Stochastic and Deterministic Tensorization for Blind Signal Separation

Otto Debals and Lieven De Lathauwer

Department of Electrical Engineering (ESAT) – STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium; Group Science, Engineering and Technology, KU Leuven Kulak, E. Sabbelaan 53, 8500 Kortrijk Belgium; iMinds Medical IT, KU Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium.
`Otto.Debals@esat.kuleuven.be`, `Lieven.DeLathauwer@kuleuven-kulak.be`

Abstract. Given an instantaneous mixture of some source signals, the blind signal separation (BSS) problem consists of the identification of both the mixing matrix and the original sources. By itself, it is a non-unique matrix factorization problem, while unique solutions can be obtained by imposing additional assumptions such as statistical independence. By mapping the matrix data to a tensor and by using tensor decompositions afterwards, uniqueness is ensured under certain conditions. Tensor decompositions have been studied thoroughly in literature. We discuss the matrix to tensor step and present tensorization as an important concept on itself, illustrated by a number of stochastic and deterministic tensorization techniques.

Keywords: blind source separation, independent component analysis, tensorization, canonical polyadic decomposition, block term decomposition, higher-order tensor, multilinear algebra

1 Blind signal separation and matrix data

The separation of sources from observed data is a well-known problem in signal processing, known as blind signal separation (BSS). The linear BSS problem consists of the decomposition of an observed data matrix $\mathbf{X} \in \mathbb{K}^{K\times N}$ as

$$\mathbf{X} = \mathbf{M} \cdot \mathbf{S} = \sum_{r=1}^{R} \mathbf{m}_r \cdot \mathbf{s}_r^{\mathrm{T}}, \tag{1}$$

in which $\mathbf{M} \in \mathbb{K}^{K\times R}$ is the mixing matrix and $\mathbf{S} \in \mathbb{K}^{R\times N}$ is the observed source matrix. The vector $\mathbf{m}_r$ is the rth column of $\mathbf{M}$ and $\mathbf{s}_r^{\mathrm{T}}$ is the rth row of $\mathbf{S}$. For each signal N samples are available. The set $\mathbb{K}$ stands for either $\mathbb{R}$ or $\mathbb{C}$. Furthermore, additive noise can be represented by a matrix $\mathbf{N} \in \mathbb{K}^{K\times N}$.

Equation (1) is a decomposition of the data matrix $\mathbf{X}$ in rank-1 terms, where each term corresponds to the contribution of one particular source. Except in the case of a single source with $R = 1$, it is well-known that such a decomposition is not unique. Uniqueness appears by imposing additional constraints on the matrices. Acclaimed matrix decompositions with well-understood uniqueness conditions are the singular value decomposition (imposing column-wise orthogonality) and the QR and RQ factorizations (imposing triangularity and column-wise orthonormality). However, in the light of BSS, the constraints from these well-known decompositions are both too restrictive and unnatural. For instance, it is uncommon that the mixing matrix is known to be triangular, as it is uncommon that both mixing vectors and source vectors are mutually orthogonal. We are facing here what is called the factor indeterminacy problem in Factor Analysis (FA) [29]. One needs to resort to other assumptions and matrix decompositions, specifically tailored to the BSS problem.

One of the more realistic constraints for BSS is nonnegativity: nonnegative matrix factorization (NMF) is a decomposition in which the entries of the factor matrices are nonnegative [8, 27, 35, 38]. Nonnegativity is natural for concentrations, number of occurrences, pixel intensities, frequencies, etc. Sparse component analysis (SCA) is also gaining in popularity [5, 44]. In SCA, the source matrix $\mathbf{S}$ is assumed to be sparse. Note that nonnegativity in itself does not ensure uniqueness; additional sparsity is necessary too [21, 23, 26, 30]. For dense data sets, SCA is mostly applied after a sparsifying transformation such as the wavelet transformation [15].

2 Blind signal separation and tensor data

A tensor is a higher-order generalization of vectors (boldface lowercase letters) and matrices (boldface uppercase letters). It is denoted by a calligraphic letter, e.g., $\mathcal{X}$, and is a multiway array of numerical values $x_{i_1 i_2 \cdots i_N} = \mathcal{X}(i_1, i_2, \ldots, i_N)$ where $\mathcal{X} \in \mathbb{K}^{I_1 \times I_2 \times \cdots \times I_N}$. By fixing all but a single index, one obtains a mode-n vector, e.g., $\mathbf{a} = \mathcal{X}(i_1, \ldots, i_{n-1}, :, i_{n+1}, \ldots, i_N) \in \mathbb{K}^{I_n}$. A diagonal tensor only has nonzeros on the entries of which all the indices are equal.

The third-order counterpart of (1) is a decomposition of a tensor $\mathcal{X} \in \mathbb{K}^{I \times J \times K}$ in R rank-1 terms:

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r = \mathcal{I} \cdot_1 \mathbf{A} \cdot_2 \mathbf{B} \cdot_3 \mathbf{C}, \tag{2}$$

in which $\otimes$ denotes the tensor (outer) product, $\cdot_i$ denotes the tensor-matrix product in the ith mode and $\mathcal{I}$ denotes a diagonal tensor with ones on the diagonal and zeros elsewhere. For all index values, we have that $x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr}$. Eq. (2) gives a polyadic decomposition (PD) of $\mathcal{X}$. If R is minimal, it is defined as the rank of $\mathcal{X}$ and the decomposition is called a canonical polyadic decomposition (CPD). It has been proven that the CPD is unique under relatively mild conditions, typically expressing that the rank-1 terms are "sufficiently different" while not necessitating additional constraints such as nonnegativity [34, 19, 20].

Recently, the block term decomposition (BTD) has been introduced [12, 14]. Instead of decomposing a tensor in rank-1 terms, it is written as a linear combination of tensors with low multilinear rank. The multilinear rank of a tensor $\mathcal{X}$

is an N-tuple $(R_1, R_2, \ldots, R_N)$ with R_n the mode-n rank, defined as the dimension of the subspace spanned by the mode-n vectors of $\mathcal{X}$. A special instance of the BTD is the decomposition of a third-order tensor $\mathcal{X} \in \mathbb{K}^{I \times J \times K}$ in rank-$(L_r, L_r, 1)$ terms for which uniqueness under mild conditions has been proven too [12, 13]. We then have

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{E}_r \otimes \mathbf{c}_r = \sum_{r=1}^{R} (\mathbf{A}_r \mathbf{B}_r^{\mathrm{T}}) \otimes \mathbf{c}_r, \tag{3}$$

with matrices $\mathbf{E}_r = \mathbf{A}_r \mathbf{B}_r^{\mathrm{T}} \in \mathbb{K}^{I \times J}$ of rank L_r. The matrices $\mathbf{A}_r \in \mathbb{K}^{I \times L_r}$ and $\mathbf{B}_r \in \mathbb{K}^{J \times L_r}$ have full column rank, and we have nonzero $\mathbf{c}_r \in \mathbb{K}^{K}$ for all r.

Tensor methods for BSS receive their success from the uniqueness of tensor decompositions such as the CPD and the BTD. These are becoming standard tools for BSS and have been applied in many domains such as telecommunication, array processing and chemometrics [7, 32, 33, 42].

3 Tensorization of matrix data

Tensor techniques require the presence of tensor data. Matrix data obviously remain more common than tensor data. Nevertheless, the techniques may still be used for BSS after the data matrix is mapped to a tensor. The mapping to the tensor domain translates the assumptions made for BSS, with the subsequent tensor decompositions having the possibility of ensuring uniqueness. While the uniqueness and algorithms of tensor decompositions have received a lot of attention lately, we discuss different tensorization techniques. A clear overview is necessary to benefit from the advantages of tensor techniques for matrix data.

What is essential about the mappings, is that linear transformations are used that map the sources to matrices or tensors that (approximately) have low (multilinear) rank under a certain working hypothesis. The (multi)linearity of the transformation is necessary to retain a linear mixture of the sources and avoid the introduction of inseparable terms, while the low-rank structure enables us to apply the tensor decompositions of the previous section.

In a first subsection, we discuss a stochastic tensorization technique using higher-order statistics. The second subsection describes the use of parameter variation for tensorization, illustrated with second-order statistics. Three different deterministic techniques relying on Hankelization, Löwnerization and segmentation are discussed in Sections 3.3, 3.4 and 3.5, respectively. Note that the uses of higher-order statistics and second-order statistics for BSS are well known, both applying tensorization in a different way. For each tensorization technique described, the multilinearity, working hypothesis, applied tensor decomposition and higher-order representation of each source are reported.

3.1 Higher-Order Statistics

Higher-order statistics (HOS) are fundamental for independent component analysis (ICA), in which one separates the observations in mutually statistically independent sources. This technique for BSS is highly renowned and has been applied in a diversity of domains [6, 9, 10, 36, 37]. Within the different types of higher-order statistics, especially cumulants are compelling. They are able to separate non-Gaussian, mutually independent sources. For simplicity, we assume stationary, identically distributed signals. Consider a zero-mean stochastic signal vector $\mathbf{u}(t)$. We give the explicit definition of the fourth-order cumulant:

$$\begin{aligned}\left(\mathcal{C}_{\mathbf{u}}^{(4)}\right)_{i_1 i_2 i_3 i_4} \triangleq\ & \mathrm{E}\left\{u_{i_1} u_{i_2}^* u_{i_3}^* u_{i_4}\right\} - \mathrm{E}\left\{u_{i_1} u_{i_2}^*\right\} \mathrm{E}\left\{u_{i_3}^* u_{i_4}\right\} \\ & - \mathrm{E}\left\{u_{i_1} u_{i_3}^*\right\} \mathrm{E}\left\{u_{i_2}^* u_{i_4}\right\} - \mathrm{E}\left\{u_{i_1} u_{i_4}\right\} \mathrm{E}\left\{u_{i_2}^* u_{i_3}^*\right\}. \end{aligned} \quad (4)$$

Cumulants have very interesting properties enabling the use of tensor decompositions for BSS [37]. First of all, the expression in Eq. (4) satisfies multilinearity (it gives a quadrilinear mapping) as requested from the introduction of the section: if $\mathbf{x}(t) = \mathbf{M}\mathbf{s}(t) + \mathbf{n}(t)$ then in the fourth-order case we have:

$$\mathcal{C}_{\mathbf{x}}^{(4)} = \mathcal{C}_{\mathbf{s}}^{(4)} \cdot_1 \mathbf{M} \cdot_2 \mathbf{M}^* \cdot_3 \mathbf{M}^* \cdot_4 \mathbf{M} + \mathcal{C}_{\mathbf{n}}^{(4)}. \quad (5)$$

Second, higher-order cumulants of a Gaussian variable are zero. Under the assumption of Gaussian noise, $\mathcal{C}_{\mathbf{n}}^{(4)}$ from Eq. (5) becomes a zero tensor.

The working hypothesis in ICA with HOS is that the sources are non-Gaussian and mutually statistically independent. Then, the higher-order source cumulant $\mathcal{C}_{\mathbf{s}}^{(4)}$ from Eq. (5) is a diagonal tensor, with kurtoses κ_{s_r} as diagonal entries for $1 \leq r \leq R$. Hence, under the working hypothesis, Eq. (5) admits a CPD with a rank R:

$$\mathcal{C}_{\mathbf{x}}^{(4)} = \sum_{r=1}^{R} \kappa_{s_r} \otimes \mathbf{m}_r \otimes \mathbf{m}_r^* \otimes \mathbf{m}_r^* \otimes \mathbf{m}_r + \mathcal{C}_{\mathbf{n}}^{(4)}, \quad (6)$$

with $\mathbf{M}$ satisfying the uniqueness conditions. The separation of the source vectors and mixing vectors in Eq. (1) has been translated to the identification of rank-1 terms in (6) as each source contributes a rank-1 term to the CPD.

A variant of applying a CPD in (6) is to use a maximal diagonalization technique [9] or the joint approximate diagonalization of eigenmatrices method (JADE) [6]. They are used in conjunction with a prewhitening step using the second-order covariance matrix.

3.2 Parameter Variation

Given some matrix data, one can perform a (multilinear) transformation depending upon a parameter to generate a set of matrices. After stacking them, a third-order tensor is obtained which can be decomposed to identify the underlying unknown components. It is used in the decoupling of multivariate polynomials [22] but also in BSS with the second-order blind identification (SOBI) algorithm [2] and variants. In SOBI, the set of matrices consists of lagged covariance matrices. Let us define

$$\mathbf{C}_{\mathbf{u}}(\tau) = \mathrm{E}\left\{\mathbf{u}(t)\mathbf{u}(t+\tau)^{\mathrm{H}}\right\} \quad (7)$$

as the covariance matrix with a lag τ of a stochastic signal vector $\mathbf{u}(t)$. Observe that Eq. (7) gives a bilinear transformation: if $\mathbf{x}(t) = \mathbf{M}\mathbf{s}(t)+\mathbf{n}(t)$, then $\mathbf{C_x}(\tau) = \mathbf{M} \cdot \mathbf{C_s}(\tau) \cdot \mathbf{M}^{\mathrm{H}} + \mathbf{C_n}(\tau)$. For multiple lags $\tau_1, \dots, \tau_L$ we then have:

$$\begin{cases} \mathbf{C_x}(\tau_1) = \mathbf{M} \cdot \mathbf{C_s}(\tau_1) \cdot \mathbf{M}^{\mathrm{H}} + \mathbf{C_n}(\tau_1), \\ \qquad\qquad \vdots \\ \mathbf{C_x}(\tau_L) = \mathbf{M} \cdot \mathbf{C_s}(\tau_L) \cdot \mathbf{M}^{\mathrm{H}} + \mathbf{C_n}(\tau_L). \end{cases} \tag{8}$$

The working hypothesis made by SOBI is that the source signals are mutually uncorrelated but individually correlated for the different lags $\tau_1, \dots, \tau_L$[1]. Then, the corresponding lagged covariance matrices of the sources are diagonal matrices. Hence, the matrices $\mathbf{M}$ and $\mathbf{M}^*$ simultaneously diagonalize the lagged covariance matrices of $\mathbf{x}(t)$ in (8) [11]. Let us define $\sigma^2_{s_r}(\tau_l)$ as the autocovariance of source $s_r(t)$ for the given lag τ_l. We collect them for each source in a vector $\boldsymbol{\sigma}^2_{s_r} \in \mathbb{K}^L$ for all τ_l, $1 \leq l \leq L$. By stacking $\mathbf{C_x}(\tau_l)$ in the third dimension of a tensor $\mathcal{C}_\mathbf{x}$ and assuming the noise level is low, a CPD emerges:

$$\mathcal{C}_\mathbf{x} = \sum_{r=1}^{R} \mathbf{m}_r \otimes \mathbf{m}_r^* \otimes \boldsymbol{\sigma}^2_{s_r} + \mathcal{C}_\mathbf{n} = \mathcal{I} \cdot_1 \mathbf{M} \cdot_2 \mathbf{M}^* \cdot_3 \boldsymbol{\Sigma} + \mathcal{C}_\mathbf{n}, \tag{9}$$

in which $\boldsymbol{\Sigma} \in \mathbb{K}^{L \times R}$ contains the columns $\boldsymbol{\sigma}^2_{s_r}$ for $1 \leq r \leq R$. Note that each source contributes a rank-1 term to $\mathcal{C}_\mathbf{x}$. In [11], the connection between simultaneous matrix diagonalization and CPD is discussed.

A variant for nonstationary sources of the SOBI tensorization method is the stacking of a set of covariance matrices computed for different time frames [39].

3.3 Hankelization

Consider an exponential signal $f(k) = az^k$ arranged in a Hankel matrix $\mathbf{H}$. The matrix appears to have rank 1:

$$\mathbf{H} = \begin{bmatrix} f(0) & f(1) & f(2) & \cdots \\ f(1) & f(2) & f(3) & \cdots \\ f(2) & f(3) & f(4) & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix} = a \begin{bmatrix} 1 \\ z \\ z^2 \\ \vdots \end{bmatrix} \begin{bmatrix} 1 & z & z^2 & \cdots \end{bmatrix}. \tag{10}$$

These simple exponential functions can be generalized to exponential polynomials, which are functions that can be written as sums and/or products of exponentials, sinusoids and/or polynomials. They have a broad relevance: for (multidimensional) harmonic retrieval, direction-of-arrival estimation, sinusoidal carriers in telecommunication, etc. [24, 40, 31, 41]. Furthermore, they can be used to model various signal shapes. The idea is analogous to the approximation of functions with the well-known Taylor series expansion. Figures 1 and 2 show approximations of a sigmoid and Gaussian function through Hankelization.

[1] Note that the autocorrelation is not required for each source for each of the lags.

It has been shown that for an exponential polynomial signal of degree δ, the corresponding Hankel matrix will have rank δ [43]. The Hankel tensorization technique for BSS exists in mapping each row of the observed data matrix $\mathbf{X}$ from (1) to a Hankel matrix which is being stacked in a third-order tensor $\mathcal{H}_{\mathbf{X}}$. Defining $\mathbf{H}_{\mathbf{s}_r}$ as the Hankel matrix of the rth source $\mathbf{s}_r$, we have because of linearity that:

$$\mathcal{H}_{\mathbf{X}} = \sum_{r=1}^{R} \mathbf{H}_{\mathbf{s}_r} \otimes \mathbf{m}_r = \sum_{r=1}^{R} (\mathbf{A}_r \mathbf{B}_r^{\mathrm{T}}) \otimes \mathbf{m}_r. \tag{11}$$

The latter transition is based on the working hypothesis that the rth source can be approximated by an exponential polynomial of (low) degree L_r. Each matrix $\mathbf{H}_{\mathbf{s}_r}$ has (low) rank L_r then, and we have full column rank matrices $\mathbf{A}_r \in \mathbb{K}^{I \times L_r}$ and $\mathbf{B}_r \in \mathbb{K}^{J \times L_r}$. Hence, after the Hankel-tensorization (or Hankelization), a decomposition in rank-$(L_r, L_r, 1)$ terms like in Eq. (11) can be applied. Instead of each source contributing a rank-1 tensor, it contributes a tensor with low multilinear rank, namely $(L_r, L_r, 1)$.

3.4 Löwnerization

Another class of functions suitable for BSS is the set of rational functions, able to take on a very wide range of shapes. An illustration is given in Figure 1 and Figure 2 by approximating a sigmoid and Gaussian function. Rational functions have the same connection with Löwner matrices as exponential polynomials have with Hankel matrices [1, 25]. Given a function $f(t)$ sampled on $N = I + J$ points which are divided in two distinct point sets $X = \{x_1, \ldots, x_I\}$ and $Y = \{y_1, \ldots, y_J\}$, we define the entries of the Löwner matrix $\mathbf{L} \in \mathbb{K}^{I \times J}$ as follows:

$$\forall i, j : \quad l_{i,j} = \frac{f(x_i) - f(y_j)}{x_i - y_j}. \tag{12}$$

It has been shown in [16, 17] that an equivalent formulation as in Eq. (11) can be made: because of the linearity of the Löwner transformation, the tensor $\mathcal{L}_{\mathbf{X}}$, obtained by mapping every row of the observed data matrix $\mathbf{X}$ to a Löwner matrix and stacking these matrices, can be written as a linear combination of the Löwner matrices of the sources. Under the working hypothesis that the rth source can be modeled as a rational function of (low) degree L_r, the corresponding Löwner matrix will have (low) rank L_r. Like in the Hankel case, a BTD is obtained where the rth source contributes a rank-$(L_r, L_r, 1)$ term to $\mathcal{L}_{\mathbf{X}}$.

3.5 Segmentation

Segmentation is a general term used to denote the reshaping of a vector into a matrix, i.e., extracting small segments and stacking them after each other. Consider the following exponential vector: $\begin{bmatrix} 1 \; z \; z^2 \; z^3 \; z^4 \; z^5 \end{bmatrix}$. If it is reshaped to a matrix, the latter has rank 1:

$$\begin{bmatrix} 1 \; z \; z^2 \; z^3 \; z^4 \; z^5 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & z & z^2 \\ z^3 & z^4 & z^5 \end{bmatrix} = \begin{bmatrix} 1 \\ z^3 \end{bmatrix} \begin{bmatrix} 1 \; z \; z^2 \end{bmatrix}. \tag{13}$$

Focusing on BSS, let us now reshape the kth row of the observed data matrix $\mathbf{X} \in \mathbb{K}^{K \times N}$ to a matrix $\mathbf{E}_{x_k} \in \mathbb{K}^{I \times J}$ with $N = I \times J$ for $k = 1, \ldots, K$, and stack these matrices in a tensor $\mathcal{X}$. The transformation is clearly linear. Let us start from the assumption that the segmented matrix of each source has rank 1, as in Eq. (13). One obtains the following CPD:

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{E}_{s_r} \otimes \mathbf{m}_r = \sum_{r=1}^{R} \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{m}_r. \tag{14}$$

with rank-1 matrices $\mathbf{E}_{s_r} = \mathbf{a}_r \otimes \mathbf{b}_r$ and vectors $\mathbf{a}_r \in \mathbb{K}^I$ and $\mathbf{b}_r \in \mathbb{K}^J$. This is equivalent to stating that the rth source signal can be written as a Kronecker product $\mathbf{a}_r^{\mathrm{T}} \otimes \mathbf{b}_r^{\mathrm{T}}$ for $r = 1, \ldots, R$, with the Kronecker product for row vectors $\mathbf{u} \in \mathbb{K}^{1 \times I}$, $\mathbf{v} \in \mathbb{K}^{1 \times J}$ defined as $\mathbf{u} \otimes \mathbf{v} = \begin{bmatrix} u_1\mathbf{v} \; u_2\mathbf{v} \; \cdots \; u_I\mathbf{v} \end{bmatrix}$.

Although the hypothesis is fulfilled when the sources are, for instance, exponential functions, it is quite restrictive. By increasing the assumed rank $L_r \geq 1$ of the reshaped matrices $\mathbf{E}_{s_r}$, we obtain a BTD in rank-$(L_r, L_r, 1)$ terms:

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{E}_{s_r} \otimes \mathbf{m}_r = \sum_{r=1}^{R} (\mathbf{A}_r \mathbf{B}_r^{\mathrm{T}}) \otimes \mathbf{m}_r, \tag{15}$$

with matrices $\mathbf{A}_r \in \mathbb{K}^{I \times L_r}$ and $\mathbf{B}_r \in \mathbb{K}^{J \times L_r}$. Adding a subscript l to denote the lth column of the matrices $\mathbf{A}_r$ and $\mathbf{B}_r$, the working hypothesis now becomes that the source signals can be modeled as, or approximated by, sums of Kronecker products: $\mathbf{s}_r = \sum_{l=1}^{L_r} \mathbf{a}_{r,l}^{\mathrm{T}} \otimes \mathbf{b}_{r,l}^{\mathrm{T}}$. An example of a source exactly displaying this structure is a sine wave, which can be written as a sum of two Kronecker products. Other functions can be approximated too, e.g. sigmoid and Gaussian functions, illustrated in Figures 1 and 2. While each source contributed a rank-1 term to $\mathcal{X}$ for the first hypothesis, it now contributes a term with low multilinear rank $(L_r, L_r, 1)$.

Note that because of the segmentation and the structure of the low-rank decompositions, a nonnegligible compression is obtained in the number of underlying variables. This is especially useful for big data systems, with many observed samples or signals. The technique has been described in [3, 4] for large-scale BSS problems, including a generalization for higher-order segmentation. Segmentation of signal vectors to matrices or tensors has been successfully applied in various domains before, such as biomedical signal processing [18] and scientific computing for large-scale models with high dimensions and a very high number of numerical values [28].

4 Discussion and Conclusion

In many techniques for blind signal separation (BSS), multilinear algebra is used to recover the mixing vectors and the original source signals. Given only an observed data matrix, a transformation is made to higher-order structures called tensors. This paper introduces the tensorization step as an important concept

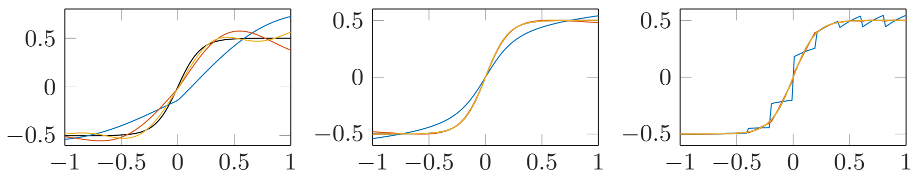

Fig. 1. Approximation of a sigmoid function $f(t) = \frac{1}{1+e^{-10t}}$. It is sampled uniformly 100 times in $[-1, 1]$ (——). To the left, an approximation with exponential polynomials is used by Hankelizing the samples. In the middle, Löwnerization is applied. To the right, segmentation with $I = J = 10$ is used. The tensorized matrix is approximated by a low-rank matrix through truncation of the singular value decomposition, after which the underlying signal is calculated from this low-rank matrix. Approximations for ranks $R = 1$ (——), $R = 2$ (——) and $R = 3$ (——) are shown.

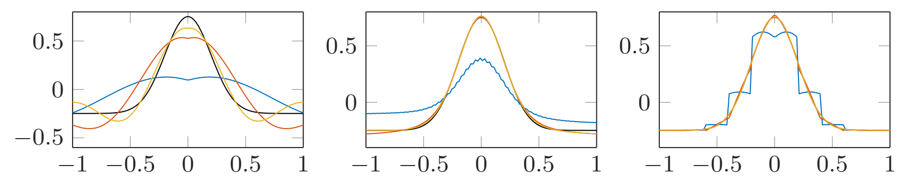

Fig. 2. Approximation of a Gaussian function $f(t) = e^{-\frac{1}{2}(5t)^2}$, sampled uniformly 100 times in $[-1, 1]$ (——). An equal procedure as in Figure 1 is used, with Hankelization (left), Löwnerization (middle) and segmentation (right) for ranks $R = 1$ (——), $R = 2$ (——) and $R = 3$ (——).

by itself, as many results concerning tensorization have appeared in the literature in a disparate manner and have not been discussed as such. Higher-order statistics and second-order statistics, for example, are well-known to solve BSS, but apply tensorization in a significantly different way. Many links to multilinear algebra from other existing BSS techniques have not yet been established. Because of space limitations, the presentation of the idea has been restricted to instantaneous mixtures of one-dimensional sources. A following paper will discuss generalizations such as multidimensional sources or convolutive mixtures.

Acknowledgments. The research is funded by (1) a Ph.D. grant of the Agency for Innovation by Science and Technology (IWT), (2) Research Council KU Leuven: CoE PFV/10/002 (OPTEC), (3) F.W.O.: projects G.0830.14N and G.0881.14N, (4) the Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO II, Dynamical systems, control and optimization, 2012-2017), (5) EU: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (no. 339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

References

1. Antoulas, A.C., Anderson, B.D.O.: On the scalar rational interpolation problem. IMA Journal of Mathematical Control and Information 3(2-3), 61–88 (1986)
2. Belouchrani, A., Abed-Meraim, K., Cardoso, J., Moulines, E.: A blind source separation technique using second-order statistics. IEEE Transactions on Signal Processing 45(2), 434–444 (1997)
3. Boussé, M., Debals, O., De Lathauwer, L.: Deterministic blind source separation using low-rank tensor approximations (April 2015), Internal Report 15-59, ESAT-STADIUS, KU Leuven, Belgium
4. Boussé, M., Debals, O., De Lathauwer, L.: A novel deterministic method for large-scale blind source separation (January 2015), Internal Report 15-07, ESAT-STADIUS, KU Leuven, Belgium
5. Bruckstein, A.M., Donoho, D.L., Elad, M.: From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM Review 51(1), 34–81 (2009)
6. Cardoso, J.F., Souloumiac, A.: Blind beamforming for non-gaussian signals. Radar and Signal Processing, IEE Proceedings F 140(6), 362–370 (1993)
7. Cichocki, A., Mandic, D., Phan, A.H., Caiafa, C., Zhou, G., Zhao, Q., De Lathauwer, L.: Tensor decompositions for signal processing applications: From two-way to multiway component analysis. IEEE Signal Processing Magazine 32(2), 145–163 (2015)
8. Cichocki, A., Zdunek, R., Phan, A., Amari, S.: Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation. Wiley (2009)
9. Comon, P.: Independent component analysis, a new concept? Signal Processing 36(3), 287–314 (Apr 1994)
10. Comon, P., Jutten, C.: Handbook of blind source separation: Independent component analysis and applications. Academic Press (2010)
11. De Lathauwer, L.: A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization. SIAM Journal on Matrix Analysis and Applications 28(3), 642–666 (2006)
12. De Lathauwer, L.: Decompositions of a higher-order tensor in block terms — Part II: Definitions and uniqueness. SIAM Journal on Matrix Analysis and Applications 30(3), 1033–1066 (2008)
13. De Lathauwer, L.: Blind separation of exponential polynomials and the decomposition of a tensor in rank-$(L_r, L_r, 1)$ terms. SIAM Journal on Matrix Analysis and Applications 32(4), 1451–1474 (2011)
14. De Lathauwer, L., Nion, D.: Decompositions of a higher-order tensor in block terms — Part III: Alternating least squares algorithms. SIAM Journal on Matrix Analysis and Applications 30(3), 1067–1083 (2008)
15. De Vos, M.: Decomposition methods with applications in neuroscience. Ph.D. thesis, KU Leuven (2009)
16. Debals, O., Van Barel, M., De Lathauwer, L.: Blind signal separation of rational functions using Löwner-based tensorization. IEEE Proceedings on International Conference on Acoustics, Speech and Signal Processing (April 2015), Accepted for publication
17. Debals, O., Van Barel, M., De Lathauwer, L.: Löwner-based blind signal separation of rational functions with applications (March 2015), Internal Report 15-44, ESAT-STADIUS, KU Leuven, Belgium

18. Deburchgraeve, W., Cherian, P., De Vos, M., Swarte, R., Blok, J., Visser, G.H., Govaert, P., Van Huffel, S.: Neonatal seizure localization using PARAFAC decomposition. Clinical Neurophysiology 120(10), 1787–1796 (2009)
19. Domanov, I., De Lathauwer, L.: On the uniqueness of the canonical polyadic decomposition of third-order tensors — Part I: Basic results and uniqueness of one factor matrix. SIAM Journal on Matrix Analysis and Applications 34(3), 855–875 (2013)
20. Domanov, I., De Lathauwer, L.: On the uniqueness of the canonical polyadic decomposition of third-order tensors — Part II: Uniqueness of the overall decomposition. SIAM Journal on Matrix Analysis and Applications 34(3), 876–903 (2013)
21. Donoho, D., Stodden, V.: When does non-negative matrix factorization give a correct decomposition into parts? In: Advances in neural information processing systems (2003)
22. Dreesen, P., Ishteva, M., Schoukens, J.: Decoupling multivariate polynomials using first-order information and tensor decompositions. SIAM J. Matrix Anal. Appl. (2015), Accepted for publication
23. Eggert, J., Korner, E.: Sparse coding and NMF. IEEE Proceedings of International Joint Conference on Neural Networks 4, 2529–2533 (2004)
24. Elad, M., Milanfar, P., Golub, G.H.: Shape from moments — an estimation theory perspective. IEEE Transactions on Signal Processing 52(7), 1814–1829 (2004)
25. Fiedler, M.: Hankel and Löwner matrices. Linear Algebra and its Applications 58, 75–95 (1984)
26. Georgiev, P., Theis, F., Cichocki, A.: Sparse component analysis and blind source separation of underdetermined mixtures. IEEE Transactions on Neural Networks 16(4), 992–996 (July 2005)
27. Gillis, N.: Nonnegative matrix factorization: Complexity, algorithms and applications. Ph.D. thesis, UCL (2011)
28. Grasedyck, L.: Polynomial approximation in hierarchical Tucker format by vector-tensorization. Citeseer (2010)
29. Harman, H.H.: Modern factor analysis. Univ. of Chicago Press (1976), third ed.
30. Hoyer, P.O.: Non-negative matrix factorization with sparseness constraints. The Journal of Machine Learning Research 5, 1457–1469 (2004)
31. Jiang, T., Sidiropoulos, N.D., ten Berge, J.M.: Almost-sure identifiability of multidimensional harmonic retrieval. IEEE Transactions on Signal Processing 49(9), 1849–1859 (2001)
32. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Review 51(3), 455–500 (2009)
33. Kroonenberg, P.: Applied multiway data analysis, vol. 702. Wiley-Interscience (2008)
34. Kruskal, J.B.: Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. Linear Algebra and its Applications 18(2), 95 – 138 (1977)
35. Lee, D., Seung, H., et al.: Learning the parts of objects by non-negative matrix factorization. Nature 401(6755), 788–791 (1999)
36. McCullagh, P.: Tensor methods in statistics, vol. 161. Chapman and Hall London (1987)
37. Nikias, C.L., Petropulu, A.P.: Higher-order spectra analysis: A nonlinear signal processing framework. PTR Prentice Hall, Englewood Cliffs, NJ (1993)
38. Paatero, P., Tapper, U.: Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics 5(2), 111–126 (1994)

39. Pham, D.T., Cardoso, J.F.: Blind separation of instantaneous mixtures of nonstationary sources. IEEE Transactions on Signal Processing 49(9), 1837–1848 (2001)
40. Roemer, F., Haardt, M., Del Galdo, G.: Higher order SVD based subspace estimation to improve multi-dimensional parameter estimation algorithms. In: Fortieth Asilomar Conference on Signals, Systems and Computers. pp. 961–965. IEEE (2006)
41. Sidiropoulos, N.D.: Generalizing Caratheodory's uniqueness of harmonic parameterization to N dimensions. IEEE Transactions on Information Theory 47(4), 1687–1690 (2001)
42. Smilde, A.K., Bro, R., Geladi, P., Wiley, J.: Multi-way analysis with applications in the chemical sciences. Wiley Chichester, UK (2004)
43. Vandevoorde, D.: A fast exponential decomposition algorithm and its applications to structured matrices. Ph.D. thesis, Rensselaer Polytechnic Institute, Troy, NY (1998)
44. Zibulevsky, M., Pearlmutter, B.: Blind source separation by sparse decomposition in a signal dictionary. Neural computation 13(4), 863–882 (2001)