Block Component Analysis, a New Concept for Blind Source Separation

Lieven De Lathauwer

Katholieke Universiteit Leuven Campus Kortrijk,
E. Sabbelaan 53, 8500 Kortrijk, Belgium
Lieven.DeLathauwer@kuleuven-kortrijk.be
http://homes.esat.kuleuven.be/~delathau/

Abstract. The fact that the decomposition of a matrix in a minimal number of rank-1 terms is not unique, leads to a basic indeterminacy in factor analysis. Factors and loadings are only unique under certain assumptions. Working in a multilinear framework has the advantage that the decomposition of a higher-order tensor in a minimal number of rank-1 terms (its Canonical Polyadic Decomposition (CPD)) is unique under mild conditions. We have recently introduced Block Term Decompositions (BTD) of a higher-order tensor. BTDs write a given tensor as a sum of terms that have low multilinear rank, without having to be rank-1. In this paper we explain how BTDs can be used for factor analysis and blind source separation. We discuss links with Canonical Polyadic Analysis (CPA) and Independent Component Analysis (ICA). Different variants of the approach are illustrated with examples.

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

1 Algebraic Tools

We start with a few basic definitions from multilinear algebra. These are subsequently used to define two tensor decompositions.

Definition 1. *A* mode-n vector *of an Nth-order tensor $\mathcal{T} = [t_{i_1 i_2 \ldots i_N}]$ is a vector obtained by varying the n-th index and keeping the other indices fixed.*

Definition 2. *The* multilinear rank *of an Nth-order tensor is the N-tuplet consisting of the dimension of the space spanned by the mode-1 vectors, the dimension of the space spanned by the mode-2 vectors, and so on.*

Definition 3. *The* (tensor) outer product $\mathcal{A} \otimes \mathcal{B}$ *of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_P}$ and a tensor $\mathcal{B} \in \mathbb{C}^{J_1 \times J_2 \times \ldots \times J_Q}$ is the tensor defined by $(\mathcal{A} \otimes \mathcal{B})_{i_1 i_2 \ldots i_P j_1 j_2 \ldots j_Q} = a_{i_1 i_2 \ldots i_P} b_{j_1 j_2 \ldots j_Q}$, for all values of the indices.*

For instance, the outer product $\mathcal{T}$ of three vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ is defined by $t_{ijk} = a_i b_j c_k$.

Definition 4. *An Nth-order tensor has* rank 1 *iff it equals the outer product of N nonzero vectors.*

Definition 5. *The* rank *of a tensor $\mathcal{T}$ is the minimal number of rank-*1 *tensors that yield $\mathcal{T}$ in a linear combination.*

We can now define a first basic tensor decomposition.

Definition 6. *A* Canonical Polyadic Decomposition (CPD) *of a rank-R tensor $\mathcal{T} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ is a decomposition of $\mathcal{T}$ in a sum of R rank-1 terms:*

$$\mathcal{T} = \sum_{r=1}^{R} \mathbf{a}_r^{(1)} \otimes \mathbf{a}_r^{(2)} \otimes \cdots \otimes \mathbf{a}_r^{(N)} \ . \tag{1}$$

The decomposition was for the first time used for data analysis in [3] and [14], where it was called Canonical Decomposition (CANDECOMP) and Parallel Factor Decomposition (PARAFAC), respectively. The term CPD, where "CP" may also stand for "CANDECOMP/PARAFAC", is now becoming more common. An important advantage over the decomposition of a matrix in rank-1 terms, is that CPD of a higher-order tensor is unique under mild conditions, see [11, 16, 17, 20, 21] and references therein. (Uniqueness is up to permutation of terms and scaling/counterscaling of factors within a term.) For algorithms, see [11, 16, 22, 23] and references therein.

Consider a third-order tensor $\mathcal{T} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ that has CPD

$$\mathcal{T} = \sum_{r=1}^{R} \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r \ . \tag{2}$$

Define $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \ldots \ \mathbf{a}_R] \in \mathbb{C}^{I_1 \times R}$, $\mathbf{B} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \ldots \ \mathbf{b}_R] \in \mathbb{C}^{I_2 \times R}$ and $\mathbf{C} = [\mathbf{c}_1 \ \mathbf{c}_2 \ \ldots \ \mathbf{c}_R] \in \mathbb{C}^{I_3 \times R}$. Eq. (2) is often written as

$$\mathbf{T}_{:,:,i_3} = \mathbf{A} \cdot \mathrm{diag}(c_{i_3 1}, c_{i_3 2}, \ldots, c_{i_3 R}) \cdot \mathbf{B}^T, \qquad 1 \leqslant i_3 \leqslant I_3 \ , \tag{3}$$

in which we use MATLAB colon notation. We see that all slices $\mathbf{T}_{:,:,i_3}$ are linear combinations of the same rank-1 terms $\mathbf{a}_r \mathbf{b}_r^T$, $1 \leqslant r \leqslant R$, where the coefficients are given by the entries of $\mathbf{C}$.

In [8, 9, 13] we introduced Block Term Decompositions (BTD) of a higher-order tensor. BTDs are a generalization of CPD. A specific case is the following.

Definition 7. *A* decomposition of a tensor $\mathcal{T} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ in a sum of rank-$(L_r, L_r, 1)$ terms, $1 \leqslant r \leqslant R$, *is a decomposition of $\mathcal{T}$ of the form*

$$\mathcal{T} = \sum_{r=1}^{R} (\mathbf{A}_r \cdot \mathbf{B}_r^T) \otimes \mathbf{c}_r \ , \tag{4}$$

in which each of the matrices $\mathbf{A}_r \in \mathbb{C}^{I_1 \times L_r}$ and $\mathbf{B}_r \in \mathbb{C}^{I_2 \times L_r}$ has linearly independent columns and in which the vectors $\mathbf{c}_r \in \mathbb{C}^{I_3}$ are nonzero, $1 \leqslant r \leqslant R$. We assume that R is minimal.

Conditions under which this decomposition is unique, have been established in [8–10]. (Here, uniqueness is up to permutation of terms, scaling/counterscaling of factors within a term, and post-multiplication of $\mathbf{A}_r$ by a square nonsingular matrix $\mathbf{W}_r$ provided $\mathbf{B}_r^T$ is pre-multiplied by $\mathbf{W}_r^{-1}$, $1 \leqslant r \leqslant R$.) Algorithms have been presented in [13, 18, 19, 22]. Note that $(L_r, L_r, 1)$ is the multilinear rank of the r-th term.

Define $\mathbf{A} = [\mathbf{A}_1\ \mathbf{A}_2\ \ldots\ \mathbf{A}_R] \in \mathbb{C}^{I_1 \times \sum_r L_r}$, $\mathbf{B} = [\mathbf{B}_1\ \mathbf{B}_2\ \ldots\ \mathbf{B}_R] \in \mathbb{C}^{I_2 \times \sum_r L_r}$ and $\mathbf{C} = [\mathbf{c}_1\ \mathbf{c}_2\ \ldots\ \mathbf{c}_R] \in \mathbb{C}^{I_3 \times R}$. Eq. (4) can also be written as

$$\mathbf{T}_{:,:,i_3} = \mathbf{A} \cdot \mathrm{diag}(c_{i_3 1}\mathbf{I}_{L_1 \times L_1}, c_{i_3 2}\mathbf{I}_{L_2 \times L_2}, \ldots, c_{i_3 R}\mathbf{I}_{L_R \times L_R}) \cdot \mathbf{B}^T, \quad 1 \leqslant i_3 \leqslant I_3\ . \tag{5}$$

All slices $\mathbf{T}_{:,:,i_3}$ are linear combinations of the same rank-L_r matrices $\mathbf{A}_r\mathbf{B}_r^T$, $1 \leqslant r \leqslant R$, where the coefficients are given by the entries of $\mathbf{C}$.

In the next section we explain how CPD and decomposition in rank-$(L_r, L_r, 1)$ terms can be used for blind source separation.

2 Block Component Analysis: the Concept

Factor analysis and blind source separation aim at decomposing a data matrix $\mathbf{X} \in \mathbb{C}^{K \times N}$ into a sum of interpretable rank-1 terms:

$$\mathbf{X} = \mathbf{A} \cdot \mathbf{S}^T = \sum_{r=1}^{R} \mathbf{a}_r \mathbf{s}_r^T\ . \tag{6}$$

Here, $\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_R] \in \mathbb{C}^{K \times R}$ is the unknown mixing matrix and the columns of $\mathbf{S} = [\mathbf{s}_1\ \mathbf{s}_2\ \ldots\ \mathbf{s}_R] \in \mathbb{C}^{N \times R}$ are the unknown sources. (We consider the noiseless case for clarity of exposition.) Since the decomposition of a matrix is not unique, some assumptions need to be made. In Independent Component Analysis (ICA) the most important assumption is that the sources are mutually statistically independent [5, 6, 15].

If we dispose of a data tensor, then things are simpler, in the sense that the decomposition in rank-1 terms is unique under mild conditions, as mentioned above. This uniqueness makes CPD a powerful tool for data analysis [4, 17, 21]. ICA can actually be seen as a form of Canonical Polyadic Analysis (CPA). Namely, algebraic methods for ICA typically rely on the CPD of a higher-order cumulant tensor or a third-order tensor in which a set of covariance matrices is stacked. The links are explicitly discussed in [11].

The crucial observation on which Block Component Analysis (BCA) is based, is that also the constraints in CPA are in a certain sense restrictive. Namely, in (3) the matrix slices are decomposed in terms that are rank-1, i.e., they consist of the outer product of two vectors. One could wish to decompose the slices in terms that just have low rank, since the latter enable the modelling of more general phenomena. As explained, this corresponds to the decomposition of a tensor in rank-$(L_r, L_r, 1)$ terms, which is still unique under certain conditions. Probably CPA owes much of its success to rank-1 terms that capture the essence

of components that actually are more complex. In such cases it could be of interest to check whether BCA provides more detail.

BCA can be applied to matrix data as well. First, the data need to be tensorized. For instance, one could map the rows of $\mathbf{X}$ to $(I \times J)$ Hankel matrices, with $I + J - 1 = N$, yielding a tensor $\mathcal{X} \in \mathbb{C}^{I \times J \times K}$. Formally, we define:

$$(\mathcal{X})_{ijk} = (\mathbf{X})_{k,i+j-1}, \qquad 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K. \tag{7}$$

Since the mapping is linear, we have:

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{H}_r \otimes \mathbf{a}_r, \tag{8}$$

in which $\mathbf{H}_r \in \mathbb{C}^{I \times J}$ is the Hankel matrix associated with the r-th source, $1 \leqslant r \leqslant R$. An interesting property is that, for a sufficient number of samples, Hankel matrices associated with exponential polynomials have low rank. Exponential polynomials are functions that can be written as sums and/or products of exponentials, sinusoids and/or polynomials. BCA allows the blind separation of such signals, provided decomposition (8) is unique. Uniqueness conditions guarantee that the components are sufficiently different to allow separation, which in turn implies a bound on the number of components one can deal with. Also, there is a trade-off between complexity, measured by rank L_r, and number of components. For theory underlying the blind separation of exponential polynomials by means of a decomposition in rank-$(L_r, L_r, 1)$ terms, we refer to [10].

Hankelization is just one way to tensorize matrix data. What is essential is that we use a linear transformation that maps the sources to matrices that (approximately) have low rank. Possible alternatives are spectrograms, wavelet representations, etc. For comparison we repeat that in ICA the problem is typically tensorized through the computation of higher-order statistics or sets of second-order statistics.

In the next section we illustrate the principle of BCA by means of examples.

3 Illustration

3.1 Toy Example: Audio

We consider the following sources: $\mathbf{s}_1$ consists of samples 50–80 of the chirp demo signal and $\mathbf{s}_2$ consists of samples 250–280 of the train demo signal in MATLAB (version 7.13). These two signals are shown in Fig. 1. The singular values of the corresponding Hankel matrices $\mathbf{H}_1, \mathbf{H}_2 \in \mathbb{R}^{16 \times 16}$ are shown in Fig. 2. We see that $\mathbf{H}_1$ and $\mathbf{H}_2$ can be very well approximated by low-rank matrices. The entries of $\mathbf{A} \in \mathbb{R}^{5 \times 2}$ are drawn from a zero-mean unit-variance Gaussian distribution. Hankelization of $\mathbf{X} \in \mathbb{R}^{5 \times 31}$ yields a tensor $\mathcal{X}^{(H)} \in \mathbb{R}^{16 \times 16 \times 5}$. We also map $\mathbf{X}$ to a tensor $\mathcal{X}^{(W)} \in \mathbb{R}^{40 \times 31 \times 5}$ by means of the biorthogonal spline wavelet 1.3 [7]. This transformation maps every observed time signal to a (scale $\times$ time) matrix, where we take the scale values equal to $0.8/(0.05s)$, $1 \leqslant s \leqslant 40$. The singular

values of the wavelet representations $\mathbf{W}_1, \mathbf{W}_2 \in \mathbb{R}^{40\times 31}$ of the two sources are shown in Fig. 2. Tensors $\mathcal{X}^{(H)}$ and $\mathcal{X}^{(W)}$ are decomposed in a sum of a rank-$(L_1, L_1, 1)$ and a rank-$(L_2, L_2, 1)$ term. We conduct a Monte Carlo experiment consisting of 100 runs for different values of L_1 and L_2. The mean and median Signal-to-Interference Ratio (SIR) are shown in Table 1. This table demonstrates that BCA allows one to accurately separate the sources. Moreover, the choice of L_1 and L_2 turns out not to be very critical. The ICA algorithm in [5] yields a mean and median SIR of only 15 dB, due to the fact that in this toy example not enough samples are available to allow the reliable estimation of statistics.

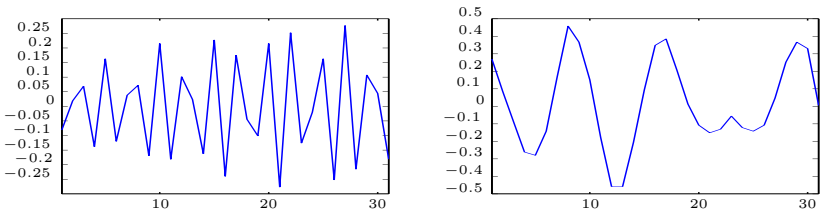


Fig. 1. Chirp (left) and train (right) audio source.

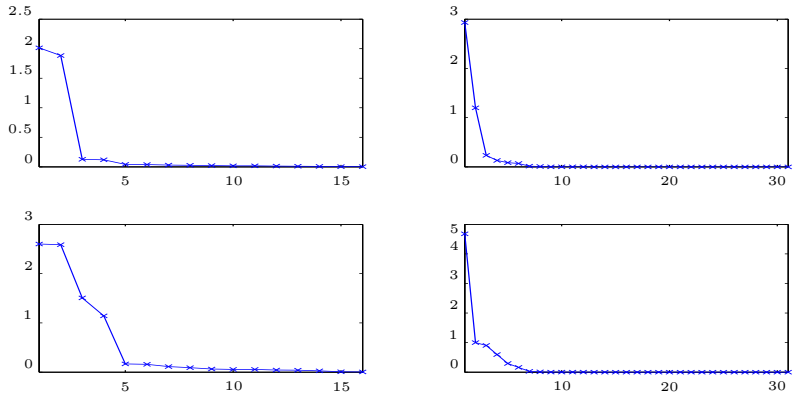


Fig. 2. Left: singular values of Hankel matrices $\mathbf{H}_1$ (top) and $\mathbf{H}_2$ (bottom). Right: singular values of wavelet matrices $\mathbf{W}_1$ (top) and $\mathbf{W}_2$ (bottom).

We next add zero-mean Gaussian noise to the observations and investigate the effect of the Signal-to-Noise Ratio (SNR) on the quality of the separation. We conduct a new Monte Carlo simulation consisting of 100 runs. The value of $L_1 = L_2 = L$ is varied between 1 and 4. The results are shown in Fig. 3. A rank-1 structure turns out to be too simple, at least in the Hankel case.

In the Hankel setting, the signals that correspond to rank-1 matrices are complex exponentials (one frequency, one damping factor). A rank-1 term is

L_1 / L_2	1	2	3	4	5	6	7
1	20 (49)	48 (47)	49 (49)	37 (51)	20 (51)	15 (19)	15 (13)
	43 (43)	41 (41)	19 (41)	19 (41)	16 (16)	14 (13)	13 (12)
2	48 (47)	47 (47)	49 (50)	48 (49)	44 (51)	17 (38)	16 (22)
	41 (41)	46 (46)	47 (49)	48 (52)	33 (33)	17 (18)	14 (17)
3	49 (49)	49 (50)	49 (49)	47 (48)	23 (49)	20 (47)	19 (45)
	19 (41)	47 (49)	46 (46)	27 (41)	25 (32)	18 (33)	13 (12)
4	37 (51)	48 (49)	47 (48)	47 (47)	47 (48)	20 (46)	18 (44)
	19 (41)	48 (52)	27 (41)	52 (52)	16 (21)	47 (48)	13 (12)
5	20 (51)	44 (51)	23 (49)	47 (48)	45 (48)	29 (46)	16 (44)
	16 (16)	33 (33)	25 (32)	16 (21)	13 (13)	28 (35)	12 (12)
6	15 (19)	17 (38)	20 (47)	20 (46)	29 (46)	25 (46)	33 (47)
	14 (13)	17 (18)	18 (33)	47 (48)	28 (35)	46 (47)	17 (25)
7	15 (13)	16 (22)	19 (45)	18 (44)	16 (44)	33 (47)	24 (44)
	13 (12)	14 (17)	13 (12)	13 (12)	12 (12)	17 (25)	17 (20)

Table 1. Mean (median) SIR [dB] in the noiseless audio example, as a function of L_1 and L_2 (top: Hankel-based BCA, bottom: wavelet-based BCA).

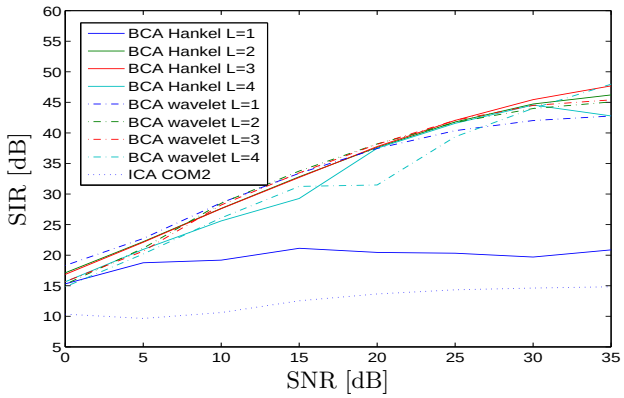


Fig. 3. Mean SIR as a function of SNR in the audio example.

sometimes called an atom, since it is a constituent element that cannot be split into smaller parts. In this terminology, CPA consists of splitting a data tensor into atoms. On the other hand, one could say that sounds or melodies, having a certain spectral content, correspond to molecules rather than atoms. BCA is then the separation at the level of molecules.

3.2 Application in Wireless Communication

In spread-spectrum systems that employ an antenna array at the receiver, the received data are naturally represented by the third-order tensor that shows the signal along the temporal, spectral and spatial axis. In [20] it was shown for Direct Sequence - Code Division Multiple Access (DS-CDMA) systems that, in simple propagation scenarios that do not cause Inter-Symbol-Interference (ISI), every user contributes a rank-1 term to the received data. Consequently, in a non-cooperative setting multiple access can be realized through the computation

of a CPD. In propagation scenarios that do involve ISI, rank-1 terms are a too restrictive model. It was shown in [12] that, when reflections only take place in the far field of the receive array, multiple access can be realized through the computation of a decomposition in rank-$(L_r, L_r, 1)$ terms. In [18] a more general type of BTD was used to deal with cases where reflections do not only take place in the far field. The same ideas can be applied to other systems with at least triple diversity.

4 Discussion and Conclusion

CPA makes a strong assumption on the components that one looks for, namely, that they are rank-1. In the analysis of text data, web documents, biomedical data, images, ... it is often questionable whether this assumption is satisfied. Low (multilinear) rank may be a better approximation of reality. In this paper we introduced BCA as an analysis technique based on the computation of BTDs. BCA can be used for the analysis of matrix data, after these have been tensorized. To this end, one can compute statistics, like in ICA, but one can also consider Hankel representations, wavelet representations, etc. Deterministic variants of BCA may be useful for the analysis of short data sequences.

BCA is related to Sparse Component Analysis (SCA) [1]. In SCA, the sources are low-dimensional in the sense that they are most often zero. In BCA, the sources have a low intrinsic dimension, characterized by multilinear rank. BCA is also related to compressive sensing [2]. In compressive sensing, low intrinsic dimensionality is used for compact signal representation. In BCA, it is used as the basis for signal separation.

In this paper we limited ourselves to the decomposition in rank-$(L_r, L_r, 1)$ terms. In [8, 9, 13] more general types of BTD were introduced, which allow a more general analysis.

Acknowledgments. Research supported by: (1) Research Council K.U.Leuven: GOA-MaNet, CoE EF/05/006 Optimization in Engineering (OPTEC), CIF1 and STRT1/08/023, (2) F.W.O.: Research Communities ICCoS, ANMMM and MLDM, (3) the Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, "Dynamical systems, control and optimization", 2007–2011), (4) EU: ERNSI.

References

1. Bruckstein, A.M., Donoho, D.L., Elad, M.: From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM Rev. 51(1), 34–81 (2009)
2. Candes, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. on Information Theory 52(2), 489–509 (Feb 2006)
3. Carroll, J.D., Chang, J.J.: Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart–Young" decomposition. Psychometrika 35(3), 283–319 (Sep 1970)

4. Cichocki, A., Zdunek, R., Phan, A.H., Amari, S.I.: Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation. John Wiley & Sons (2009)
5. Comon, P.: Independent Component Analysis, a new concept? Signal Processing 36(3), 287–314 (Apr 1994)
6. Comon, P., Jutten, C. (eds.): Handbook of Blind Source Separation, Independent Component Analysis and Applications. Academic Press (2010)
7. Daubechies, I.: Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 61. SIAM (1994)
8. De Lathauwer, L.: Decompositions of a higher-order tensor in block terms — Part I: Lemmas for partitioned matrices. SIAM J. Matrix Anal. Appl. 30, 1022–1032 (2008)
9. De Lathauwer, L.: Decompositions of a higher-order tensor in block terms — Part II: Definitions and uniqueness. SIAM J. Matrix Anal. Appl. 30, 1033–1066 (2008)
10. De Lathauwer, L.: Blind separation of exponential polynomials and the decomposition of a tensor in rank-$(L_r, L_r, 1)$ terms. SIAM J. Matrix Anal. Appl. 32(4), 1451–1474 (Dec 2011)
11. De Lathauwer, L.: A short introduction to tensor-based methods for factor analysis and blind source separation. In: Proc. 7th Int. Symp. on Image and Signal Processing and Analysis (ISPA 2011). pp. 558–563. Dubrovnik, Croatia (Sep 4–6, 2011)
12. De Lathauwer, L., de Baynast, A.: Blind deconvolution of DS-CDMA signals by means of decomposition in rank-$(1, L, L)$ terms. IEEE Trans. on Signal Processing 56(4), 1562–1571 (Apr 2008)
13. De Lathauwer, L., Nion, D.: Decompositions of a higher-order tensor in block terms — Part III: Alternating least squares algorithms. SIAM J. Matrix Anal. Appl. 30, 1067–1083 (2008)
14. Harshman, R.A.: Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. UCLA Working Papers in Phonetics 16 (1970)
15. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley & Sons (2001)
16. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Rev. 51(3), 455–500 (Sep 2009)
17. Kroonenberg, P.M.: Applied Multiway Data Analysis. Wiley (2008)
18. Nion, D., De Lathauwer, L.: Block component model based blind DS-CDMA receivers. IEEE Trans. on Signal Processing 56(11), 5567–5579 (Nov 2008)
19. Nion, D., De Lathauwer, L.: An enhanced line search scheme for complex-valued tensor decompositions. Application in DS-CDMA. Signal Processing 88(3), 749–755 (Mar 2008)
20. Sidiropoulos, N.D., Giannakis, G.B., Bro, R.: Blind PARAFAC receivers for DS-CDMA systems. IEEE Trans. Signal Process. 48(3), 810–823 (Mar 2000)
21. Smilde, A., Bro, R., Geladi, P.: Multi-way Analysis with Applications in the Chemical Sciences. John Wiley & Sons, Chichester, U.K. (2004)
22. Sorber, L., Van Barel, M., De Lathauwer, L.: Optimization-based algorithms for tensor decompositions: Canonical Polyadic Decomposition, decomposition in rank-$(L_r, L_r, 1)$ terms and a new generalization. Tech. rep. 2011-182, ESAT-SISTA, K.U.Leuven (Leuven, Belgium)
23. Tomasi, G., Bro, R.: A comparison of algorithms for fitting the PARAFAC model. Comp. Stat. & Data Anal. 50(7), 1700–1734 (2006)