KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Boussé, M., Debals, O., and De Lathauwer, L. (2017) **A Tensor-Based Method for Large-Scale Blind Source Separation using Segmentation** IEEE Transactions on Signal Processing, vol. 65, no. 2, p. 346-358
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	http://dx.doi.org/10.1109/TSP.2016.2617858
Journal homepage	https://signalprocessingsociety.org/publications-resources/ieee-transactions-signal-processing
Author contact	Martijn.Bousse@esat.kuleuven.be + 32 (0)16 32 78 10
Abstract	Many real-life signals are compressible, meaning that they depend on much fewer parameters than their sample size. In this paper we use low-rank matrix or tensor representations for signal compression. We propose a new deterministic method for blind source separation that exploits the low-rank structure, enabling a unique separation of the source signals and providing a way to cope with large-scale data. We explain that our method reformulates the blind source separation problem as the computation of a tensor decomposition, after reshaping the observed data matrix into a tensor. This deterministic tensorization technique is called segmentation and is closely related to Hankel-based tensorization. We apply the same strategy to the mixing coefficients of the blind source separation problem, as in many large-scale applications the mixture is also compressible because of many closely located sensors. Moreover, we combine both strategies, resulting in a general technique

	that allows us to exploit the underlying compactness of the sources and the mixture simultaneously. We illustrate the techniques for fetal electrocardiogram extraction and direction-of-arrival estimation in large-scale antenna arrays.
IR	https://lirias.kuleuven.be/handle/123456789/554595

(article begins on next page)

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2016.2617858

A Tensor-Based Method for Large-Scale Blind Source Separation Using Segmentation

Martijn Boussé, *Student Member, IEEE*, Otto Debals, *Student Member, IEEE*, Lieven De Lathauwer, *Fellow, IEEE*

***Abstract*—Many real-life signals are compressible, meaning that they depend on much fewer parameters than their sample size. In this paper we use low-rank matrix or tensor representations for signal compression. We propose a new deterministic method for blind source separation that exploits the low-rank structure, enabling a unique separation of the source signals and providing a way to cope with large-scale data. We explain that our method reformulates the blind source separation problem as the computation of a tensor decomposition, after reshaping the observed data matrix into a tensor. This deterministic tensorization technique is called segmentation and is closely related to Hankel-based tensorization. We apply the same strategy to the mixing coefficients of the blind source separation problem, as in many large-scale applications the mixture is also compressible because of many closely located sensors. Moreover, we combine both strategies, resulting in a general technique that allows us to exploit the underlying compactness of the sources and the mixture simultaneously. We illustrate the techniques for fetal electrocardiogram extraction and direction-of-arrival estimation in large-scale antenna arrays.**

***Index Terms*—Blind source separation, higher-order tensor, tensor decomposition, low-rank approximation, big data**

I. Introduction

In blind source separation (BSS) one tries to reconstruct a set of unobserved sources based only on a set of observed signals. In this paper, the latter are unknown linear instantaneous mixtures of the sources. Applications can be found in telecommunications, signal processing and biomedical sciences [1]–[4]. In general, there is no unique solution to the BSS problem, hence, one imposes additional assumptions.

A well-known BSS method, called independent component analysis (ICA), assumes statistically independent sources [4]. Several ICA methods use higher-order statistics (HOS) in order to tensorize the BSS problem and then apply a tensor decomposition to uniquely identify the sources. Recently, a class of deterministic methods has been proposed that do not use (higher-order) statistics but assume that the sources can be modeled as, e.g., exponential polynomials or rational functions [5], [6]. Specific tensorization techniques can be used, such as Hankel-based or Löwner-based tensorization [7]. The source signals can then be uniquely recovered by block component analysis (BCA). BCA is a framework based on block term decompositions which was introduced in [8]–[10]. These methods, as well as the method we propose here, go further than dictionary-based methods. In the latter, one defines *a priori* a fixed signal dictionary in which one assumes the sources can be described sparsely and then one exploits this sparse representability to identify the sources [11], [12]. Here, we do not need an initial dictionary.

In this paper, we introduce a new method for BSS that exploits the fact that many real-life signals are compressible, i.e., the fact that they can be described in terms of much fewer parameters than the actual number of samples [13], [14]. One way of representing signals in a (possibly very) compact way is a (higher-order) low-rank approximation of a tensorized version of the signal [15]. This can be interpreted as approximating the original signals by sums of Kronecker products of smaller vectors. This strategy is similar to tensor-based scientific computing in high dimensions [15]–[17], which has allowed one to solve problems in a number of unknowns that exceeds the number of atoms in the universe. It is used in a novel way for BSS in this paper and is a key idea to handle large-scale BSS problems, i.e., problems with many sensors and/or samples. In particular, we use a deterministic tensorization technique, called segmentation, that reshapes each observed signal into a matrix (tensor) and stacks them into a (higher-order) tensor. The latter can be interpreted as a compact version of the Hankel-based tensorization mentioned above. We show that the BSS problem boils down to the computation of a decomposition of the tensor obtained by segmentation if the sources exhibit the hypothesized low-rank structure. This yields a unique solution to the BSS problem and provides a way to cope with large-scale problems where conventional methods fall short. Also, it is illustrated that our method allows the separation of underdetermined mixtures, i.e., the separation of more sources than observed signals.

We can apply the same strategy to the mixing coefficients of the BSS problem (instead of the source signals) following a similar argument. Indeed, in the context of big data, we see a large increase in the number of sensors and/or sensor density in fields such as biomedical sciences and sensor array

Manuscript received April 1, 2015; revised August 12, 2015; accepted October 2, 2016.

Copyright © 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

This research is funded by (1) a Ph.D. grant of the Agency for Innovation by Science and Technology (IWT), (2) Research Council KU Leuven: C1 project C16/15/059-nD, CoE PFV/10/002 (OPTEC). (3) FWO: project: G.0830.14N, G.0881.14N, (4) the Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO, 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.

Martijn Boussé, Otto Debals, and Lieven De Lathauwer are with the Department of Electrical Engineering (ESAT), KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium (e-mail: martijn.bousse@kuleuven.be and otto.debals@kuleuven.be, lieven.delathauwer@kuleuven.be). Otto Debals and Lieven De Lathauwer are also with the Group of Science, Engineering and Technology, KU Leuven Kulak, E. Sabbelaan 53, B-5800 Kortrijk, Belgium.

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2016.2617858

processing [18], [19]. The mixing coefficients are in that case often smoothly varying because of the many closely located sensors, allowing a (higher-order) low-rank approximation of a tensorized version of the mixing vectors. Conventional methods such as ICA fall short in a large-scale setting because of the exponential dependence on the order of the statistics. Exploiting low-rank structure on the mixing level was briefly discussed in [20]. In this paper, we go further: we apply the strategy on the sources, as described above, but also apply it on both the sources and the mixture simultaneously. The latter is a natural extension that results into a more general method that exploits the hypothesized low-rank structure of the simultaneously tensorized source and mixing level of the BSS problem, enabling a unique solution for large-scale BSS.

We illustrate the proposed methods with two applications. First, we have the separation of the fetal and maternal electrocardiogram (ECG) from multilead cutaneous potential recordings. Our method allows a clear separation of the two sources. Second, we have direction-of-arrival (DOA) estimation for large uniform linear arrays in both a line-of-sight and multipath setting. Our methods provides accurate estimates, even for close DOAs. In very large-scale applications, the arrays, however, are typically non-uniform but this is outside the scope of this paper; here, we focus on the main principles.

In the remainder of this section we introduce the notation and basic definitions. In Sections II and III, we introduce a new BSS method that exploits the hypothesized compressibility of the sources and mixing vectors, respectively. We combine both strategies in Section IV. Simulations and applications are presented in Section V. Finally, we conclude in Section VI.

A. Notation and definitions

Tensors, denoted by calligraphic letters (e.g., $\mathcal{A}$), are higher-order generalizations of vectors and matrices, denoted by bold lowercase (e.g., $\mathbf{a}$) and bold uppercase (e.g., $\mathbf{A}$) letters, respectively. The $(i_1, i_2, \ldots, i_N)$th entry of an Nth-order tensor $\mathcal{A} \in \mathbb{K}^{I_1 \times I_2 \times \cdots \times I_N}$, with $\mathbb{K}$ meaning $\mathbb{R}$ or $\mathbb{C}$, is denoted by $a_{i_1 i_2 \ldots i_N}$. The nth element in a sequence is indicated by a superscript between parentheses (e.g., $\{\mathbf{A}^{(n)}\}_{n=1}^{N}$). The unit vector with a one in the ith row is denoted as $\mathbf{e}_i$.

A mode-n vector of a tensor $\mathcal{A} \in \mathbb{K}^{I_1 \times I_2 \times \cdots \times I_N}$ is defined by fixing every index except the nth, e.g., $\mathbf{a}_{i_1 \cdots i_{n-1} : i_{n+1} \cdots i_N}$, and is a natural extension of the rows and columns of a matrix. The mode-n unfolding of $\mathcal{A}$ is a matrix $\mathbf{A}_{(n)}$ with the mode-n vectors as its columns (following the ordering convention in [21]). The vectorization of $\mathcal{A}$, denoted as $\text{vec}(\mathcal{A})$, maps each element $a_{i_1 i_2 \cdots i_N}$ onto $\text{vec}(\mathcal{A})_j$ with $j = 1 + \sum_{k=1}^{N} (i_k - 1) J_k$ and $J_k = \prod_{m=1}^{k-1} I_m$. The outer and Kronecker product are denoted by $\circ$ and $\otimes$, respectively, and are related through a vectorization: $\text{vec}\,(\mathbf{a} \circ \mathbf{b}) = \mathbf{b} \otimes \mathbf{a}$. A frontal slice of a third-order tensor $\mathcal{X} \in \mathbb{K}^{I \times J \times K}$, denoted by $\mathbf{X}_k$, is obtained by fixing the last index.

B. Tensor decompositions

An Nth-order tensor has rank one if it can be written as the outer product of N nonzero vectors. The rank of a tensor is defined as the minimal number of rank-1 terms that generate the tensor as their sum. The multilinear rank of an Nth-order tensor is equal to the tuple of mode-n ranks, which are defined as the ranks of the mode-n unfoldings of the tensor.

Definition 1. A *polyadic decomposition* (PD) writes an Nth-order tensor $\mathcal{A} \in \mathbb{K}^{I_1 \times I_2 \times \cdots \times I_N}$ as a sum of R rank-1 terms:

$$\mathcal{A} = \sum_{r=1}^{R} \mathbf{u}_r^{(1)} \circ \mathbf{u}_r^{(2)} \circ \cdots \circ \mathbf{u}_r^{(N)}. \tag{1}$$

The columns of the factor matrices $\mathbf{U}^{(n)} \in \mathbb{K}^{I_n \times R}$ are equal to the factor vectors $\mathbf{u}_r^{(n)}$ for $r = 1, \ldots, R$. The PD is called *canonical* (CPD) when R is equal to the rank of $\mathcal{A}$.

The CPD is a powerful model for several applications within signal processing, biomedical sciences, computer vision, data mining and machine learning [21]–[23]. The decomposition is *essentially unique* if it is unique up to trivial permutation of the rank-1 terms and scaling and counterscaling of the factors in the same rank-1 term. In general, no unique solution exists in the matrix case without additional assumptions for $R > 1$. In the higher-order case, we typically expect uniqueness under rather mild conditions. Consider a third-order tensor of rank R and size $I \times J \times K$ with factor matrices $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$. Kruskal's condition states that the CPD is unique if [24]:

$$2R + 2 \leq k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}}. \tag{2}$$

The k-rank of a matrix $\mathbf{A}$ equals the largest number $k_{\mathbf{A}}$ such that any $k_{\mathbf{A}}$ columns of $\mathbf{A}$ are linearly independent. Condition (2) is deterministic in the sense that uniqueness is guaranteed for a particular choice of factor matrices satisfying the condition. Generic uniqueness conditions consider uniqueness with probability one when the entries of the factor matrices are drawn from absolutely continuous probability density functions. For example, condition (2) implies generic uniqueness if $2R+2 \leq \min(I, R) + \min(J, R) + \min(K, R)$ as the k-rank of a generic matrix equals its smallest dimension. In general, milder conditions than Kruskal's can be obtained. Let us for instance consider the case where at least one of the tensor dimensions is not strictly smaller than R. For example, the CPD is generically unique for $\mathbb{K} = \mathbb{C}$ if [25], [26]:

$$R \leq (I-1)(J-1), \quad 3 \leq I \leq J, \quad \text{and} \quad R \leq K. \tag{3}$$

More generally, the CPD is generically unique (with a few known exceptions) if [27]:

$$R \leq \left\lceil \frac{IJK}{I+J+K-2} \right\rceil - 1 \quad \text{and} \quad IJK \leq 15000, \tag{4}$$

with $\lceil x \rceil$ the smallest integer not less than x. The bound on the number of entries IJK has only been verified numerically up to 15000 but is assumed to hold for larger number of entries as well. Condition (4) is equivalent with (3) for $R \leq K$ and $3 \leq I \leq J$.

Note that condition (4) involves the ratio between the number of entries in the tensor and the number of parameters in a rank-1 term (compensated for scaling). The condition states that the decomposition is unique with probability one if the number of entries is (strictly) larger than the number of parameters, i.e., if the tensor is (minimally) compressible.

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2016.2617858

Our working assumption to solve the large-scale BSS problem is based on this compressibility, as will be explained further. We expect even milder uniqueness conditions when $N > 3$ [28], [29]. An overview and state-of-the-art deterministic and generic uniqueness conditions for higher-order tensors are given in [25], [27], [28], [30]–[35] and references therein. For a short introduction to CPD uniqueness we refer to [23, Section IV].

Definition 2. A *block term decomposition* (BTD) of a third-order tensor $\mathcal{X} \in \mathbb{K}^{I\times J\times K}$ *in multilinear rank-*$(L_r, L_r, 1)$ *terms* for $r = 1, \ldots, R$ is a decomposition of the form:

$$\mathcal{X} = \sum_{r=1}^{R} (\mathbf{A}_r \mathbf{B}_r^{\mathrm{T}}) \circ \mathbf{c}_r, \tag{5}$$

in which $\mathbf{A}_r \in \mathbb{K}^{I\times L_r}$ and $\mathbf{B}_r \in \mathbb{K}^{J\times L_r}$ have full column rank L_r and $\mathbf{c}_r$ is nonzero.

These block terms are more general than the simple rank-1 terms of a third-order PD. Hence, they allow the modeling of more complex phenomena, see e.g., [8], [36]. Other BTDs and associated uniqueness results can be found in [5], [9], [10].

II. Large-scale blind source separation via low-rank sources

We derive a new BSS method that exploits the hypothesized compressibility of the sources. We show that this is possible by applying a particular deterministic tensorization technique to the observed data matrix. Decomposition of the resulting tensor allows us to uniquely retrieve the mixing vectors and the sources. In Subsections II-A, II-B, and II-C, we define BSS, motivate the working hypothesis, and derive our method.

A. Blind Source Separation

We use a linear and instantaneous data model for BSS [4]:

$$\mathbf{X} = \mathbf{M}\mathbf{S} + \mathbf{N}, \tag{6}$$

with $\mathbf{X} \in \mathbb{K}^{M\times K}$ and $\mathbf{S} \in \mathbb{K}^{R\times K}$ containing K samples of each of the M observed and R source signals, respectively; $\mathbf{M} \in \mathbb{K}^{M\times R}$ is the mixing matrix and $\mathbf{N} \in \mathbb{K}^{M\times K}$ is the additive noise. The goal of BSS is to retrieve the unknown mixing vectors in $\mathbf{M}$ and/or the unknown sources in $\mathbf{S}$, given only the observed data $\mathbf{X}$. In the derivation of our method we ignore the noise $\mathbf{N}$ for notational simplicity, its influence will be further investigated in Section V by means of simulations.

The proposed method reshapes each observed signal, i.e., each row of $\mathbf{X}$, into a matrix and stacks them into a third-order tensor. This is illustrated in Figure 1. If the matricized sources admit a low-rank representation, the BSS problem can be solved uniquely by decomposing the tensorized observed data. In general, we reshape each row into an Nth-order tensor and stack them into an $(N + 1)$th-order tensor. As such, the parsimonious low-rank models enable very large signal compressions, allowing one to tackle large-scale problems. In general, no unique solution to (6) exists without additional assumptions. By assuming that the source signals are low-rank signals, which can be written as sums of Kronecker products of smaller vectors, the problem can be reformulated as a tensor decomposition. As a decomposition of a higher-order tensor is unique under mild conditions as discussed in Subsection I-B, the working assumption enables a unique solution of (6) under the same conditions.

B. Low-rank sources

Many real-life signals are compressible. For example, many common types of signals can be expressed in a basis such that the coefficients decay according to a power law [37]. In a large-scale setting, the amount of information contained in the signal can often be represented by a number of parameters that is much smaller than the total number of entries because there is some structure in the data [38]. Such compressible signals can often be represented in a very compact way by a low-rank approximation of a tensor representation [15], [39]; we call them *low-rank signals*. It is this notion that is the key to our approach: it enables a unique separation of the sources and identification of the mixing vectors. Moreover, it provides a way to cope with large-scale BSS problems because of the large reduction in the number of parameters. We show that our working hypothesis holds exactly for exponential polynomials.

Consider $f(t) = az^t$ evaluated in $t = 0, 1, \ldots, 5$. The resulting vector is reshaped into a (3×2) matrix $\mathbf{S}$ of rank 1:

$$\mathbf{S} = a \begin{pmatrix} 1 & z^3 \\ z & z^4 \\ z^2 & z^5 \end{pmatrix} = a \begin{pmatrix} 1 \\ z \\ z^2 \end{pmatrix} \begin{pmatrix} 1 & z^3 \end{pmatrix}. \tag{7}$$

The (3×4) Hankelized version $\mathbf{H}$ of the same vector is [5]:

$$\mathbf{H} = a \begin{pmatrix} 1 & z & z^2 & z^3 \\ z & z^2 & z^3 & z^4 \\ z^2 & z^3 & z^4 & z^5 \end{pmatrix} = \begin{pmatrix} 1 \\ z \\ z^2 \end{pmatrix} \begin{pmatrix} 1 & z & z^2 & z^3 \end{pmatrix}. \tag{8}$$

It is well-known that if the original signal is exponential, then $\mathbf{H}$ has rank one, as illustrated. One can see that the columns of $\mathbf{S}$ are a subset of the columns of $\mathbf{H}$. Hence, if $\mathbf{H}$ has rank one, then clearly $\mathbf{S}$ also has rank one. Consider now a vector $\mathbf{f} \in \mathbb{K}^K$ defined by the underlying function $f(t)$ as $f_k = f(t_k)$, $1 \le k \le K$, using equidistant samples. We reshape $\mathbf{f}$ into a $(I \times J)$ matrix $\mathbf{S}$ such that $\mathrm{vec}(\mathbf{S}) = \mathbf{f}$ with $K = IJ$. Consider also a Hankelized version $\mathbf{H} \in \mathbb{K}^{I\times J_h}$ such that $h_{ij_h} = f_{i+j_h-1}$ with $K = I + J_h - 1$. Hence, we have that $\mathbf{S} = \mathbf{H}\mathbf{Q}$ with $\mathbf{Q} \in \mathbb{K}^{J_h\times J}$ the selection matrix defined by $\mathbf{q}_j = \mathbf{e}_{(j-1)I+1}$ for $j = 1, \ldots, J$. One can verify that the matrix $\mathbf{Q}$ selects all distinct columns of $\mathbf{H}$, by comparing, e.g., the matrices in (7) and (8). It is clear that if $\mathbf{H}$ has low rank then $\mathbf{S}$ has low rank as well, while $\mathbf{S}$ offers a more compact representation than $\mathbf{H}$. It is known that $\mathbf{H}$ has low rank if the underlying functions are sums of a limited number of exponential and trigonometric terms. This fact extends to the larger class of exponential polynomials [5]. The latter allows one to model a wide range of signals in many applications, e.g., the autonomous behavior of linear systems can be described by (complex) exponential and, if we admit coinciding poles, exponential polynomials. In Table I we show the coinciding (exact) rank values of $\mathbf{H}$ and $\mathbf{S}$ for several common (exponential) polynomials; by combining such functions one can model a wide variety of signals. For

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2016.2617858

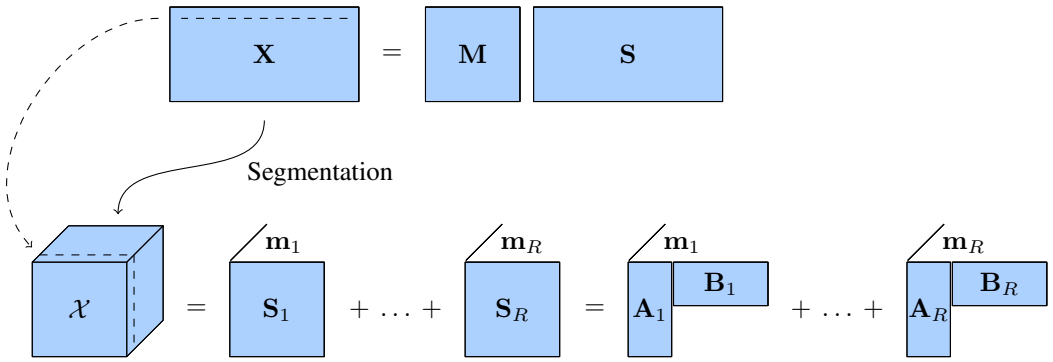


Fig. 1. Illustration of segmentation: each row of the observed data matrix $\mathbf{X}$ is reshaped into a matrix and then stacked into a tensor $\mathcal{X}$. The reshaped sources appear in the first and second mode, and the mixing vectors appear in the third mode. The BSS problem boils down to a BTD in multilinear rank-$(L_r, L_r, 1)$ terms if the reshaped sources allow a low-rank representation, enabling a unique separation of the sources and identification of the mixing vectors.

TABLE I
RANK $r(\mathbf{H})$ OF THE HANKELIZED VERSION OF SEVERAL (EXPONENTIAL) POLYNOMIALS $f(t)$. IF $\mathbf{H}$ HAS LOW RANK THEN THE $(I \times J)$ RESHAPED VERSION $\mathbf{S}$ HAS LOW RANK AS WELL (IF $R < \text{MIN}(I, J)$). THE LATTER, HOWEVER, PROVIDES A MUCH MORE COMPACT REPRESENTATION FOR $f(t)$ THAN THE FORMER. ($p_r(t)$ IS A POLYNOMIAL OF DEGREE Q_r.)

$f(t)$	$r(\mathbf{H})$	$f(t)$	$r(\mathbf{H})$
az^t	1	$\sum_{r=1}^{R} a_r z_r^t$	R
$a\sin(bt)$ $a\cos(bt)$	2	$\sum_{r=1}^{R} a_r \sin(b_r t)$	$2R$
$az^t \sin(bt)$	2	$\sum_{r=1}^{R} a_r z_r^t \sin(b_r t)$	$2R$
$p(t) = \sum_{q=0}^{Q} a_q t^q$	$Q+1$	$\sum_{r=1}^{R} p_r(t)$	$\sum_{r=1}^{R} Q_r + R$
$p(t)z^t$	$Q+1$	$\sum_{r=1}^{R} p_r(t) z_r^t$	$\sum_{r=1}^{R} Q_r + R$
$p(t)\sin(at)$	$Q+2$	$\sum_{r=1}^{R} p_r(t) \sin(a_r t)$	$\sum_{r=1}^{R} Q_r + 2R$
$p(t)z^t \sin(at)$	$Q+2$	$\sum_{r=1}^{R} p_r(t) z_r^t \sin(a_r t)$	$\sum_{r=1}^{R} Q_r + 2R$

example, a sine is a linear combination of two (complex conjugated) exponentials and, hence, admits a rank-2 model. Note that, while exponential polynomials can be represented by low-rank matrices, the latter allow the representation of a much larger family of signals than only exponential signals. Moreover, Hankel matrices are often ill-conditioned [40], so that the numerical rank can be significantly smaller than the theoretical one.

So far we have discussed signals that admit an exact low-rank representation. However, our approach also works well for more general compressible signals. A reshaped version of the latter often admits an approximate low-rank model as illustrated in Figure 2. Assume we approximate $\mathbf{S}$ by a rank-R matrix $\tilde{\mathbf{S}} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r$, then the approximation error on the original function $\mathbf{f} = \text{vec}(\mathbf{S})$ is:

$$\|\mathbf{f} - \text{vec}(\tilde{\mathbf{S}})\|_{\text{F}}^2 = \|\mathbf{f} - \sum_{r=1}^{R} \mathbf{b}_r \otimes \mathbf{a}_r\|_{\text{F}}^2. \tag{9}$$

Recall from Subsection I-A that a Kronecker product equals a vectorized outer product. We can make the approximation

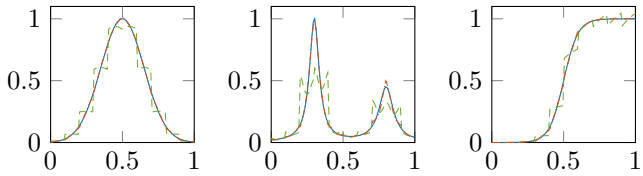


Fig. 2. Low-rank approximation of a reshaped smooth function often provides a good representation. This is illustrated for a Gaussian (left), a rational function (middle), and a sigmoid (right) sampled uniformly 100 times in $[0, 1]$ (——). The original functions are reshaped into a (10×10) matrix and then approximated by a low-rank matrix by truncating the SVD. The reconstructed functions are obtained by vectorizing this low-rank matrix. One can clearly see that the functions can be better approximated by a rank-2 (-·-·-) than a rank-1 (- - - -) approximation.

error (9) as small as desired by increasing R. Since (9) is just a vectorized version of $\|\mathbf{S} - \tilde{\mathbf{S}}\|_{\text{F}}^2$, Eckart–Young's theorem provides an upper bound on the approximation error [41]. Namely, the least-squares error on the representation of the signal $\mathbf{f}$ is the sum of the squares of the discarded singular values of $\mathbf{S}$. The singular value spectrum of $\mathbf{S}$ is often fast decaying, and hence the signal $\mathbf{f}$ often admits a good representation of the form (9) for low R. It is outside the scope of this paper to investigate in general under which conditions on the signal $\mathbf{f}$ the error in (9) is small. However, we do provide explicit bounds by focusing on signals that admit a good polynomial approximation. We emphasize that these are only bounds, as 1) polynomials are only a special case of exponential polynomials and 2) the latter are only a special case of functions that yield a low-rank matrix $\mathbf{S}$. As such, assume that we approximate the underlying function $f(t)$ of $\mathbf{f}$ with a Taylor polynomial $p(t)$ of degree $R-1$ around $t = t_*$. Assuming $f(t)$ and its derivatives up to order R are continuous, which is satisfied for smooth signals, Taylor's theorem provides the following element-wise upper bound on the error in (9):

$$|f(t) - p(t)| \leq \frac{f_{\max}}{R!} |t - t_*|^R \tag{10}$$

with $f_{\max} = \max_u f^{(R)}(u)$, $u \in (t_*, t)$ and $f^{(R)}$ the Rth derivative of f. The corresponding matrix $\tilde{\mathbf{S}}$ of $p(t)$ has rank R, see Table I; hence, (10) is a bound on the error of the rank-R approximation of $\mathbf{S}$. Signals with rapidly converging

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2016.2617858

Taylor series admit an approximate low-rank model, hence, only a small R is needed for a good approximation. A general polynomial approximation $p(t)$ in K uniformly sampled points in the interval $[a, b]$ gives the following upper bound on (9):

$$\|\mathbf{f} - \text{vec}(\tilde{\mathbf{S}})\|_{\text{F}}^2 \leq \left(\frac{h^R}{4R}\bar{f}\right)^2$$

with $h = (b-a)/K$, $\bar{f} = \max_u f^{(R)}(u)$, $u \in [a, b]$, and $f^{(R)}$ the Rth derivative of f. Similar results can be derived for other types of approximations, e.g., a polynomial approximation in Chebyshev points. In Section V, we illustrate our strategy for real-life signals as well, showing that our working hypothesis is valid for a variety of signals and applications.

In this paper we also reshape signals into higher-order tensors, going further than the Hankel strategy from [5] and enabling an even more compact representation. In tensor-based scientific computing one often reshapes a function up to a $(2 \times 2 \times \cdots \times 2)$ tensor of very high order to achieve maximal compression for a fixed rank R [15], [39]. Here, we allow much more freedom in the choice of the reshaping parameters, which enables a trade-off between the approximation error in (9) and the compression rate, see Subsection V-E.

Let us now describe the strategy more formally. Suppose one reshapes the rth source $\mathbf{s}_r$ in (6) into a $(I \times J)$ matrix $\mathbf{S}_r$ such that $\text{vec}(\mathbf{S}_r) = \mathbf{s}_r$ with $K = IJ$. Note that this is the same as stacking different decimated versions of the signal in the rows of a matrix. If the rth reshaped (or matricized) source $\mathbf{S}_r$ admits a rank-1 representation, which is our working hypothesis, we have that $\mathbf{S}_r = \mathbf{a}_r \circ \mathbf{b}_r$ with $\mathbf{a}_r \in \mathbb{K}^I$ and $\mathbf{b}_r \in \mathbb{K}^J$, as, e.g., in (7). In general, however, this model is too restrictive. The reshaped sources may admit, or better be approximated by a low-rank representation, as is, e.g., the case for a sine and the functions in Figure 2, respectively. Hence, we have that $\mathbf{S}_r = \sum_{l_r=1}^{L_r} \mathbf{a}_{l_r r} \circ \mathbf{b}_{l_r r}$. Note that this means that we assume that the sources can be written as a sum of Kronecker products: $\mathbf{s}_r = \text{vec}\,(\mathbf{S}_r) = \sum_{l_r=1}^{L_r} \mathbf{b}_{l_r r} \otimes \mathbf{a}_{l_r r}$. This strategy enables a compact representation of the sources, see Table II. Indeed, the number of parameters is one order of magnitude lower than the finite length K if $I \approx J$.

More generally, we can reshape the sources into a higher-order tensor, enabling a more compact representation. Suppose we reshape the rth source $\mathbf{s}_r$ into an Nth-order tensor $\mathcal{S}_r \in \mathbb{K}^{I_1 \times I_2 \times \cdots \times I_N}$ such that $\text{vec}(\mathcal{S}_r) = \mathbf{s}_r$ with $K = \prod_{n=1}^N I_n$. If the rth reshaped (or tensorized) source $\mathcal{S}_r$ admits a (higher-order) low-rank representation, we have that:

$$\mathcal{S}_r = \sum_{l_r=1}^{L_r} \mathbf{u}_{l_r r}^{(1)} \circ \mathbf{u}_{l_r r}^{(2)} \circ \cdots \circ \mathbf{u}_{l_r r}^{(N)}, \tag{11}$$

in which $\mathbf{u}_{l_r r}^{(n)} \in \mathbb{K}^{I_n}$ for $n = 1, \ldots, N$, where the number of rank-1 terms L_r can differ between sources. Note that this is a PD as in (1). This means that the sources can be modeled, or approximated, by sums of $(N-1)$ Kronecker products [15]:

$$\mathbf{s}_r = \text{vec}\,(\mathcal{S}_r) = \sum_{l_r=1}^{L_r} \mathbf{u}_{l_r r}^{(N)} \otimes \mathbf{u}_{l_r r}^{(N-1)} \otimes \cdots \otimes \mathbf{u}_{l_r r}^{(1)}, \tag{12}$$

In general, the number of parameter decreases logarithmically in the number of Kronecker products N (i.e., the order of the representation) and increases proportionally with the number of rank-1 terms L_r, see Table II. For example, if $I_n = I$ for $n = 1, \ldots, 3$, then $K = I^3$ and only $O(3L_r I)$ parameters are needed. The possibly large compressions indicate the applicability of this strategy for large-scale BSS problems.

TABLE II
RESHAPING $\mathbf{s}_r$ IN (6) INTO $\mathcal{S}_r \in \mathbb{K}^{I_1 \times I_2 \times \cdots \times I_N}$ AND THEN USING A RANK-L_r REPRESENTATION LEADS TO A CONSIDERABLE COMPRESSION. IF $N = 2$, WE USE I AND J. THE NUMBER OF PARAMETERS DECREASES LOGARITHMICALLY IN N AND INCREASES PROPORTIONALLY WITH L_r.

	K	for general I_n	$I_n \approx I$, for all n
$N = 2$	IJ	$L_r(I + J - 1)$	$O(L_r I)$
$N > 2$	$\prod_{n=1}^N I_n$	$L_r(\sum_{n=1}^N I_n - N + 1)$	$O(L_r N I)$

C. Decomposition

We now demonstrate how the BSS problem in (6) can be reformulated as the computation of a tensor decomposition when the sources admit a low-rank representation. Let us start as follows: each row of $\mathbf{X}$ is reshaped into a $(I \times J)$ matrix as described earlier and then stacked into a third-order tensor $\mathcal{X} \in \mathbb{K}^{I \times J \times M}$ such that $\text{vec}\,(\mathbf{X}_m) = \mathbf{x}_m$. In other words, the mth matricized observed signal is equal to the mth frontal slice of $\mathcal{X}$. Since the tensorization is a linear operation, the M reshaped observed signals are linear combinations of the R reshaped sources $\mathbf{S}_r \in \mathbb{K}^{I \times J}$. As such, we have that:

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{S}_r \circ \mathbf{m}_r. \tag{13}$$

We denote this deterministic tensorization technique by *segmentation*; see Figure 1 for an illustration. Now assume that the rth reshaped source in (13) admits a rank-1 representation, i.e., $\mathbf{S}_r = \mathbf{a}_r \circ \mathbf{b}_r$ for $r = 1, \ldots, R$, then we have that:

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

Equation (14) is a CPD as defined in (1). Consequently, the BSS problem boils down to the computation of a CPD of a third-order tensor in R rank-1 terms. Analogously, if the reshaped sources admit a low-rank representation, the BSS problem boils down to a BTD in multilinear rank-$(L_r, L_r, 1)$ terms, as in (5) and illustrated in Figure 1. References to uniqueness results for both cases have been mentioned in Subsection I-B. We insist that the compressibility of the sources has enabled their blind separation.

More generally, we can reshape each observed signal into a $(I_1 \times I_2 \times \cdots \times I_N)$ Nth-order tensor as described earlier and then stack it into a $(N+1)$th-order tensor $\mathcal{X} \in \mathbb{K}^{I_1 \times I_2 \times \cdots \times I_N \times M}$. As such, the mth tensorized observed signal is equal to the mth Nth-order "frontal slice" of $\mathcal{X}$:

$$\mathcal{X} = \sum_{r=1}^{R} \mathcal{S}_r \circ \mathbf{m}_r, \tag{15}$$

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2016.2617858

If the reshaped sources $\mathcal{S}_r \in \mathbb{K}^{I_1 \times I_2 \times \cdots \times I_N}$ allow a low-rank representation as in (11), we have:

$$\mathcal{X} = \sum_{r=1}^{R} \left(\sum_{l_r=1}^{L_r} \mathbf{u}_{l_r r}^{(1)} \circ \mathbf{u}_{l_r r}^{(2)} \circ \cdots \circ \mathbf{u}_{l_r r}^{(N)} \right) \circ \mathbf{m}_r, \qquad (16)$$

which is a decomposition in R (rank-$L_r \circ$ vector) terms [42]. It is a more general decomposition because it boils down to a CPD of a higher-order tensor as in (1) if $L_r = 1$ for all r. Also, it boils down to a BTD in multilinear rank-$(L_r, L_r, 1)$ terms as in (5) if $N = 2$, i.e., if $\mathcal{X}$ is a third-order tensor. In that case, the factor matrices $\mathbf{U}_r^{(1)}$ and $\mathbf{U}_r^{(2)}$ of the rth term are not unique, but their products are (up to scaling and permutation). On the other hand, for $N > 2$, the factor matrices $\mathbf{U}_r^{(n)}$ are unique under mild conditions because they form a rank-L_r PD of an Nth-order tensor. We will exploit this in the DOA estimation application in Subsection V-G.

The proposed method simultaneously determines both the mixing vectors *and* the sources by 1) simply reshaping the data (using segmentation) and 2) exploiting the fact that many real-life signals admit a (higher-order) low-rank representation. As such, the BSS problem boils down to a tensor decomposition and 3) we can benefit from mild uniqueness properties. Moreover, 4) it is applicable for large-scale BSS problems, i.e., large K, as is clear from the possibly huge compressions as indicated above. However, this is not necessarily a significant advantage compared to existing methods like ICA. The latter has only a linear dependence on K and even benefits from large K accuracy-wise because the K samples are used to estimate statistics. Finally, 5) the method is deterministic, meaning that it does not use (higher-order) statistics, hence, it also works well if the number of samples is small and/or if the sources are not statistically independent. This is a difference with statistical methods such as ICA.

III. Large-scale blind source separation via low-rank mixing vectors

In the previous section we exploited the fact that many real-life (source) signals admit a low-rank representation. This is also a natural assumption for the mixing vectors if one considers, e.g., many sensors and/or high sensor density; we call them *low-rank mixing vectors* analogous to low-rank sources. Such problems arise in biomedical sciences, e.g., wireless body area networks (WBANs) using electroencephalography (EEG) [18] and electrocorticography (ECoG) [43] with high spatial resolution, or neural dust with thousands of miniature sensors (neural probes) dispersed throughout the brain [44]. Moreover, one often encounters mixing matrices with Vandermonde structure [45], i.e., each reshaped mixing vector has exactly rank one. An example are uniform linear (ULAs) and rectangular arrays (URAs) with far-field sources that emit narrowband signals [46]–[48]. Here, we also see a trend towards large-scale antennas, also known as massive MIMO [19], [49]. If the signals propagate through several distinct paths, e.g., due to reflections or scattering [50], each reshaped mixing vector has low rank. If the sources are located in the near-field, the Vandermonde structure is only approximate which can be accommodated by a low-rank approximation.

Exploitation of the underlying compactness of such low-rank mixing vectors amounts to a comparable method as in Section II, which has been briefly addressed in [20]. Let us illustrate the analogy with the previous section more clearly: each *column* (cf. above) of $\mathbf{X}$ is reshaped into a $(I \times J)$ matrix with $M = IJ$ and then stacked into a third-order tensor $\mathcal{X} \in \mathbb{K}^{I \times J \times K}$. Next, assume the reshaped mixing vectors admit a rank-1 representation, which is our working hypothesis, i.e., $\mathbf{M}_r = \text{unvec}(\mathbf{m}_r) = \mathbf{a}_r \circ \mathbf{b}_r$ for $r = 1, \ldots, R$. Hence,

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{s}_r. \qquad (17)$$

Note that this boils down to applying the same strategy as before on the transposed observed data matrix. The generalization to higher-order low-rank representations is straightforward. The same analysis as in Subsection II-C applies, but now we segment the mixing vectors and exploit the fact that they possibly admit a (higher-order) low-rank representation. Moreover, the method has several advantages over ICA: ICA methods based on (full) HOS are infeasible when M is large as the number of entries in Qth-order statistics is $O(M^Q)$. Also, our method can handle Gaussian random sources in contrast to ICA (if the mixing vectors indeed exhibit some low-rank structure) [4]. Finally, the method imposes only mild conditions (via the uniqueness conditions) on the sources in contrast to existing methods, e.g., linear independence instead of statistical independence as in ICA.

IV. Large-scale blind source separation using twofold segmentation

In the previous two sections we either reshaped the sources *or* the mixing vectors and then exploited the hypothesized low-rank structure. However, as we have illustrated before, both the mixing vectors *and* the sources may admit such a higher-order low-rank representation. Hence, a natural extension is to use both strategies simultaneously. For instance, one often has sinusoidal sources, which admit a rank-2 representation, in ULAs of which the Vandermonde mixing vectors admit a rank-1 representation. To the best of our knowledge, this is the first time that tensorization is used on both levels of the BSS problem and more generally in matrix factorization.

By exploiting the underlying compactness on both levels, we are again able to reformulate the BSS problem as the computation of a tensor decomposition. Let us start with reshaping each column of $\mathbf{X}$ into a $(I_1 \times I_2)$ matrix with $M = I_1 I_2$ and stacking them in an intermediate third-order tensor $\mathcal{Y} \in \mathbb{K}^{I_1 \times I_2 \times K}$. Note that the (i_1, i_2)th mode-3 vector of $\mathcal{Y}$ equals the $(i_1 + (i_2 - 1)I_1)$th row of $\mathbf{X}$. Each mode-3 vector of $\mathcal{Y}$ (i.e., row of $\mathbf{X}$) is subsequently reshaped into a $(J_1 \times J_2)$ matrix with $K = J_1 J_2$, which overall yields a fourth-order tensor $\mathcal{X} \in \mathbb{K}^{I_1 \times I_2 \times J_1 \times J_2}$. Hence, we have that:

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{M}_r \circ \mathbf{S}_r. \qquad (18)$$

We denote this by *twofold segmentation* (cf. Sections II and III). Let us now assume that both the reshaped mixing vectors and sources admit a rank-1 representation. In that case,

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2016.2617858

it is easy to see that (18) is a CPD of a fourth-order tensor in R rank-1 terms. More generally, if the segmented mixing vectors and sources allow a low-rank representation, we have:

$$\mathcal{X} = \sum_{r=1}^{R} (\mathbf{A}_r \mathbf{B}_r^{\mathrm{T}}) \circ (\mathbf{C}_r \mathbf{D}_r^{\mathrm{T}}), \tag{19}$$

in which $\mathbf{A}_r \in \mathbb{K}^{I_1 \times L_r}$ and $\mathbf{B}_r \in \mathbb{K}^{I_2 \times L_r}$ have full column rank L_r and $\mathbf{C}_r \in \mathbb{K}^{J_1 \times P_r}$ and $\mathbf{D}_r \in \mathbb{K}^{J_2 \times P_r}$ have full column rank P_r. Note that the ranks L_r and P_r can be different for each r and do not necessarily have the same value inside the rth term. This is a new kind of decomposition: $\mathcal{X}$ is decomposed in a sum of R (rank-$L_r \circ$ rank-P_r) terms.

More generally, we can reshape each row and column of $\mathbf{X}$ into $\mathcal{S}_r \in \mathbb{K}^{J_1 \times J_2 \times \cdots \times J_{N_s}}$ and $\mathcal{M}_r \in \mathbb{K}^{I_1 \times I_2 \times \cdots \times I_{N_m}}$ such that $\mathrm{vec}(\mathcal{M}_r) = \mathbf{m}_r$ and $\mathrm{vec}(\mathcal{S}_r) = \mathbf{s}_r$, respectively, with $M = \prod_{n_m=1}^{N_m} I_{n_m}$ and $K = \prod_{n_s=1}^{N_s} J_{n_s}$, analogous to the single segmentation case in (15). As such, we have that:

$$\mathcal{X} = \sum_{r=1}^{R} \mathcal{M}_r \circ \mathcal{S}_r.$$

Analogous to (16), the reshaped mixing vectors and sources can both admit a low-rank representation. Hence, we have that:

$$\mathcal{X} = \sum_{r=1}^{R} \left(\sum_{l_r=1}^{L_r} \circ_{n_m=1}^{N_m} \mathbf{u}_{l_r r}^{(n_m)} \right) \circ \left(\sum_{p_r=1}^{P_r} \circ_{n_s=1}^{N_s} \mathbf{v}_{l_r r}^{(n_s)} \right),$$

in which $\mathbf{u}_{l_r r}^{(n_m)} \in \mathbb{K}^{I_{n_m}}$ and $\mathbf{v}_{p_r r}^{(n_s)} \in \mathbb{K}^{J_{n_s}}$. In comparison with (19), the block factors $\mathbf{U}_r^{(n)}$ and/or $\mathbf{V}_r^{(n)}$ are unique under mild conditions if $N_m > 2$ and/or $N_s > 2$. The reason is the same as for the single segmentation case, see Subsection II-C.

The proposed method offers 1) a framework to exploit the low-rank structure of both the reshaped mixing vectors and sources; the same analysis as in the previous sections applies. Again, we reformulate the BSS problem as the computation of a tensor decomposition, hence, 2) we can benefit from the mild uniqueness properties. More specifically, it boils down to the computation of a new and more general decomposition. As such, 3) the method is applicable in a big data setting: it can handle both large sample sizes and large numbers of sensors efficiently, see Table II. Furthermore, 4) the method is deterministic, hence, it is not needed *per se* to have a large number of samples. Finally, 5) only mild, and natural, assumptions are imposed on the mixing vectors and the sources. We simply exploit the low-rank structure which is often present in real-life signals as explained above.

V. Simulations and Applications

In Subsection V-A, we give an example of the separation of two low-rank sources and the separation of two low-rank sources that are mixed with low-rank mixing vectors. Subsection V-B demonstrates the separation of more sources than observed signals. We investigate the influence of the noise and the sample size in Subsection V-C. Subsection V-D shows how well one can approximate the reshaped mixing vectors and/or sources for varying rank and SNR. In Subsection V-E we analyze the influence of the choice of reshaping dimensions. Finally, in the last two subsections, we illustrate the proposed methods with fetal electrocardiogram extraction and direction-of-arrival estimation in large-scale uniform linear arrays.

We use the `segmentize` command from Tensorlab to apply segmentation to the observed data matrices [51]. The CPD and BTD in multilinear rank-$(L_r, L_r, 1)$ terms can typically be computed algebraically by means of a generalized eigenvalue decomposition [10], [42], [52]. The algebraic solution is exact in the noiseless case and a good initialization for optimization-based methods in the noisy case. In this paper, we use least squares optimization-based algorithms `cpd` and `lll` to fit the decomposition to the data until a sufficiently high accuracy is attained. During the computation, it is theoretically possible that degeneracy occurs [53], [54]. For example, the magnitude of some terms grows without bounds but with opposite sign, resulting in a poor solution but a good fit. Degeneracy can be avoided in several ways such as increasing the number of rank-1 terms or imposing orthogonality or non-negative constraints on the factor matrices [22], [54]–[56]. The decompositions in (rank-$L_r \circ$ vector) and (rank-$L_r \circ$ rank-P_r) terms are computed with two adapted versions of `cpd_nls` called `lvec_nls` and `lp_nls`, respectively, and are available upon request. For very large tensors, one can resort to large-scale algorithms as described in [17], [57], [58].

The mixing vectors and sources can only be determined up to scaling and permutation, i.e., the standard indeterminacies in BSS. Hence, in order to compute the error they are first optimally scaled and permuted with respect to the true ones. The relative error is then defined as the relative difference in Frobenius norm, i.e., we have relative error $\epsilon_{\mathbf{A}} = ||\mathbf{A} - \hat{\mathbf{A}}||_{\mathrm{F}} / ||\mathbf{A}||_{\mathrm{F}}$ with $\hat{\mathbf{A}}$ an optimally scaled and permuted estimate of $\mathbf{A}$. We use Gaussian additive noise unless indicated otherwise.

A. General experiments

First, we illustrate the method proposed in Section II. Consider $R = 2$ low-rank sources: $s_1(t) = e^{-t}$ and $s_2(t) = \sin(4\pi t)$ with $K = 4096$ equidistant samples in $[0, 1]$. They are mixed into $M = 3$ observed signals using $\mathbf{M} = [0.5,\ 2;\ 2,\ -3;\ 1,\ 0.5]$. We use a second-order ($N = 2$) rank-1 ($L_1 = 1$) and rank-2 ($L_2 = 2$) approximation for the first and second source, respectively, with $I = J = 64$. Note that the approximation of the first and second source requires only 127 and 254 values, respectively, see Table II. This is the maximal reduction for a second-order approximation. Namely, we have a compression of $1 - L_r \frac{I+J-N+1}{M}$, i.e., 96.90% and 93.80% for the first and second source, respectively. The perfectly recovered sources are shown in Figure 3.

Second, we illustrate the method proposed in Section IV. Consider $R = 2$ low-rank sources: $s_1(t) = e^{-t} + e^{t} - e^{0.5t}$ and $s_2(t) = 2e^{-t}$ with $K = 4096$ equidistant samples in $[0, 1]$. The sources are mixed with two low-rank mixing vectors: $m_1(\xi) = \sin(2\pi\xi)$ and $m_2(\xi) = e^{-2\xi}\sin(6\pi\xi)$ with $M = 4096$ equidistant samples in $[0, 1]$. We use a third-order ($N_s = 3$) rank-3 ($P_1 = 3$) and rank-1 ($P_2 = 1$) approximation for the first and second source, respectively, with $J_1 = J_2 = J_3 = 16$. Furthermore, we use a second-order ($N_m = 2$) rank-2 approximation for both mixing vectors

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2016.2617858

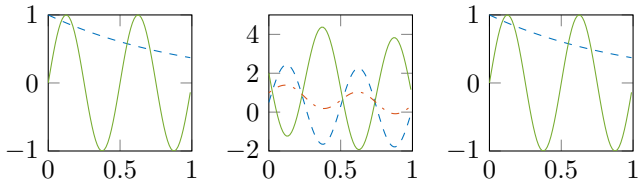


Fig. 3. Results for the first experiment of Subsection V-A. Left: the two original sources. Middle: the observed signals. Right: the recovered sources.

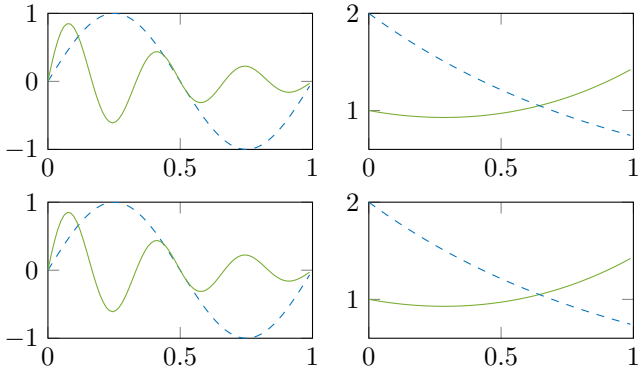


Fig. 4. Results for the second experiment of Subsection V-A. Top: original mixing vectors (left) and sources (right). Bottom: perfectly recovered mixing vectors (left) and sources (right).

$(L_1 = L_2 = 2)$ with a non-optimal choice of the segmentation parameters: $I_1 = 128$ and $I_2 = 32$. Hence, we decompose the $(128 \times 32 \times 16 \times 16 \times 16)$ segmented version of $\mathbf{X}$ into a sum of a (rank-2 $\circ$ rank-3) and a (rank-2 $\circ$ rank-1) term. The approximation of the rth mixing vector requires only $L_r(I_1 + I_2 - N_m + 1)$ values, i.e., a compression of $1 - L_r \frac{I_1+I_2-N_m+1}{M} = 92.19\%$, although this is not the maximal compression. Higher compression can be attained by increasing the order. For instance, the approximation of the rth source consists of only $P_r(J_1 + J_2 + J_3 - N_s + 1)$ values, i.e., a compression of $1 - P_r \frac{J_1+J_2+J_3-N_s+1}{M}$. Specifically, we have a compression of 96.63% and 98.88% for the first and second source, respectively. We further investigate the choice of I_{n_m} and J_{n_s} in Subsection V-E. The perfectly recovered factors are shown in Figure 4.

B. Underdetermined mixture

We illustrate the separation of more sources than observed signals. Consider $R = 3$ complex exponential source signals $s_r(t) = e^{2\pi i r t}$ for $r = 1, \ldots, R$ which are mixed into $M = 2$ observed signals using $\mathbf{M} = [-1,\ 0.5,\ 2;\ 0.5,\ 1,\ 0.5]$. We take $K = 4096$ uniformly discretized samples in $[0, 1]$. We use a second-order $(N = 2)$ rank-1 approximation for both sources with $I_1 = I_2 = 64$. The real part of the recovered sources is shown in Figure 5: perfect reconstruction is obtained.

C. Noise and sample length

First, we investigate the influence of the noise and the sample size K for the method of Section III. Consider a setup in which we have $M = 4096$ sensors and $R = 2$

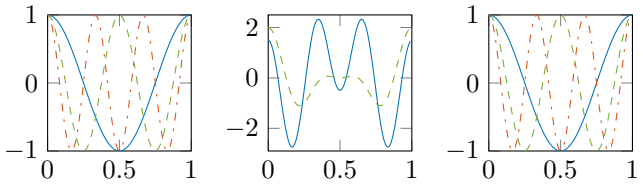


Fig. 5. Results for the underdetermined mixture. The real part of the three original sources (left), the two observed signals (middle), and the recovered sources (right).

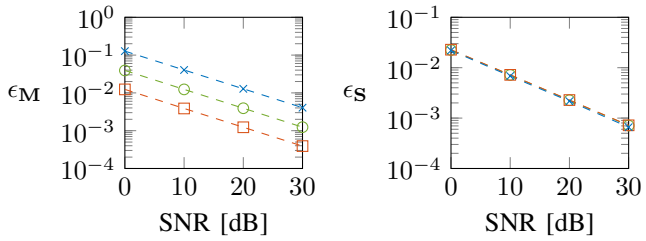


Fig. 6. Median across 100 experiments of the relative error on the mixing vectors (left) and the sources (right) as a function of SNR for $K = 10^1$ (- × -), 10^2 (- ○ -), and 10^3 (- □ -). The mixing vectors are well conditioned.

i.i.d. zero-mean unit-variance Gaussian random sources of length $K = \{10^1, 10^2, 10^3\}$. We construct the low-rank mixing vectors as the vectorization of a second-order $(N = 2)$ rank-2 $(L_1 = 2)$ and rank-3 $(L_2 = 3)$ tensor using (12) with zero-mean unit-variance Gaussian random factor vectors and $I = J = 64$. Hence, we use a second-order rank-2 and rank-3 approximation with $I = J = 64$, respectively. In Figure 6, we report the relative error on the mixing vectors $\epsilon_{\mathbf{M}}$ and the sources $\epsilon_{\mathbf{S}}$; note that the results are very accurate in comparison with the SNR. Although the method is deterministic, it is beneficial to increase K under noisy conditions. However, K can be (very) low in comparison to typical values in ICA. (Note that in this particular example, ICA cannot be used since the sources are Gaussian.) $\epsilon_{\mathbf{S}}$ does not improve for increasing K because one also has to estimate longer source signals. Similar results can be obtained for the method of Section II when increasing the number of sensors M under noisy conditions.

Next, consider a similar setup as in the previous experiment but now with the following rank-1 mixing vectors: $m_1(\xi) = e^{0.5\xi}$ and $m_2(\xi) = e^{-2\xi}$ with $\xi \in [0, 1]$. We use a second-order $(N = 2)$ rank-1 approximation for both mixing vectors $(L_1 = L_2 = 1)$ with $I = J = 64$. The results are shown in Figure 7: in comparison with Figure 6, there is some loss of accuracy on the mixing vectors and much clearer on the sources. This is due to the condition of the problem: in the previous experiment, the mixing vectors are approximately orthogonal and have about the same size $(\|\mathbf{m}_1\|/\|\mathbf{m}_2\| \approx 0.8)$, while now the angle is 37.11° and $\|\mathbf{m}_1\|/\|\mathbf{m}_2\| = 2.65$. Hence, the computation of the decomposition is more difficult and the estimates less accurate.

D. Low-rank approximation

We investigate the influence of deviations from a second-order rank-1 structure on the relative error as follows. De-

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2016.2617858

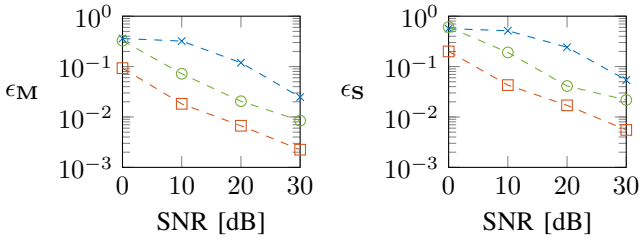


Fig. 7. Median across 100 experiments of the relative error on the mixing vectors (left) and the sources (right) as a function of SNR for $K = 10^1$ (- ×-), 10^2 (- ○-), and 10^3 (- □-). The mixing vectors are ill-conditioned.

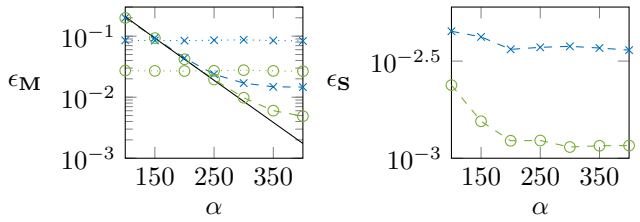


Fig. 8. Median across 100 experiments of the relative error on the mixing vectors (left), extracted from (17) (dashed) and computed via the inverse of $\hat{\mathbf{S}}$ (dotted), and the sources (right) for varying rank-1-ness α and an SNR of 15 dB (cross) and 25 dB (circle). The error bound given by the Eckart-Young theorem is shown in solid.

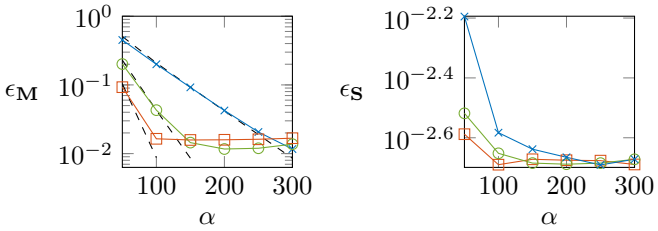


Fig. 9. Median across 100 experiments of the relative error on the mixing vectors (left) and the sources (right) for varying rank-1-ness α of the mixing vectors and 20 dB SNR with $L_1 = L_2 = 1$ (—×—), $L_1 = L_2 = 2$ (—○—), and $L_1 = L_2 = 3$ (—□—). The error estimate given by the Eckart-Young theorem is shown with dashed lines. Note the small scale of $\epsilon_{\mathbf{S}}$ (right).

fine each mixing vector as the vectorization of a random matrix with exponentially decaying singular values, i.e., $\mathbf{m_r} = \text{vec}\left(\mathbf{U_r}\text{diag}\left(\boldsymbol{\sigma}\right)\mathbf{V_r}\right)$ with $\boldsymbol{\sigma} = e^{-\alpha\boldsymbol{\xi}}$ and $\boldsymbol{\xi}$ a vector containing $\min\left(I, J\right)$ equidistant samples in $[0, 1]$. $\mathbf{U}_r$ and $\mathbf{V}_r$ are random orthogonal matrices of compatible dimensions. The exponential decay of the singular values is controlled with α which is a measure for the rank-1-ness of the mixing vectors: increasing α leads to more rank-1-like mixing vectors and vice-versa. We take $R = 2$ i.i.d. zero-mean unit-variance Gaussian random sources of length $K = 10$ and use a second-order ($N = 2$) rank-L_r approximation with $I = J = 64$.

Figure 8 shows the relative errors $\epsilon_{\mathbf{M}}$ and $\epsilon_{\mathbf{S}}$ as a function of α for an SNR of 15 dB and 25 dB using $L_1 = L_2 = 1$. Note that an estimate of the mixing matrix $\hat{\mathbf{M}}$ can be obtained from the decomposition, i.e., from (17) for this particular case, in the way explained above. However, one can also estimate it via the noisy observed data matrix and the pseudo-inverse of the estimated source matrix: $\hat{\mathbf{M}} = \mathbf{X}\hat{\mathbf{S}}^{\dagger}$. The figure illustrates that $\epsilon_{\mathbf{M}}$ decreases for increasing α until it stagnates due to noise. One can also see that, for large α, $\hat{\mathbf{M}}$ computed via the pseudo-inverse is less accurate than directly extracting $\hat{\mathbf{M}}$ from (17) and imposing rank-1 structure. However, for small α, the opposite is true. Indeed, for decreasing α, the mixing vectors become less rank-1 like and our rank-1 model cannot attain a better estimate than the one given by Eckart–Young's theorem [41]. Also, note that the sources are estimated more accurately than the mixing vectors: the noise on the sources is more averaged out because this factor is much shorter in the decomposition ($K \ll I, J$) [59].

Figure 9 shows the relative errors for several choices of L_r. One can observe that for increasing L_r, the relative error decreases in the case of small α, i.e., in the case of little rank-1-like mixing vectors. On the other hand, little is lost through overmodeling (i.e., choosing L_r too large) for large α. In fact, we overmodel less than conventional methods as we exploit the low-rank structure. Hence, the choice of L_r is not so critical, see [5], [6]. In this case one also knows that the multilinear rank of $\mathcal{X}$ is bounded by $(\sum_{r=1}^{R} L_r, \sum_{r=1}^{R} L_r, R)$.

E. Compression versus accuracy

We investigate the trade-off between compression and accuracy which will lead to a better understanding on how to choose the segmentation parameters I_{n_m} and/or I_{n_s}. We do this by examining the accuracy of a low-rank approximation of various segmentations of a real-life EEG signal with a sample rate of 500 Hz. More precisely, we reshape the EEG signal of length $K = 2^{14}$ into a $(I \times J)$ matrix with $I = 2^q$ and vary $q = 2, \ldots, 12$, then $J = 2^{14-q}$ such that $K = IJ$. Subsequently, we approximate the reshaped signal with a rank-L model with $L = \{1, 2, 3\}$.

In Figure 10, we plot the normalized number of parameters $\hat{K} = L(I + J)/K$ versus the relative error ϵ of the rank-L approximation. We see a clear trade-off between compression and accuracy, hence, what is considered a "good" choice of parameters will depend on the needs in a particular application. First of all, the curves are not symmetric since segmentation is not symmetric in the modes that it creates. Note that one can easily improve the accuracy without affecting the compression rate by switching the values of I and J such that $I < J$ rather than $I > J$ for the same rank. For fixed I and J, increasing the rank can greatly improve the accuracy, e.g., when $I \ll J$ (left part of Figure 10). The original signal and two particular approximations are shown in Figure 11. Note the relative error decreased from 0.68 to 0.096 by taking $I < J$ and increasing L for the second approximation. On the other hand, the compression reduced from 96.88% to 86.72%.

In general, a good choice of the parameters will depend on the application. If compression is the objective, one should choose $I \approx J$ and L not too large. If, on the other hand, accuracy is the objective, one can try other choices of I and J and maybe a higher rank L. In practice, one can try a particular choice of parameters, perform a similar analysis as here on the estimated sources, and further refine the choice from there.

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2016.2617858

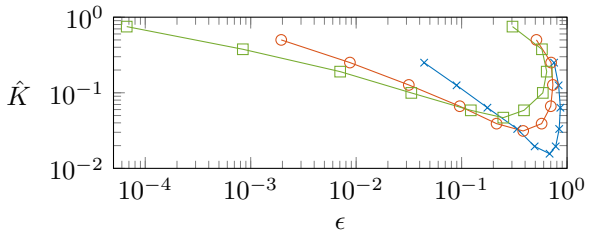


Fig. 10. Normalized number of parameters $\hat{K}$ as a function of the relative error of a rank-1 (– × –), rank-2 (– ○ –), and rank-3 (– □ –) approximation of a segmented real-life EEG signal of length $K = 2^{14}$. The signal is reshaped into a $(I \times J)$ matrix with $I = 2^q$ and $J = 2^{14-q}$ such that $K = IJ$ with $q = 2, \ldots, 12$ and q increasing from left to right on the curve.

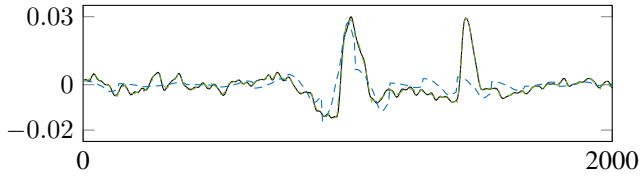


Fig. 11. Visualization of the original EEG signal (——) and two approximations. The latter are obtained by first reshaping the original signal into a $(2^7 \times 2^7)$ and $(2^5 \times 2^9)$ matrix, respectively, and then approximating them by a rank-1 (– – – –) and rank-2 (–·–·–) matrix by truncating the singular value decomposition. The reconstructed signals are obtained by vectorizing these low-rank matrices. Only the first 2000 samples are shown. The rank-2 approximation is much better than the rank-1 as is also clear from Figure 10.

F. Fetal electrocardiogram extraction

We use the method of Section II for the extraction of the antepartum fetal electrocardiogram (FECG) from multi-lead cutaneous (i.e., recorded on the mother's skin) potential recordings. The FECG is important for analyzing the health and condition of the fetus. The elimination of the mother's dominant heartbeat in the ECG can be seen as a BSS problem and one can use methods such as ICA [60]. ICA, however, falls short when only a few samples or heartbeats are available. FECG extraction is not a large-scale problem, but it is useful to illustrate a few features of our approach. Our method is applicable here because the typical QRS complexes in the ECG admit a low-rank approximation. In other words, we show that representability by a small number of parameters can be used as a ground for blind ECG signal separation. We illustrate our method for a real-life dataset.

The dataset contains eight observed signals, of which five abdominal and three thoracic; the dataset is available from DaISY[1]. Data acquisition and preprocessing is described in [61]. The sampling rate is 250 Hz. We only use the first 500 samples and scale each signal to unit norm. Each observed signal is segmented into a (25×20) matrix and the overall data set is stacked into a $(25 \times 20 \times 8)$ tensor. We use a rank-5 approximation for each source ($L_1 = L_2 = L_3 = L = 5$). At least three sources are needed to extract the FECG; this is also the case for ICA [60]. We use this particular segmentation as to maximize the compression which is only an arbitrary choice. We determined L by a trial-and-error approach starting from a rank-10 approximation and then decreasing L. Little is lost by choosing a larger L anyway, see Subsection V-D. Figure 12 shows two recovered sources. One can verify that the heartbeats of the fetus are no longer visible in the ECG of the mother and vice versa, i.e, we have a clear separation. The frequency of the FECG is typically twice as high as the frequency of the MECG, which can be observed as well.

[1]Available from http://homes.esat.kuleuven.be/~smc/daisy/daisydata.html

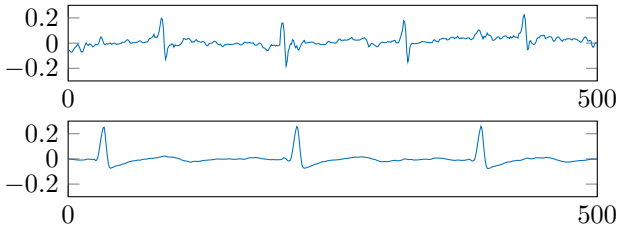


Fig. 12. Visualization of the two recovered sources in the FECG experiment. Notice a clear separation of the fetal (above) and maternal (below) ECG.

G. Direction of arrival estimation

We use the method of Section IV for direction-of-arrival (DOA) estimation of signals impinging on a ULA. Applications include radar, sonar, wireless communications, and seismic exploration. Recently, there has been a trend towards large-scale array processing [19]. Our method is able to cope with a large number of sensors, where other methods fall short. We compare our results with two well-known DOA estimation methods, MUSIC and ESPRIT, in several scenarios [46].

Consider a ULA that consists of M uniformly spaced and omnidirectional antennas receiving signals from R narrowband sources located in the far field. In that case, the problem can be described by (6) with the mixing vectors defined element-wise as $m_{mr} = \theta_r^{m-1}$ with $\theta_r = e^{-2\pi i \Delta \sin(\alpha_r)\lambda^{-1}}$. Δ is the inter-element spacing, the angle α_r to the normal is the rth DOA (i.e., $-90^\circ \leq \alpha_r \leq 90^\circ$), and λ denotes the wavelength. Note that the mixing vectors are Vandermonde vectors: $\mathbf{m}_r = \begin{bmatrix}1 & \theta_r & \theta_r^2 & \cdots & \theta_r^{M-1}\end{bmatrix}^{\mathrm{T}}$, hence, they admit a rank-1 representation [29], [48]. In a multipath setting, the mixing vectors are defined element-wise as $m_{mr} = \sum_{l=1}^{L_r} \theta_{lr}^{m-1}$ (ignoring path losses for simplicity), with L_r the number of paths for the rth source, admitting a low-rank representation. If the sources are located in the near field, the mixing vectors no longer admit a rank-1 representation but can still be well approximated by a low-rank model. If one also uses low-rank source models, we can use the method of Section IV.

First, consider a ULA with $M = 64$ sensors and Δ equal to halve the wavelength. Although our method is applicable for a large number of sensors, we choose M rather small so we can compare with MUSIC and ESPRIT. The latter two methods have to compute a $M \times M$ covariance matrix and then apply an eigenvalue decomposition (EVD). These steps can be computationally expensive because they have a complexity of $\mathcal{O}(M^2K)$ and $\mathcal{O}(M^3)$, respectively, rendering such methods infeasible for large M and K. Moreover, MUSIC has to evaluate the MUSIC spectrum for many angles in order to estimate the DOAs accurately. Here, we evaluated the MUSIC spectrum in 10^4 equidistant angles in $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Note that the number of evaluation points bounds the attainable accuracy. Consider

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2016.2617858

$R = 2$ low-rank sources: $s_r(t) = \sin(10\pi rt)$ with $K = 1024$ equidistant samples in $[0, 1]$. The sources are in line-of-sight and impinge on the ULA with $\alpha_{11} = 32°$ and $\alpha_{12} = 34°$. We use a second-order ($N_s = 2$) rank-2 ($P_1 = P_2 = 2$) approximation for both sources with $J_1 = J_2 = 32$ and a second-order ($N_m = 2$) rank-1 ($L_1 = L_2 = 1$) approximation for both mixing vectors with $I_1 = I_2 = 8$. Note that the model of the sources and mixing vectors requires only 126 and 15 values instead of 1024 and 64, respectively, see Table II. This results in a compression of $1 - P_r \frac{J_1+J_2-N_s+1}{K} = 87.70\%$ and $1-L_r \frac{I_1+I_2-N_m+1}{M} = 76.56\%$, respectively. In Figure 13 (left), we report the median of the relative errors on the DOAs ϵ_α. It is clear that the dedicated methods estimate the DOAs more accurately than our method. On the other hand, by exploiting the low-rank structure, we show that it is still possible to get fairly accurate estimates in comparison with well-known dedicated methods. Moreover, our method is applicable for large M.

In a second experiment, we add a third source ($R = 3$) that impinges on the ULA from two different paths ($L_3 = 2$): $\alpha_{13} = -15°$ and $\alpha_{23} = 67°$. We use a third-order ($N_m = 3$) rank-1 and rank-2 approximation for the first two and last mixing vector, respectively, with $I_1 = I_2 = I_3 = 4$. We choose $N_m > 2$ such that the different DOAs of the third source can be found directly from the estimated vectors $\mathbf{u}_{13}^{(1)}$ and $\mathbf{u}_{23}^{(1)}$ (instead of the column space of $\mathbf{S}_3$), see the discussion of uniqueness in Subsection II-C. Note that one simply has to increase the rank L_r in order to cope with a multipath source, while MUSIC and ESPRIT need additional spatial smoothing [62]. The results are shown in Figure 13 (right).

In a third experiment, we use the same setup as in the first experiment but with two near-field sources defined by a DOA and range relative to the first antenna: $\alpha_1 = -17°$, $w_1 = 2(M-1)\Delta$, $\alpha_2 = 41°$, and $w_2 = 3(M-1)\Delta$. We compare our results with a two-dimensional version of MUSIC [63]. Figure 14 shows the median of the relative errors on the DOAs ϵ_α and the ranges ϵ_w. MUSIC estimates both the DOA and range more accurately but is even more computationally expensive because now one has to evaluate a two-dimensional spectrum for many angles and ranges. Here, we used 10^2 equidistant angles and ranges in $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and $[5, 12]$, respectively. In order to cope with near-field sources in our approach, one simply has to increase the rank L_r.

The final experiment uses the same setup as the first experiment but now with $M = 9$ and $K = 100$ with $J_1 = J_2 = 10$ and $I_1 = I_2 = 3$. As can be seen from Figure 15, MUSIC fails to distinguish close DOAs when only a few samples are available and the SNR is low [46]. A small number of sensors M flattens the peaks in the MUSIC spectrum, making the problem more difficult. Our method can still estimate the DOAs accurately in such a setup because it is deterministic, performing even better than ESPRIT.

VI. Conclusion

In this paper, we have introduced a new method for BSS that exploits the fact that many real-life signals are compressible. We expressed this by assuming that the tensorized sources

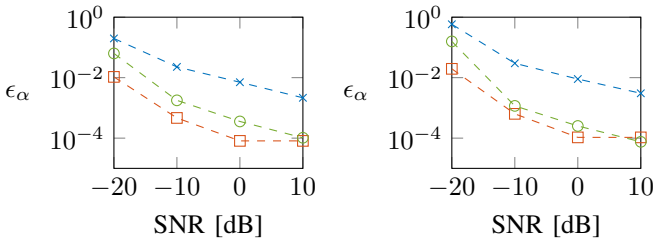


Fig. 13. Median across 100 experiments of the relative error on the DOAs as a function of SNR for the line-of-sight (left) and multipath (right) experiment using segmentation (– ×–), ESPRIT (– ○–), and MUSIC (– □–).

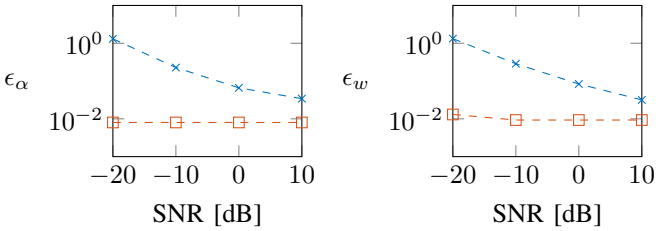


Fig. 14. Median across 100 experiments of the relative error on the DOAs (left) and ranges (right) as a function of SNR for the near field experiment using segmentation (– ×–) and MUSIC (– □–).

can be well approximated by a low-rank model. In other words, we assume that the sources can be well approximated by sums of Kronecker products of smaller vectors. As such, we have demonstrated that, if the sources indeed admit such a low-rank representation/approximation, the BSS problem boils down to the computation of a decomposition of the resulting tensorized observed data matrix. It is precisely the compressibility, which is essential in large-scale problems, that makes it very likely that the tensor decomposition is unique. Hence, our method provides a unique solution to the BSS problem and a way to cope with large-scale problems. Furthermore, we applied the same strategy to the mixing level motivated by an increasing number of sensors and sensor density in fields such as biomedical sciences and array processing. Moreover, combining both strategies simultaneously allowed the exploitation of low-rank structure on both levels of the BSS problem. We have illustrated our methods with two applications: FECG and DOA estimation for large-scale ULAs. We note that it is possible to impose constraints on the sources and/or mixture when applicable, e.g., statistical

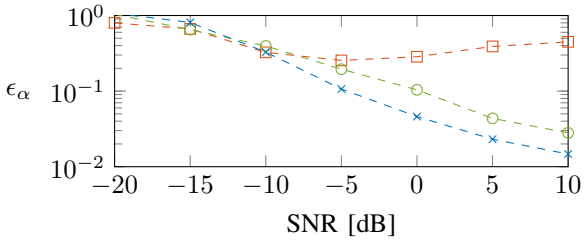


Fig. 15. Median across 100 experiments of the relative error on the DOAs as a function of SNR for the small-scale line-of-sight experiment using segmentation (– ×–), ESPRIT (– ○–), and MUSIC (– □–).

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2016.2617858

independence of the sources as in ICA. Such variants are out of the scope of this paper. Although we focused on the CPD for modeling the tensorized sources and/or mixture, it is possible to consider other tensor models such as tensor trains (TTs) and hierarchical Tucker [15]. The latter are often used in tensor-based scientific computing because they combine large compression rates with good numerical properties. For the CPD of very large tensors, algorithms such as the ones in [17], [57], [58] can be used.

REFERENCES

[1] A. Cichocki and S. Amari, *Adaptive blind signal and image processing: Learning algorithms and applications*. John Wiley Chichester, 2002, vol. 1.

[2] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, no. 4–5, pp. 411–430, June 2000.

[3] C. J. James and C. W. Hesse, "Independent component analysis for biomedical signals," *Physiological Measurement*, vol. 26, no. 1, pp. R15–R39, 2005.

[4] P. Comon and C. Jutten, *Handbook of blind source separation: Independent component analysis and applications*. Academic press, 2009.

[5] L. De Lathauwer, "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*, vol. 32, no. 4, pp. 1451–1474, 2011.

[6] O. Debals, M. Van Barel, and L. De Lathauwer, "Löwner-based blind signal separation of rational functions with applications," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 1909–1918, 2016.

[7] O. Debals and L. De Lathauwer, "Stochastic and deterministic tensorization for blind signal separation," in *Latent Variable Analysis and Signal Separation*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2015, vol. 9237, pp. 3–13.

[8] L. De Lathauwer, "Block component analysis, a new concept for blind source separation," in *Latent Variable Analysis and Signal Separation*, ser. Lecture Notes in Computer Science. Springer Berlin/Heidelberg, 2012, vol. 7191, pp. 1–8.

[9] ——, "Decompositions of a higher-order tensor in block terms — Part I: Lemmas for partitioned matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1022–1032, Sept. 2008.

[10] ——, "Decompositions of a higher-order tensor in block terms — Part II: Definitions and uniqueness," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1033–1066, Sept. 2008.

[11] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural computation*, vol. 13, no. 4, pp. 863–882, Apr. 2001.

[12] Y. Li, A. Cichocki, S. Amari, and S. Shishkin, "Analysis of sparse representation and blind source separation," *Neural computation*, vol. 16, no. 6, pp. 1193–1234, June 2004.

[13] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[14] C. F. Caiafa and A. Cichocki, "Multidimensional compressed sensing and their applications," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 3, no. 6, pp. 355–380, Nov.–Dec. 2013.

[15] L. Grasedyck, D. Kressner, and C. Tobler, "A literature survey of low-rank tensor approximation techniques," *GAMM-Mitteilungen*, vol. 36, no. 1, pp. 53–78, Feb. 2013.

[16] I. Oseledets, "Tensor-train decomposition," *SIAM Journal of Scientific Computing*, vol. 33, no. 5, pp. 2295–2317, Sep. 2011.

[17] N. Vervliet, O. Debals, L. Sorber, and L. De Lathauwer, "Breaking the curse of dimensionality using decompositions of incomplete tensors: Tensor-based scientific computing in big data analysis," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 71–79, Sept. 2014.

[18] A. Bertrand, "Distributed signal processing for wireless EEG sensor networks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 6, pp. 923–935, Dec. 2015.

[19] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[20] M. Boussé, O. Debals, and L. De Lathauwer, "A novel deterministic method for large-scale blind source separation," in *Proceedings of the 23rd European Signal Processing Conference (EUSIPCO 2015, Nice, France)*, Aug. 2015, pp. 1935–1939.

[21] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, Aug. 2009.

[22] A. Cichocki, D. P. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. F. Caiafa, and A. H. Phan, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145–163, Mar. 2015.

[23] N. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *Technical Report 16-34, ESAT-STADIUS, KU Leuven, Leuven, Belgium*, 2016, (submitted).

[24] J. B. Kruskal, "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra and its Applications*, vol. 18, no. 2, pp. 95–138, 1977.

[25] I. Domanov and L. De Lathauwer, "Generic uniqueness conditions for the canonical polyadic decomposition and INDSCAL," *SIAM Journal on Matrix Analysis and Applications*, vol. 36, no. 4, pp. 1567–1589, Nov. 2015.

[26] L. Chiantini and G. Ottaviani, "On generic identifiability of 3-tensors of small rank," *SIAM Journal on Matrix Analysis and Applications*, vol. 33, no. 3, pp. 1018–1037, 2012.

[27] L. Chiantini, G. Ottaviani, and N. Vannieuwenhoven, "An algorithm for generic and low-rank specific identifiability of complex tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 35, no. 4, pp. 1265–1287, 2014.

[28] M. Sørensen and L. De Lathauwer, "Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank-$(L_{r,n}, L_{r,n}, 1)$ terms — Part I: Uniqueness," *SIAM Journal on Matrix Analysis and Applications*, vol. 36, no. 2, pp. 496–522, Apr. 2015.

[29] N. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays," *Journal of Chemometrics*, vol. 14, no. 3, pp. 229–329, Mar. 2000.

[30] I. Domanov and L. De Lathauwer, "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*, vol. 34, no. 3, pp. 855–875, July-Sept. 2013.

[31] ——, "On the uniqueness of the canonical polyadic decomposition of third-order tensors — Part II: Uniqueness of the overal decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 3, pp. 876–903, July-Sept. 2013.

[32] ——, "Canonical polyadic decomposition of third-order tensors: Reduction to generalized eigenvalue decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 35, no. 2, pp. 636–660, Apr.-May 2014.

[33] ——, "Canonical polyadic decomposition of third-order tensors: Relaxed uniqueness conditions and algebraic algorithm," *arXiv:1501.07251*, 2015.

[34] L.-H. Lim and P. Comon, "Blind multilinear identification," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 1260–1280, Feb. 2014.

[35] I. Domanov and L. De Lathauwer, "Generic uniqueness of a structured matrix factorization and applications in blind source separation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 701–711, June 2016.

[36] L. De Lathauwer and A. De Baynast, "Blind deconvolution of DS-CDMA signals by means of decomposition in rank-$(1, L, L)$ terms," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1562–1571, Apr. 2008.

[37] Y. C. Eldar and G. Kutyniok, Eds., *Compressed sensing: Theory and applications*. Cambridge University Press, 2012.

[38] L. N. Trefethen and D. Bau, *Numerical Linear Algebra*. SIAM, 1997.

[39] B. N. Khoromskij, "$\mathcal{O}(d\log N)$-quantics approximation of N-d tensors in high-dimensional numerical modeling," *Constructive Approximation*, vol. 34, no. 2, pp. 257–280, Oct. 2011.

[40] E. E. Tyrtyshnikov, "How bad are hankel matrices?" *Numerische Mathematik*, vol. 67, pp. 261–269, Mar. 1994.

[41] G. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, Sept. 1936.

[42] L. Sorber, M. Van Barel, and L. De Lathauwer, "Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-$(L_r, L_r, 1)$ terms, and a new generalization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 695–720, Apr. 2013.

[43] B. Rubehn, C. Bosman, R. Oostenveld, P. Fries, and T. Stiegliz, "A MEMS-based flexible multichannel ECoG-electrode array," *Journal of Neural Engineering*, vol. 6, no. 3, pp. 1–10, 2009.

[44] D. Seo, J. M. Carmena, J. M. Rabaey, E. Alon, and M. M. Maharbiz, "Neural dust: An ultrasonic, low power solution for chronic brain-machine interfaces," *arXiv preprint arXiv:1307.2196*, July 2013.

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TSP.2016.2617858

[45] M. Sørensen and L. De Lathauwer, "Blind signal separation via tensor decomposition with Vandermonde factor: Canonical polyadic decomposition," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5507–5519, Nov. 2013.
[46] H. Krim and M. Viberg, "Two decades of array signal processing: The parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.
[47] N. Sidiropoulos, R. Bro, and G. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Transactions on Signal Processing*, vol. 48, no. 8, pp. 2377–2388, Aug. 2000.
[48] S. Miron, Y. Song, D. Brie, and K. Wong, "Multilinear direction finding for sensor-array with multiple scales of invariance," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 3, pp. 2057–2070, July 2015.
[49] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO performance evaluation based on measured propagation data," *IEEE Transactions on Wireless Communications*, vol. 14, no. 7, pp. 3899–3911, Mar. 2015.
[50] W. C. Jakes, *Microwave mobile communications*. New York: Wiley, 1974.
[51] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, "Tensorlab 3.0," Mar. 2016. [Online]. Available: http://www.tensorlab.net/
[52] L. Sorber, M. Van Barel, and L. De Lathauwer, "Structured data fusion," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 4, pp. 586–600, June 2015.
[53] V. D. Silva and L.-H. Lim, "Tensor rank and the ill-posedness of the best low-rank approximation problem," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1084–1127, 2008.
[54] W. Krijnen, T. Dijkstra, and A. Stegeman, "On the non-existence of optimal solutions and the occurrence of "degeneracy" in the candecomp/parafac model," *Psychometrika*, vol. 73, no. 3, pp. 431–439, 2008.
[55] M. Sørensen, L. De Lathauwer, P. Comon, S. Icart, and L. Deneire, "Canonical polyadic decomposition with orthogonality constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 33, no. 4, pp. 1190–1213, Oct.-Dec. 2012.
[56] L.-H. Lim and P. Comon, "Nonnegative approximations of nonnegative tensors," *Journal of Chemometrics*, vol. 23, no. 7–8, pp. 432–441, 2009.
[57] N. Vervliet and L. De Lathauwer, "A randomized block sampling approach to canonical polyadic decomposition of large-scale tensors," *IEEE Journal on Selected Topics in Signal Processing*, vol. 10, pp. 284–295, 2016.
[58] N. D. Sidiropoulos, E. E. Papalexakis, and C. Faloutsos, "Parallel randomly compressed cubes : A scalable distributed architecture for big tensor decomposition," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 57–70, Sept. 2014.
[59] B. De Moor, "The singular value decompositions and long and short spaces of noisy matrices," *IEEE Transactions on Signal Processing*, vol. 41, no. 9, pp. 2826–2838, Sept. 1993.
[60] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Fetal electrocardiogram extraction by blind source subspace separation," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 5, pp. 567–572, May 2000.
[61] D. Callaerts, "Signal separation methods based on singular value decomposition and their application to the real-time extraction of the fetal electrocardiogram from cutaneous recordings," Ph.D. dissertation, KU Leuven, Dec. 1989.
[62] T.-J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, no. 4, pp. 806–811, Aug. 1985.
[63] Y.-D. Huang and M. Barkat, "Near-field multiple source localization by passive sensor array," *IEEE Transactions on Antennas and Propagation*, vol. 39, no. 7, pp. 968–975, Jul. 1991.



Martijn Boussé received the M.Sc. in Mathematical Engineering from KU Leuven, Belgium in 2014. He is a Ph.D. candidate affiliated with the STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics of the Electrical Engineering Department (ESAT), KU Leuven and with iMinds Medical IT. His research concerns the development of tensor-based methods for blind signal separation and blind system identification in a big data context.



Otto Debals received the M.Sc. in Mathematical Engineering from KU Leuven, Belgium in 2013. He is a Ph.D. candidate affiliated with the Group Science, Engineering and Technology of Kulak, KU Leuven, with the STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics of the Electrical Engineering Department (ESAT), KU Leuven and with iMinds Medical IT. His research concerns the tensorization of matrix data, with further interests in tensor decompositions, optimization, blind signal separation and blind system identification.



Lieven De Lathauwer received the Ph.D. degree from the Faculty of Engineering, KU Leuven, Belgium, in 1997. From 2000 to 2007 he was Research Associate with the Centre National de la Recherche Scientifique, France. He is currently Professor with KU Leuven. He is affiliated with the Group Science, Engineering and Technology of Kulak, with the STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics of the Electrical Engineering Department (ESAT) and with iMinds Medical IT. He is Associate Editor of SIAM Journal on Matrix Analysis and Applications and has served as Associate Editor for IEEE Transactions on Signal Processing. His research concerns the development of tensor tools for engineering applications.

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.