compare with the tensor approach, we chose the optimal dimensionality (15×11). Following Reference [2], for the tensor approach, all possible values of I_1, I_2 have been tried with I_3 constrained

$$I_3 = N + 2 - I_1 - I_2 = 27 - I_1 - I_2 \tag{45}$$

and the result is shown in Figure 7. The best results were obtained for tensors that consist of a small number of vertical 'slices' $(I_3 = 4, 5, 6)$. For these values the results obtained by

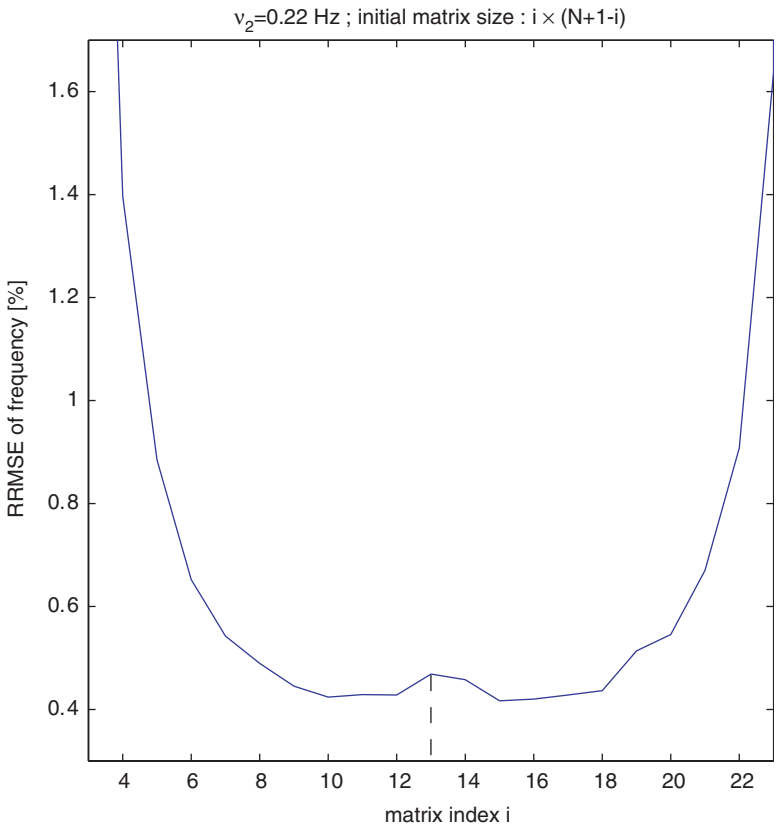

Figure 6. RRMSE of v_2 at SNR = 30 dB obtained for matrices of dimension $i \times (N + 1 - i)$. The index corresponding to the highest accuracy is $i = 15$. The performance curve for the other parameters exhibit exactly the same behaviour.

the tensor approach are consistently more accurate than the best matrix results over a large range of values for I_1 and I_2. Like in the matrix case, the best result is obtained for vertical slices that are slightly rectangular. For other values of I_3, the tensor algorithm may perform much worse than the matrix algorithm. In order to circumvent this weakness, we are currently investigating whether the optimal dimensionality may be predicted by first-order perturbation analysis [14]. Moreover, further studies showed that for well-conditioned problems (peaks well separated), the matrix and the tensor approach gives comparable results. Therefore, we cannot guarantee that the tensor approach always performs better, but it may be useful for difficult problems. At this moment we only want to show that, for certain dimensions, the tensor technique may consistently yield better results than the matrix algorithm. In this example we retained the optimal dimensionality $(14 \times 8 \times 5)$. Matrix and tensor results are compared, for varying signal-to-noise ratio (SNR), in Figure 8. The average improvement in accuracy is given in Table I.

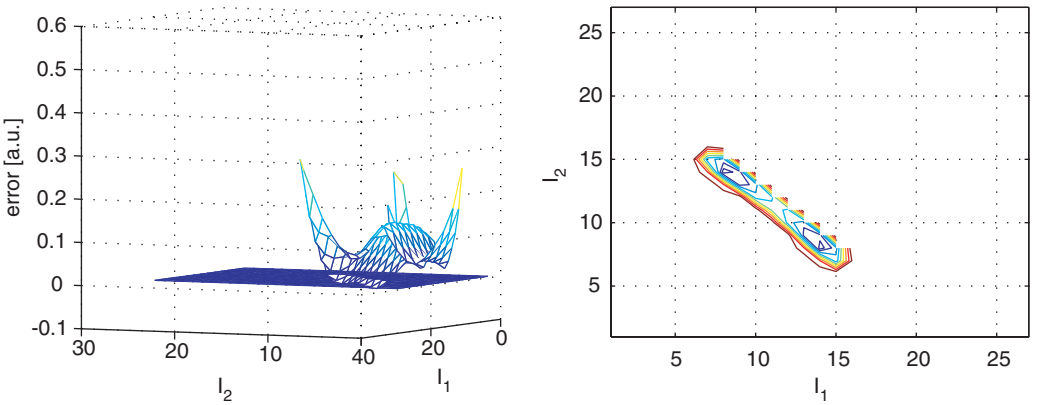

Figure 7. The left plot is the error map of v_2 at SNR = 30 dB. It shows the RRMSE obtained for tensors of dimension $I_1 \times I_2 \times I_3$ from which is subtracted the smallest RRMSE obtained for the matrix approach (see Figure 6). Therefore the indices for which the error is negative (below the plane) correspond to a better performance of the tensor approach. The right contour plot represents the 2D projection of the surface for which the values are negative. The indices corresponding to the highest accuracy are $I_1 = 14$ and $I_2 = 8$. The performance curve for the other parameters exhibit exactly the same behaviour.

For the multi-channel case model (43) has been slightly modified as follows:

$$\begin{aligned} y_n^{(q)} &= x_n^{(q)} + e_n^{(q)} \\ &= \underbrace{c_1^{(q)} \exp\{(-0.01 + 2j\pi 0.2)\, n\}}_{\text{pole 1}} + \underbrace{c_2^{(q)} \exp\{(-0.02 + 2j\pi 0.22)\, n\}}_{\text{pole 2}} + \underbrace{e_n^{(q)}}_{\text{noise}} \\ &\quad (0 \leqslant n \leqslant N = 24),\ (1 \leqslant q \leqslant Q = 12) \end{aligned} \tag{46}$$

in which $e_n^{(q)}$ is a different complex circular symmetric WGN sequence for each channel q. The complex amplitudes $c_k^{(q)}$ are drawn from a Gaussian distribution with zero-mean and unit-variance for $1 \leqslant k \leqslant K = 2$ and $1 \leqslant q \leqslant Q = 12$. For the matrix approach, we work with the matrix $\widehat{\boldsymbol{U}}$ containing the 2 dominant left singular vectors of the Hankel matrix $\boldsymbol{H}$. For the tensor approach, we work with the matrix $\widehat{\boldsymbol{U}}^{(1)}$ in the HOSVD (28) of the best rank-$(2,2,2)$ approximation of $\mathscr{H}$. Note that the philosophy of the multi-channel case is very different from the single-channel case: the data are arranged in a different way but the matrices H_q (the 'slices') are not modified and therefore the dimension of the subspace of interest in unchanged. In this case there exists strong theoretical foundations and for all dimensionalities the tensor algorithm performs better than the matrix algorithm. However, the figures have been drawn for the dimensionalities that lead to the most accurate results for the matrix algorithm. We worked with a $(13 \times 13 * 12)$-matrix $\boldsymbol{H}$ and a $(13 \times 13 \times 12)$-tensor $\mathscr{H}$. Figure 9 shows that the tensor approach consistently yields better results. Moreover the improvement increases as the noise level increases as shown in Figure 10. In the curves related to v_1 and v_2, this trend continues for $\sigma > 0.4$, although α_1 and α_2 can no longer be accurately estimated.

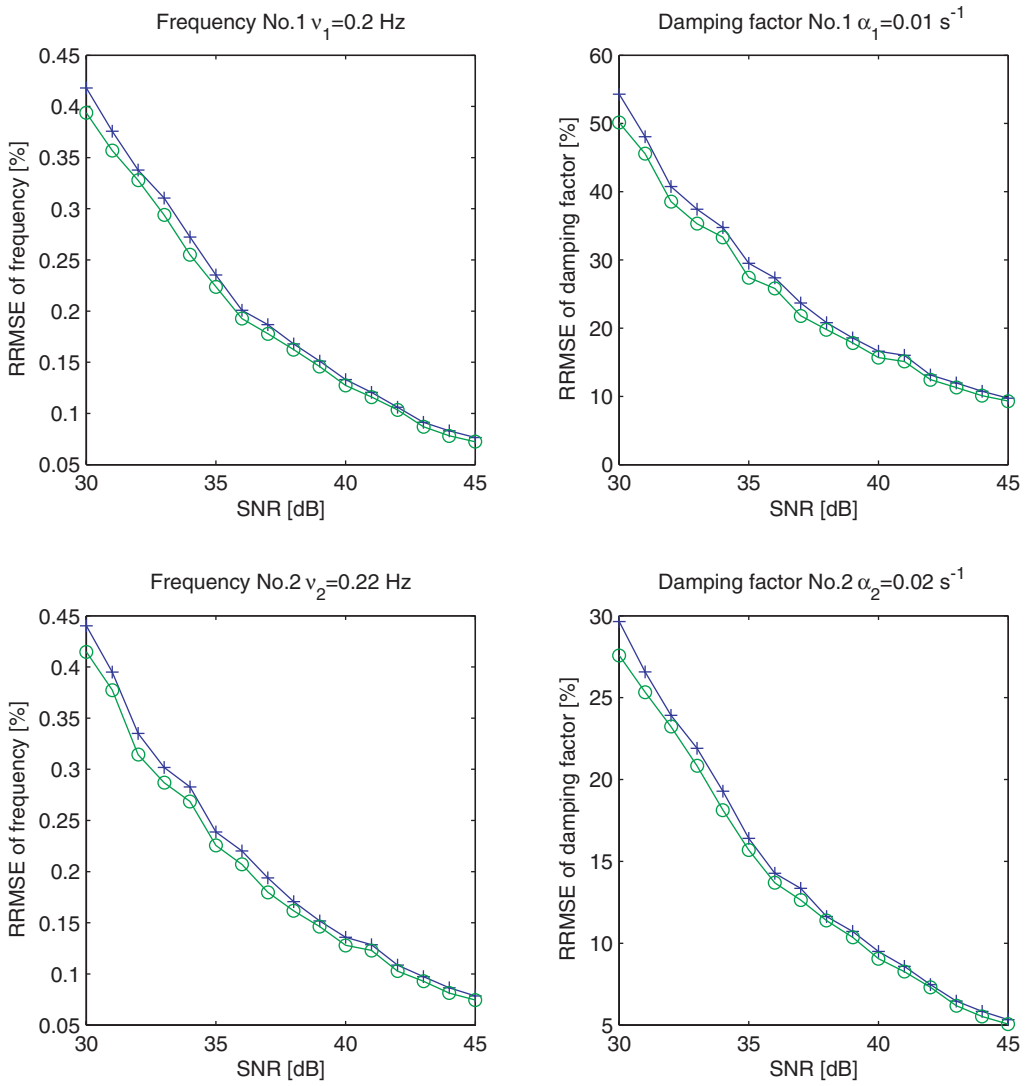

Figure 8. Comparison of the results from the matrix approach (+) and the tensor approach (∘).

Table I. Average improvement of the tensor approach compared to the matrix approach.

	Frequency (%)	Damping factor (%)
Pole 1	4.2	5.5
Pole 2	5	4.1

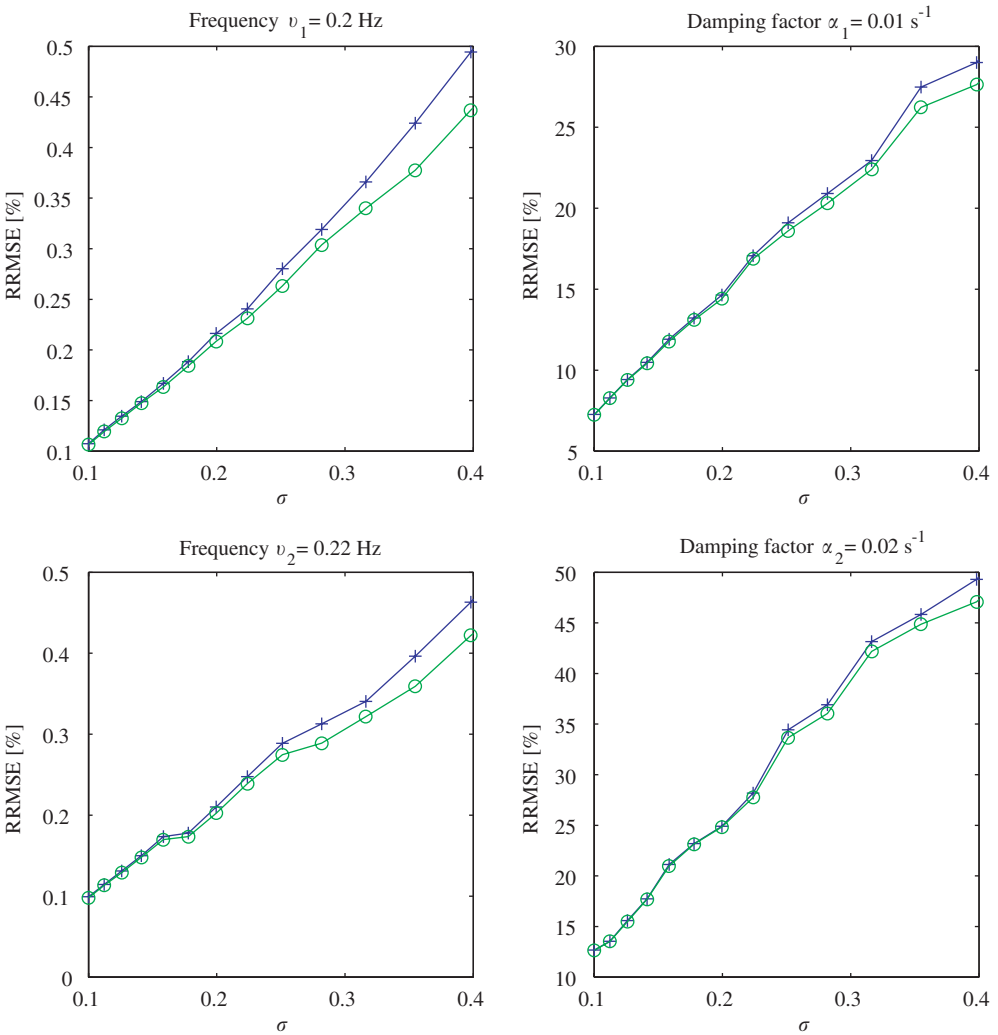

Figure 9. Plot of the RRMSE versus noise standard deviation σ. The matrix approach is displayed with '+' and the tensor is displayed with '$\circ$'.

4. CONCLUSIONS AND FURTHER RESEARCH

In this paper we have introduced the basics of a multilinear algebra-based approach to harmonic retrieval.

The basis of the approach to single-channel data is the fact that the multiplicative (Vandermonde) structure of an harmonic signal allows to represent this signal as a higher-order rank-1 tensor. We have demonstrated by means of an example that the tensor approach may yield more accurate results than its matrix counterpart. A crucial step in the technique is the choice of the tensor dimensions. We are currently investigating whether the optimal dimensionality can be predicted by means of the perturbation analysis conducted in Reference [14].

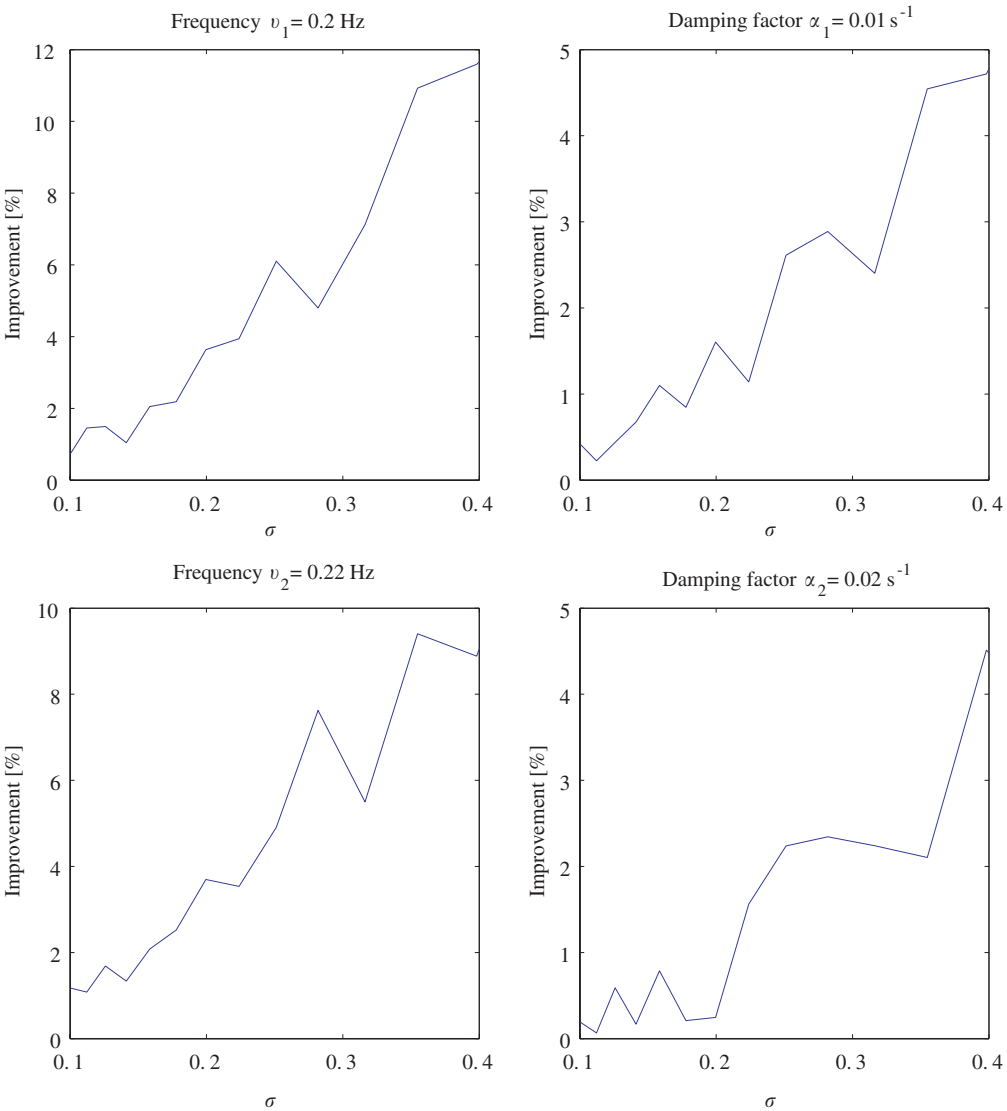

Figure 10. Plot of the improvement factor versus noise standard deviation σ.

For multichannel data it is very natural to work in a multilinear framework. One simply stacks the different Hankel matrices, corresponding to the different channels, in a third-order tensor, instead of putting them one after the other in a big matrix. The tensor algorithm allows to exploit a part of the data structure that is not being used in matrix techniques.

In this paper we have mapped the signals on tensors of order 3. In fact, the approach can be generalized to arbitrary tensor orders. The influence of the order also forms a topic of current research.

The idea presented in this paper can be used to derive a tensor version of the matrix approach to decimation-based harmonic analysis of oversampled data [15]. This result will be presented in a follow-up paper.

ACKNOWLEDGEMENTS

Lieven De Lathauwer holds a permanent research position with the French Centre National de la Recherche Scientifique (C.N.R.S.); he also holds a honorary research position with the K.U.Leuven. Sabine Van Huffel is Full Professor with the K.U.Leuven.

This work is supported by several institutions: (1) Research Council K.U.Leuven: Concerted Research Action GOA-MEFISTO-666, IDO /99/003 and /02/009, several PhD/postdoc & fellow grants, (2) Flemish Government: F.W.O. PhD/postdoc grants, the F.W.O. projects G.0240.99, G.0200.00, G.0078.01, G.0269.02, G.0407.02, G.0270.02 and the F.W.O. Research Communities ICCoS and ANMMM, I.W.T PhD grants, (3) Belgian Federal Government: Interuniversity Poles of Attraction Programmes IUAP IV-02 and IUAP V-22, (4) EU: PDT-COIL, BIOPATTERN and eTUMOUR. The scientific responsibility is assumed by the authors.

REFERENCES

1. Van Huffel S. Enhanced resolution based on minimum variance estimation and exponential data modeling. *Signal Processing* 1993; **33**(3):333–355.
2. Van Huffel S, Chen H, Decanniere C, Van Hecke P. Algorithm for time-domain NMR data fitting based on total least squares. *Journal of Magnetic Resonance* 1994; **110**(Series A):228–237.
3. Chen H, Van Huffel S, Vandewalle J. Bandpass prefiltering for exponential data fitting with known frequency region of interest. *Signal Processing* 1996; **48**:135–154.
4. Rippert L. Optical fiber for damage monitoring in carbon fiber reinforced plastic composite materials. *Ph.D. Thesis*, KU Leuven, 2003.
5. Roy R, Kailath T. ESPRIT—Estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 1989; **37**:984–995.
6. Kung SY, Arun KS, Bhaskar Rao DV. State-space and singular value decomposition-based approximation methods for the harmonic retrieval problem. *Journal of the Optical Society of America* 1983; **73**(12): 1799–1811.
7. De Lathauwer L. Signal processing based on multilinear algebra. *Ph.D. Thesis*, KU Leuven, 1997.
8. De Lathauwer L, De Moor B, Vandewalle J. A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications* 2000; **21**(4):1253–1278.
9. De Lathauwer L, De Moor B, Vandewalle J. On the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensors. *SIAM Journal on Matrix Analysis and Applications* 2000; **21**(4):1324–1342.
10. De Lathauwer L, Vandewalle J. Dimensionality reduction in higher-order signal processing and rank-$(R_1, R_2, \ldots, R_N)$ reduction in multilinear algebra. *Linear Algebra and its Applications* 2004; **391**:31–55.
11. Kruskal JB. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and satatistics. *Linear Algebra and its Applications* 1977; **18**(2):95–138.
12. Tucker LR. Some mathematical notes on three-mode factor analysis. *Psychometrika* 1966; **31**:279–311.
13. De Lathauwer L, Hoegaerts L. A Rayleigh quotient iteration for the computation of the best rank-$(R_1, R_2, \ldots, R_N)$ approximation in multilinear algebra. *Technical Report SCD-SISTA 04-003*.
14. De Lathauwer L. First-order perturbation analysis of the best rank- (R_1, R_2, R_3) approximation in multilinear algebra. *Journal of Chemometrics* 2004; **18**(1):2–11.
15. Morren G, Lemmerling P, Van Huffel S. Decimative subspace-based parameter estimation technique. *Signal Processing* 2003; **83**:1025–1033.

