

A NOVEL DETERMINISTIC METHOD FOR LARGE-SCALE BLIND SOURCE SEPARATION

Martijn Bousse^{*}

Otto Debals^{*†}

Lieven De Lathauwer^{*†}

^{*}Department of Electrical Engineering (ESAT), KU Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium. [†]Group Science, Engineering and Technology, KU Leuven Kulak, E. Sabbelaan 53, 8500 Kortrijk, Belgium.

ABSTRACT

A novel deterministic method for blind source separation is presented. In contrast to common methods such as independent component analysis, only mild assumptions are imposed on the sources. On the contrary, the method exploits a hypothesized (approximate) intrinsic low-rank structure of the mixing vectors. This is a very natural assumption for problems with many sensors. As such, the blind source separation problem can be reformulated as the computation of a tensor decomposition by applying a low-rank approximation to the tensorized mixing vectors. This allows the introduction of blind source separation in certain big data applications, where other methods fall short.

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

1. INTRODUCTION

The goal of blind source separation (BSS) is to reconstruct a collection of unobserved sources based only on a collection of observed signals. In this paper, the latter are unknown linear instantaneous mixtures of the unknown sources. Applications can be found in telecommunications, signal processing and biomedical sciences [1–3]. In general, the solution of the BSS problem is not unique. Hence, several approaches have been proposed that impose additional assumptions on the sources.

A well-known BSS method, called independent component analysis (ICA), assumes statistically independent sources [4]. Recently, a deterministic method based on block term decompositions, called block component analysis (BCA), was introduced in [5]. Elaborating on this idea,

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: CoE PFV/10/002 (OPTEC), (3) FWO: projects: 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.

new methods were proposed that assume the sources can be written as exponential polynomials or rational functions, which allows broad deterministic modeling [6–9].

Previous methods are not applicable to problems with a massive amount of sensors because they do not scale well. This is especially true for ICA, of which several variants use (full) higher-order statistics (HOS) that suffer from the curse of dimensionality [10]. In a big data setting, the mixture possibly has some smooth structure because of the many and closely located sensors. In this paper, we exploit this underlying compactness in order to cope with the large-scale aspect. A comparable strategy has proven to be very successful in the field of tensor-based scientific computing [11].

More specifically, we approximate the tensorized mixing vectors with a low-rank approximation and demonstrate that the BSS problem subsequently boils down to the computation of a tensor decomposition. This method uniquely determines both the sources and the mixing vectors of large-scale problems under mild conditions. Moreover, it imposes only mild assumptions on the sources and because of its deterministic nature it even works well if only a few samples are present.

1.1. Notation and basic 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, \dots, i_N) th entry of an N th-order tensor $\mathcal{A} \in \mathbb{K}^{I_1 \times I_2 \times \dots \times I_N}$ is denoted by $a_{i_1 i_2 \dots i_N}$ with \mathbb{K} meaning \mathbb{R} or \mathbb{C} . The n th element in a sequence is indicated by a superscript between parentheses (e.g., $\{\mathbf{a}^{(n)}\}_{n=1}^N$).

A mode- n vector of a tensor is defined by fixing every index except the n th 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 of \mathcal{A} as its columns (see [12, 13] for formal definitions). The vectorization of \mathcal{A} maps each element $a_{i_1 i_2 \dots i_N}$ 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 product of \mathcal{A} and $\mathcal{B} \in \mathbb{K}^{J_1 \times J_2 \times \dots \times J_M}$ is defined as $(\mathcal{A} \otimes \mathcal{B})_{i_1 i_2 \dots i_N j_1 j_2 \dots j_M} = a_{i_1 i_2 \dots i_N} b_{j_1 j_2 \dots j_M}$. The Kronecker product of $\mathbf{a} \in \mathbb{K}^I$ and $\mathbf{b} \in \mathbb{K}^J$ is defined as $\mathbf{a} \otimes \mathbf{b} = [\mathbf{a}_1 \mathbf{b}^T \quad \mathbf{a}_2 \mathbf{b}^T \quad \dots \quad \mathbf{a}_I \mathbf{b}^T]^T$.

1.2. Multilinear algebraic prerequisites

An N th-order tensor has rank one if it can be written as the outer product of N non-zero 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 N th-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.

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

$$\mathcal{A} = \sum_{r=1}^R \mathbf{u}_r^{(1)} \otimes \mathbf{u}_r^{(2)} \otimes \dots \otimes \mathbf{u}_r^{(N)}. \quad (1)$$

The columns of the factor matrices $\mathbf{U}^{(n)} \in \mathbb{K}^{I_n \times R}$ are equal to the factor vectors $\mathbf{u}_r^{(n)}$, $r = 1, \dots, R$. 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, machine learning and data mining [12, 13]. The decomposition is *essentially unique*, i.e., up to trivial permutations of the rank-1 terms and scalings of the factors in the same term, under rather mild conditions [14–16].

A *block term decomposition* (BTD) in rank- $(L, L, 1)$ terms writes a third-order tensor $\mathcal{X} \in \mathbb{K}^{I \times J \times K}$ as a sum of R terms with multilinear rank- $(L, L, 1)$:

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

in which $\mathbf{A}_r \in \mathbb{K}^{I \times L}$ and $\mathbf{B}_r \in \mathbb{K}^{J \times L}$ have full column rank L . These block terms are more general than the simple rank-1 terms of the PD. Hence, they allow the modeling of more complex phenomena, see e.g., [5, 17]. Other types of BTDs as well as associated uniqueness results are presented in [6, 18].

2. BLIND SOURCE SEPARATION BASED ON LOW-RANK TENSOR APPROXIMATIONS

Blind source separation (BSS) uses the following model [4]:

$$\mathbf{X} = \mathbf{M}\mathbf{S} + \mathbf{N}, \quad (2)$$

with $\mathbf{X} \in \mathbb{K}^{M \times K}$ and $\mathbf{S} \in \mathbb{K}^{R \times K}$ containing K samples of the M observed signals 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; the influence will be investigated in Subsection 3.1 by means of simulations.

2.1. Kronecker product structure

Many real-life signals are compressible, i.e., they depend on much less parameters than their finite length [19, 20]. One

way to represent a signal in a possibly compact way is a (higher-order) low-rank approximation of the tensorized signal [11, 21]. This notion is exploited in a novel way for BSS in this paper. Particularly, in a big data setting the number of sensors M and/or sensor density becomes very large, leading to possibly very smooth mixing vectors in (2) (i.e., the columns of \mathbf{M}). Exploiting this underlying compactness on the mixture level allows us to cope with large-scale BSS.

Let us illustrate this as follows: assume we approximate a vector $\mathbf{m} \in \mathbb{K}^M$ with a Kronecker product of two smaller vectors, i.e., $\mathbf{m} = \mathbf{b} \otimes \mathbf{a}$, with $\mathbf{a} \in \mathbb{K}^I$, $\mathbf{b} \in \mathbb{K}^J$ and $M = IJ$. This is actually a rank-1 approximation. Indeed, the Kronecker and outer product are related through a vectorization: $\mathbf{b} \otimes \mathbf{a} = \text{vec}(\mathbf{a} \otimes \mathbf{b})$, i.e., a (second-order) rank-1 term. The number of coefficients decreases from $M = IJ$ to $\mathcal{O}(I + J)$, which is a decrease of one order of magnitude if $I \approx J$.

More generally, larger reductions can be obtained by considering a Kronecker product of N vectors, which is the vectorization of an outer product of N vectors, i.e., $\bigotimes_{n=1}^N \mathbf{u}^{(n)} = \text{vec}(\bigotimes_{n=1}^N \mathbf{u}^{(N-n+1)})$ with $\mathbf{u}^{(n)} \in \mathbb{K}^{I_n}$, $n = 1, \dots, N$. On the other hand, a rank-1 approximation might not be sufficient; hence, we take a sum of P such Kronecker products resulting in a low-rank approximation [11]:

$$\mathbf{m} = \sum_{p=1}^P \bigotimes_{n=1}^N \mathbf{u}_p^{(n)} = \text{vec} \left(\sum_{p=1}^P \bigotimes_{n=1}^N \mathbf{u}_p^{(N-n+1)} \right), \quad (3)$$

in which $\mathbf{u}_p^{(n)} \in \mathbb{K}^{I_n}$ and $M = \prod_{n=1}^N I_n$. In other words, approximating a vector with a sum of Kronecker products amounts to a low-rank approximation of the ‘folded’ vector, i.e., a low-rank tensor approximation. Note that the number of coefficients decreases from $M = \prod_{n=1}^N I_n$ to $\mathcal{O}(P \sum_{n=1}^N I_n)$. If $I_n = I$ for $n = 1, \dots, N$, then $M = I^N$ reduces to $\mathcal{O}(PNI)$, i.e., a reduction of $N - 1$ orders of magnitude. Hence, we have a decrease in function of the number of Kronecker products and an increase proportional to the number of rank-1 terms P in the sum.

2.2. Tensorization

We now demonstrate that, if the mixing vectors can be approximated by sums of Kronecker products, the BSS problem can be reformulated as the computation of a decomposition of the tensorized observed data matrix. Let us illustrate this as follows: assume the mixing vectors have a simple Kronecker product structure, e.g., $\mathbf{m}_r = \mathbf{b}_r \otimes \mathbf{a}_r$ with $\mathbf{a}_r \in \mathbb{K}^I$, $\mathbf{b}_r \in \mathbb{K}^J$, and assume that $M = IJ$, then we have that:

$$\mathbf{X} = \sum_{r=1}^R (\mathbf{b}_r \otimes \mathbf{a}_r) \otimes \mathbf{s}_r. \quad (4)$$

Equation (4) can be tensorized by stacking each matricized column of the data matrix \mathbf{X} in a third-order tensor $\mathcal{X} \in$

$\mathbb{K}^{I \times J \times K}$. Remember that a Kronecker product is a vectorized outer product; consequently, we have that:

$$\mathcal{X} = \sum_{r=1}^R \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{s}_r. \quad (5)$$

Equation (5) is a CPD as defined in (1). Hence, the above described tensorization strategy reformulates the BSS problem as the computation of a CPD of a third-order tensor of rank R . The number of coefficients reduces from $RM = RIJ$ to $\mathcal{O}(R(I + J))$ (cf. above). The idea can be generalized to a sum of P Kronecker products of N vectors analogous to (3):

$$\mathbf{X} = \sum_{r=1}^R \left(\sum_{p=1}^P \bigotimes_{n=1}^N \mathbf{u}_{pr}^{(n)} \right) \otimes \mathbf{s}_r,$$

in which $\mathbf{u}_{pr}^{(n)} \in \mathbb{K}^{I_n}$, $n = 1, \dots, N$ and $M = \prod_{n=1}^N I_n$, where we have chosen a fixed P for each source r . Applying the same tensorization as before amounts to:

$$\mathcal{X} = \sum_{r=1}^R \left(\sum_{p=1}^P \bigotimes_{n=1}^N \mathbf{u}_{pr}^{(N-n+1)} \right) \otimes \mathbf{s}_r. \quad (6)$$

Hence, we get a decomposition of an $(N+1)$ th-order tensor in R (rank- $P \otimes$ vector) terms [22]. The number of coefficients decreases from RM to $\mathcal{O}(RP \sum_{n=1}^N I_n)$ (cf. above). Note that (6) is a rank- $(P, P, 1)$ BTD if $N = 2$.

The proposed method simultaneously and uniquely determines the mixing vectors *and* the sources under the mild uniqueness conditions of the CPD and BTD. It is 1) applicable for large M , as is clear from the possibly large reductions in the number of coefficients, and 2) is deterministic, meaning that it even works well if the number of samples K is small, which is a clear advantage compared to stochastic methods such as ICA. Moreover, 3) the method imposes only mild conditions on the sources in contrast to existing methods (e.g., linear independence instead of statistical independence).

3. RESULTS AND DISCUSSION

3.1. Simulations

Two experiments are presented in which we investigate 1) the influence of the noise level for several K with exact second-order rank-2 mixing vectors ($N = 2$ and $P = 2$) and 2) the influence of deviations from a second-order rank-1 structure ($N = 2$ and $P = 1$) for several SNRs. Both experiments have $M = 2500$ sensors, and $R = 2$ i.i.d. zero-mean unit-variance Gaussian random source signals of length K . Note that ICA techniques based on (full) HOS would be infeasible here as the number of entries in a Q th-order statistic equals $\mathcal{O}(M^Q)$. Moreover, ICA is not applicable here because of the Gaussian sources [4]. We use a second-order rank-2 and rank-1 approximation with $I = J = 50$ for the first and second experiment, respectively. The approximation consists of only

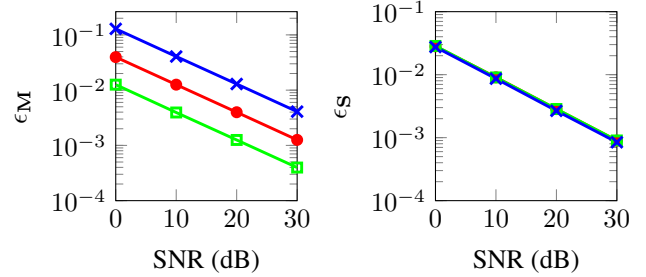


Fig. 1. Median across 100 experiments of the relative error on the mixing vectors (left) and the sources (right) as a function of SNR for $K = 10$ (\times), 100 (\bullet), and 1000 (\square). The mixing vectors satisfy a second-order rank-2 structure.

$\mathcal{O}(I + J) = \mathcal{O}(100)$ values per rank-1 term instead of $IJ = 2500$, which is the maximal reduction for the given order.

The mixing vectors and sources can only be found up to scaling and permutation, which are the standard indeterminacies in BSS. Hence, the mixing vectors and sources 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 $\epsilon_{\mathbf{A}} = \|\mathbf{A} - \hat{\mathbf{A}}\|_{\text{F}} / \|\mathbf{A}\|_{\text{F}}$ with $\hat{\mathbf{A}}$ an optimally scaled and permuted estimate for \mathbf{A} . The CPD and rank- $(L, L, 1)$ -BTB are computed using `cpd` and `btd_nls` from the MATLAB package Tensorlab, respectively, using a generalized eigenvalue decomposition as initialization for the latter [18, 23].

The results of the first experiment are shown in Figure 1. Although the method is deterministic, and therefore requires only a few samples, it is beneficial to increase K under noisy conditions in order to obtain better accuracy. However, the number of samples can still be (very) low in contrast to ICA. The relative error on the sources does not improve for increasing K because we also have to estimate longer source signals.

In the second experiment each mixing vector is defined as the vectorization of a random matrix with exponentially decaying singular values: $\mathbf{m}_r = \text{vec}(\mathbf{U}_r \text{diag}(\boldsymbol{\sigma}) \mathbf{V}_r^T)$, in which $\boldsymbol{\sigma} = e^{-\alpha \mathbf{t}}$ with \mathbf{t} containing $\min(I, J)$ uniformly discretized samples on the interval $[0, 1]$; \mathbf{U}_r and \mathbf{V}_r are random orthogonal matrices of compatible dimensions. The parameter α controls the exponential decay of the singular values: increasing α results in more rank-1-like mixing vectors and vice versa. We take $K = 20$. The relative error on the mixing vectors and the sources for several SNRs is shown in Figure 2. Increasing α decreases the relative error; this is especially clear for high SNR. The decrease in relative error eventually stagnates due to noise. The effect of α is less strong on the sources, probably because the rank-1 approximation already captures a lot of the information quite well. Also, the noise on the sources is possibly (largely) compensated because this is the shorter factor ($K \ll I, J$) in the decomposition.

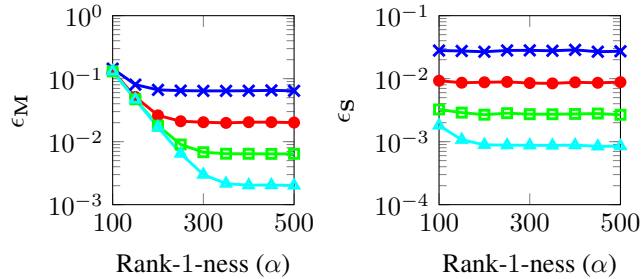


Fig. 2. Median across 100 experiments of the relative error on the mixing vectors (left) and the sources (right) for an SNR of 0 dB (\times), 10 dB (\bullet), 20 dB (\square), and 30 dB (\blacktriangle). The rank-1-ness of the mixing vectors is controlled with α .

3.2. Applications

Recent advances in biomedical sciences have led to applications with many sensors and/or high sensor density such as wireless body area networks (WBANs). These are sensor systems, located in and/or around the body, that measure several different biomedical signals such as temperature, heart-beat and neuronal activity. They are currently investigated for electroencephalography (EEG) [24] and electrocorticography (ECoG) [25] with high spatial resolution as a possible means for long-term neuromonitoring. In such cases our method can provide an efficient separation of the different sources of neuronal activity. Moreover, the data rate of the wireless connection can be reduced because of the deterministic nature of the method. The problem becomes really large-scale when we consider neural dust: here, thousands of miniature sensors, called neural probes, are dispersed throughout the brain [26].

Our method can also be applied in cases where the mixing vectors have an exact low-rank structure. An example of this are uniform linear arrays (ULAs) with far-field sources that emit narrowband signals. Here, the mixing matrix \mathbf{M} has a Vandermonde structure because of the fact that (short) time delays amount to phase shifts of the signal [27]. Then, each mixing vector can be written as a Kronecker product of some smaller vectors, e.g., $[1, a, a^2, a^3]^T = ([1, a^2] \otimes [1, a])^T$, i.e., as a rank-1 term. If the signals propagate through P distinct paths [28], e.g., due to reflections or scattering caused by objects on the path, each mixing vector is low-rank. Note that this Vandermonde structure arises often in signal processing, see [29] and references therein. Assuming a rank-1 structure in such cases is relevant and provides an efficient and deterministic processing of the data. Finally, our method can also be applied when the sources are located in the near field. In this case the Vandermonde structure is only approximate and a low-rank approximation can be applied.

3.3. Future research

In this paper we have limited ourselves to sensor data that are linear instantaneous mixtures of the sources. However, the strategy outlined above can be extended to blind system identification (BSI), also known as equalization, in which the outputs are linear convolutive mixtures of the inputs [30]. More specifically, the same strategy can be applied to the identification of finite impulse response (FIR) systems of the form: $\mathbf{x}[k] = \sum_{r=1}^R \sum_{l=0}^{L-1} \mathbf{m}_r^{(l)} s_r[k-l]$, with memory L ; note that this reduces to BSS if $L = 1$. This problem will be addressed in a follow-up paper with more simulations and applications.

4. CONCLUSION

We have presented a new deterministic method for blind source separation (BSS) of large-scale problems under mild uniqueness conditions. The method can cope with many sensors by exploiting a hypothesized low-rank structure of the mixture. This is done by approximating the mixing vectors by sums of Kronecker products or, in other words, approximating the tensorized mixing vectors by sums of higher-order rank-1 terms. This strategy allowed us to reformulate the BSS problem as the computation of a tensor decomposition. Moreover, the method is deterministic, meaning that it does not need a large number of samples. Finally, in contrast to existing methods, the method imposes only mild assumptions on the sources. The generalization of this strategy to BSI, including FIR systems, will be addressed in a future paper.

REFERENCES

- [1] A. Cichocki and S. Amari, *Adaptive blind signal and image processing: Learning algorithms and applications*, vol. 1, John Wiley Chichester, 2002.
- [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, "Block component analysis, a new concept for blind source separation," in *Latent Variable Analysis and Signal Separation*, vol. 7191 of *Lecture Notes in Computer Science*, pp. 1–8. Springer Berlin/Heidelberg, 2012.
- [6] 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.

- [7] O. Debals, M. Van Barel, and L. De Lathauwer, “Blind signal separation of rational functions using Löwner-based tensorization,” in *Proceedings of the 40th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2015, Brisbane, Australia)*, Apr. 2015, (accepted for publication).
- [8] O. Debals, M. Van Barel, and L. De Lathauwer, “Löwner-based blind signal separation of rational functions with applications,” *Technical Report 15-44, ESAT-STADIUS, KU Leuven, Leuven, Belgium*, 2015.
- [9] O. Debals and L. De Lathauwer, “Stochastic and deterministic tensorization for blind signal separation,” in *Latent Variable Analysis and Signal Separation*, Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2015, (accepted for publication).
- [10] D. L. Donoho, “Aide-Memoire. High-dimensional data analysis: The curses and blessings of dimensionality,” in *AMS Conference on Math Challenges of the 21st Century (Los Angeles, USA)*, Aug. 2000.
- [11] 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.
- [12] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” *SIAM Review*, vol. 51, no. 3, pp. 455–500, Aug. 2009.
- [13] A. Cichocki, D. P. Mandic, A. H. Phan, C. F. Caiafa, G. Zhou, Q. Zhao, and L. De Lathauwer, “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.
- [14] 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.
- [15] 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.
- [16] I. Domanov and L. De Lathauwer, “On the uniqueness of the canonical polyadic decomposition of third-order tensors — Part II: Uniqueness of the overall decomposition,” *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 3, pp. 876–903, July–Sept. 2013.
- [17] 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.
- [18] L. De Lathauwer, “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.
- [19] 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.
- [20] 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.
- [21] L. Grasedyck, “Existence and computation of low Kronecker-rank approximations for large linear systems of tensor product structure,” *Computing*, vol. 72, no. 3–4, pp. 247–265, Jan. 2004.
- [22] 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.
- [23] L. Sorber, M. Van Barel, and L. De Lathauwer, “Tensorlab v2.0,” Jan. 2014, Available online at <http://www.tensorlab.net/>.
- [24] A. Bertrand, “Distributed signal processing for wireless EEG sensor networks,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2015, (accepted for publication).
- [25] B. Rubehn, C. Bosman, R. Oostenveld, P. Fries, and T. Stieglitz, “A MEMS-based flexible multichannel ECoG-electrode array,” *Journal of Neural Engineering*, vol. 6, no. 3, pp. 1–10, 2009.
- [26] 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.
- [27] 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.
- [28] W. C. Jakes, *Microwave mobile communications*, New York: Wiley, 1974.
- [29] 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.
- [30] K. Abed-Meraim, W. Qiu, and Y. Hua, “Blind system identification,” *Proceedings of the IEEE*, vol. 85, no. 8, pp. 1310–1322, Aug. 1997.