Algorithm for imposing SOBI-type constraints on the CP model

Maarten De Vos, Lieven De Lathauwer and Sabine Van Huffel

ABSTRACT

We propose a new algorithm to impose independence constraints based on second order statistics in one mode of the Parallel Factor Analysis / Canonical Decomposition, also known as the CP model, and show with simulations that it outperforms in some cases the ordinary CP model.

I. INTRODUCTION

Blind Source Separation (BSS) or Independent Component Analysis (ICA) techniques are becoming standard concepts in signal processing in order to decompose a measurement of a mixture of signals into the source signals and corresponding mixing matrix. These techniques can be used when the data are two dimensional. When the data is three or higher dimensional, generalizations can be derived. This idea was originally investigated in [1] for analyzing functional Magnetic Resonance Imaging (fMRI) datasets and an improvement was developed in [16]. Both works use higher order statistics. In this paper we propose an algorithm to impose independence based on second order, SOBI-type statistics [2]. It should be emphasized that the work presented here solves a different problem than the work presented in [16] as the assumptions about the source signals are different.

A. Basic Definitions

Definition 1. A 3rd-order tensor $\mathcal{T}$ has rank 1 if it equals the outer product of 3 vectors A_1, B_1, C_1: $t_{ijk} = a_i b_j c_k$ for all values of the indices. The outer product of A_1, B_1 and C_1 is denoted by $A_1 \circ B_1 \circ C_1$.

Definition 2. The rank of a tensor is defined as the minimal number of rank-1 terms in which the tensor can be decomposed.

M. De Vos, L. De Lathauwer and S. Van Huffel are with the Department of Electrical Engineering (ESAT), Katholieke Universiteit Leuven, Leuven, Belgium; Research funded by PhD grants of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen); Research supported by **Research Council KUL**:GOA-AMBioRICS, CoE EF/05/006 Optimization in Engineering, IDO 05/010 EEG-fMRI, CIF1; **Flemish Government**: FWO: projects, G.0407.02 (support vector machines), G.0360.05 (EEG, Epileptic), G.0519.06 (Noninvasive brain oxygenation), FWO-G.0321.06 (Tensors/Spectral Analysis), G.0341.07 (Data fusion); IWT: PhD Grants; **Belgian Federal Science Policy Office** IUAP P6/04 ('Dynamical systems, control and optimization', 2007-2011); E-mail: maarten.devos@esat.kuleuven.be

Ł. De Lathauwer is also with the sub-faculty sciences of Katholieke Universiteit Leuven - campus Kortrijk, Kortrijk, Belgium

Definition 3. The Frobenius norm of a tensor $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$ is defined as

$$||\mathcal{T}||_F = (\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} t_{ijk}^2)^{\frac{1}{2}} \quad (1)$$

Notation Scalars are denoted by lower-case letters (a, b, ...), vectors are written as capitals (A, B, ...) (italic shaped), matrices correspond to bold-face capitals ($\mathbf{A}$, $\mathbf{B}$, ...) and tensors are written as calligraphic letters ($\mathcal{A}$, $\mathcal{B}$, ...). This notation is consistently used for lower-order parts of a given structure. For instance, the entry with row index i and column index j in a matrix $\mathbf{A}$, i.e. $(\mathbf{A})_{ij}$, is symbolized by a_{ij} (also $(A)_i = a_i$ and $(\mathcal{A})_{i_1 i_2 \ldots i_N} = a_{i_1 i_2 \ldots i_N}$). The ith column vector of a matrix $\mathbf{A}$ is denoted as A_i, i.e. $\mathbf{A} = [A_1 A_2 \ldots]$. Italic capitals are also used to denote index upper bounds (e.g. $i = 1, 2, \ldots, I$).

$\odot$ is the Khatri-Rao or column-wise Kronecker product.

B. The concept of ICA

Assume the basic linear statistical model

$$T = \mathbf{M} \cdot C^T + N \quad (2)$$

where $T \in \mathbb{R}^I$ is called the observation vector, $C^T \in \mathbb{R}^R$ the source vector and $N \in \mathbb{R}^I$ additive, temporally and spatially white noise. $\mathbf{M} \in \mathbb{R}^{I \times R}$ is the mixing matrix.

The goal of Independent Component Analysis is to estimate the mixing matrix $\mathbf{M}$, and/or the source vector C^T, given only realizations of a linear mixture of these source vectors (T). In this study, we assume that $I \geqslant R$, i.e. there are more observations than source signals.
Blind identification of $\mathbf{M}$ in (2) is only possible when some assumptions about the sources are made. One possibility is assuming that the sources are mutually statistically independent, as well as independent from the noise components and that at most one source is gaussian [5], [3].

Another possibility is assuming that the sources are mutually uncorrelated but individually correlated in time. In this case, the sources can be estimated with the SOBI algorithm [2] (Second Order Blind Identification). This algorithm has shown to be useful in several biomedical applications. Furthermore, the SOBI algorithm is based on second order statistics. This makes it possible to compute the source signals from shorter datasets.

For more details, we refer to [9], [7].

Mathematically, the SOBI algorithm works as follows. For all time lags τ the source correlation matrices are

978-1-4244-1684-4/08/$25.00 ©2008 IEEE

diagonal.

$$\mathbf{R}_t(\tau) = E\{T(t)T(t+\tau)^T\} \quad (3)$$
$$= \mathbf{M} \cdot \mathbf{R}_c(\tau) \cdot \mathbf{M}^T \qquad \forall\tau \quad (4)$$

where $\mathbf{R}_c$ is the correlation matrix of the source signals. Considering that this equation holds for all values of τ, the mixing matrix $\mathbf{M}$ is the inverse of the matrix that jointly diagonalizes all the (observation) correlation matrices.
After prewhitening, this simultaneous diagonalization can be computed by a Jacobi iteration [3]. Necessary and sufficient conditions for uniqueness of this $\mathbf{M}$ are given in [2].

C. The CP Model

1) The model: The CP model [6], [4], [14] of a three-way tensor $\mathcal{T} \in \mathbb{R}^{I\times J\times K}$ is a decomposition of $\mathcal{T}$ as a linear combination of a minimal number R of rank-1 terms:

$$\mathcal{T} = \sum_{r=1}^{R} \lambda_r \, A_r \circ B_r \circ C_r \, (+\mathcal{E}) \quad (5)$$

A pictorial representation of the CP model for third-order tensors is given in figure 1. The λ_i's can also be absorbed in one mode and the noise term is not explicitly considered.

Consider a third-order $(I \times J \times K)$ tensor $\mathcal{T}$ of which the CP model can be expressed as

$$t_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr}, \qquad \forall i,j,k \quad (6)$$

in which $\mathbf{A} \in \mathbb{R}^{I\times R}$, $\mathbf{B} \in \mathbb{R}^{J\times R}$ and $\mathbf{C} \in \mathbb{R}^{K\times R}$. Another equivalent and useful expression of the same CP model can be given in terms of the Khatri-Rao product. We assume that $\min(IJ, K) \geqslant R$.
Associate with $\mathcal{T}$ a matrix $\mathbf{T} \in \mathbb{R}^{IJ\times K}$ as follows:

$$(\mathbf{T})_{(i-1)J+j,k} = \mathcal{T}_{ijk}. \quad (7)$$

This matrix has following structure:

$$\mathbf{T} = (\mathbf{A} \odot \mathbf{B}) \cdot \mathbf{C}^T. \quad (8)$$

2) Computation of the CP decomposition: Originally, an alternating least-squares (ALS) algorithm was proposed in order to minimize the least squares cost function for the computation of the CP decomposition:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} ||\mathbf{T}^T - \mathbf{C}(\mathbf{A} \odot \mathbf{B})^T||^2. \quad (9)$$

Due to the symmetry of the model in the different modes, the updates for all modes are essentially identical with the role of the different modes shifted. Assume that $\mathbf{B}$ and $\mathbf{C}$ are fixed, the estimate of the other can be optimized with a classical linear least squares problem:

$$\min_{\mathbf{C}} ||\mathbf{T}^T - \mathbf{C}\mathbf{Z}^T||^2 \quad (10)$$

where $\mathbf{Z}$ equals $\mathbf{A} \odot \mathbf{B}$. The ALS algorithm computes the CP model by alternating between updates of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, while keeping the two other modes fixed.
Afterwards, it was also shown that the CP decomposition can be in theory computed from an eigen value decomposition (EVD) [12], [13] under certain assumptions among which the most restrictive is that $R \leqslant \min\{I, J\}$. This results in a faster computation. However, when the model is only approximately valid, this will only form the initialization of the ALS-procedure.
In [11], it is shown that the computation of the CP model, based on a simultaneous EVD is actually more robust than a single EVD. This again implied the rank condition $R \leqslant \min\{I, J\}$. As we will need this algorithm in the further developments, we review the computational scheme here. The decomposition is known when $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are computed. In this algorithm, $\mathbf{A}$ and $\mathbf{B}$ are estimated and $\mathbf{C}$ is then computed from equation (8). Substitution of (7) in (6) shows that any vector in the range of matrix $\mathbf{T}$, can be represented by an $I \times J$ matrix that can be decomposed as:

$$\mathbf{V} = \mathbf{A} \cdot \mathbf{D} \cdot \mathbf{B}^T \quad (11)$$

with $\mathbf{D}$ diagonal. If the range is spanned by K matrices $\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_K$, the computation of the canonical decomposition can be obtained by the simultaneous decomposition of the set $\{\mathbf{V}_k\}_{(1\leqslant k\leqslant K)}$.

$$\mathbf{V}_1 = \mathbf{A} \cdot \mathbf{D}_1 \cdot \mathbf{B}^T \quad (12)$$
$$\mathbf{V}_2 = \mathbf{A} \cdot \mathbf{D}_2 \cdot \mathbf{B}^T \quad (13)$$
$$\vdots$$
$$\mathbf{V}_K = \mathbf{A} \cdot \mathbf{D}_K \cdot \mathbf{B}^T \quad (14)$$

The best choice for these matrices in order to span the full range of this mapping consists of the K dominant left singular matrices of the mapping in (7) [10]. This $\{\mathbf{V}_k\}$ can thus be obtained from the singular value decomposition of $\mathbf{T}$. In order to deal with these equations in a numerically proper way, the problem can be formulated in terms of orthogonal unknowns [15], [11]. Introducing a QR-factorization $\mathbf{A} = \mathbf{Q}^T\mathbf{R}$ and an RQ-factorization $\mathbf{B}^T = \tilde{\mathbf{R}}\mathbf{Z}^T$, leads to a simultaneous generalized Schur decomposition:

$$\mathbf{Q}\mathbf{V}_1\mathbf{Z} = \mathbf{R} \cdot \mathbf{D}_1 \cdot \tilde{\mathbf{R}} \quad (15)$$
$$\mathbf{Q}\mathbf{V}_2\mathbf{Z} = \mathbf{R} \cdot \mathbf{D}_2 \cdot \tilde{\mathbf{R}} \quad (16)$$
$$\vdots$$
$$\mathbf{Q}\mathbf{V}_K\mathbf{Z} = \mathbf{R} \cdot \mathbf{D}_K \cdot \tilde{\mathbf{R}}. \quad (17)$$

This simultaneous generalized Schur decomposition can be computed by an extended QZ-iteration [15] or by a Jacobi-iteration [11]. With this decomposition, $\mathbf{A}$ and $\mathbf{B}$ are known, and $\mathcal{T}$ is decomposed with a CP model.

Recently, in [8] it is shown that the canonical components can also be obtained from a simultaneous matrix diagonalization with a much less severe restriction on R.

II. Imposing independence based on second order statistics

We now present the algorithm to compute the CP decomposition of a tensor $\mathcal{T}$ where independence is

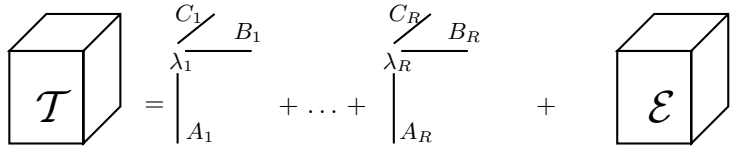

Fig. 1. Pictorial representation of the CP model

imposed to the factors in one mode. For notational convenience, we will restrict us to the three-way real case, but generalization to higher dimensions or complex tensors is straightforward. In the following, we always assume that the components of the last mode are independent. In formulas, we consider the matricized version of the real tensor $\mathcal{T}$, given by equation (8) where matrix $\mathbf{C}$ contains the independent source values, and the mixing matrix $\mathbf{M}$ equals $(\mathbf{A} \odot \mathbf{B})$.

Stack covariance matrices $\mathbf{R}_t(\tau)$ in equation (3) one after the other in a tensor $\mathcal{U}$ with dimensions $IJ \times IJ \times K$, and stack the diagonal matrices $\mathbf{R}_c(\tau)$ in a tensor $\mathcal{P}$.

With a mixing matrix $\mathbf{M} = \mathbf{A} \odot \mathbf{B}$, this tensor $\mathcal{U}$ can be rearranged into a fifth-order tensor with dimensions $I \times J \times I \times J \times K$ with CP structure:

$$\mathcal{U}_t^{(5)} = \sum_{r=1}^{R} A_r \circ B_r \circ A_r \circ B_r \circ R_r \qquad (18)$$

with $R_r = \mathcal{P}(r, r, :)$. In terms of the source correlation matrix $\mathbf{R}_c$, $R_r = [r_{c_{(r,r)}}(\tau_1), \ldots, r_{c_{(r,r)}}(\tau_k)]$. This can be seen as follows.
Define matrices $\mathbf{E_1}, \ldots, \mathbf{E_R} \in \mathbb{R}^{I \times J}$ as

$$(\mathbf{E_r})_{ij} = m_{(i-1)J+j,r} \qquad \forall i, j, r \qquad (19)$$

When the model in equation(8) is exactly satisfied, $\mathbf{E}_r$ is a rank-1 matrix and can be decomposed as

$$(\mathbf{E_r}) = A_r B_r^T \qquad r = 1, \ldots, R \qquad (20)$$

which explains the rearrangement of the tensor from order 3 to 5 in equation (18). Thus $(\mathcal{U}_t^{(5)})_{i1,i2,j1,j2,k} = a_{i1} \cdot b_{j1} \cdot a_{i2} \cdot b_{j2} \cdot (R_r)_k$.

This fifth order CP decomposition can be computed in different ways, depending on the rank R of the tensor $\mathcal{U}_t^{(5)}$. It is even not necessary to compute the full decomposition. Once $\mathbf{A}$ and $\mathbf{B}$ are known, the mixing matrix can be computed. Below, we explain how to compute $\mathbf{A}$ and $\mathbf{B}$ when the rank R is bounded by $R \leqslant \min\{I, J\}$. In a follow-up paper, we will discuss how to compute the decomposition when the rank is higher.

Finally, the independent sources can be estimated from equation 8, and the original tensor is decomposed with a trilinear decomposition with imposed independence.

A. Rank R bounded by $R \leqslant \min\{I, J\}$

In order to compute the mixing matrix $(\mathbf{A} \odot \mathbf{B})$ from (18), associate a matrix $\mathbf{H} \in \mathbb{R}^{IJ \times IJK}$ with $\mathcal{U}_t^{(5)}$ as follows:

$$\mathbf{H}_{(i-1)J+j,(k-1)IJ+(l-1)I+m} = (\mathcal{U}_t^{(5)})_{ijklm} \qquad (21)$$

This mapping can be represented by a matrix $\mathbf{H} \in \mathbb{R}^{IJ \times IJK}$:

$$\mathbf{H} = (\mathbf{A} \odot \mathbf{B}) \cdot (\mathbf{A} \odot \mathbf{B} \odot \mathbf{R})^T. \qquad (22)$$

Substituting (21) in (18) shows that any vector in the range of $\mathbf{H}$ can be represented by an $(I \times J)$ matrix that can be decomposed as:

$$\mathbf{V} = \mathbf{A} \cdot \mathbf{D} \cdot \mathbf{B}^T \qquad (23)$$

with $\mathbf{D}$ diagonal. This is very similar to equation 11. Any matrix in this range can be diagonalized by congruence with the same loading matrices $\mathbf{A}$ and $\mathbf{B}$. A possible choice of $\{\mathbf{V}_k\}_{(1 \leqslant k \leqslant K)}$ consist of 'matrix slices' obtained by fixing the 3rd to 5th index. A basis for the range of $\{\mathbf{V}_k\}$ can be obtained by a singular value decomposition. An optimal approach would be to estimate the R dominant left singular values of (22). The joint decomposition of the matrices $\{\mathbf{V}_k\}_{(1 \leqslant k \leqslant K)}$ will give a set of equations similar to equations (12) - (14). We have explained in §1.3 how to solve these equations simultaneously.

III. Simulation

In this section we illustrate the performance of the algorithm by means of numerical experiments and compare it to the standard CP model, computed with ALS.
Three autocorrelated sources are simulated by generating signals with an autoregressive model of order 5. Rank-3 tensors $\tilde{\mathcal{T}} \in \mathbb{R}^{4 \times 3 \times 100}$, of which the components in the different modes will be estimated afterwards, are generated in the following way:

$$\tilde{\mathcal{T}} = \frac{\mathcal{T}}{||\mathcal{T}||_F} + \sigma \frac{\mathcal{N}}{||\mathcal{N}||_F}, \qquad (24)$$

in which $\mathcal{T}$ exactly satisfies the CP model with the autocorrelated sources in the third mode ($\mathbf{C}$) and $\mathcal{N}$ represents gaussian noise.
In a first simulation, the entries of the two other modes ($\mathbf{A}$ and $\mathbf{B}$) are drawn from a zero-mean unit-variance gaussian distribution and we assess the accuracy when decompositions are computed from rank 3 (correct rank) and rank 4 (overestimation of the rank).
In a second simulation, we evaluated the accuracy of the decomposition when 2 vectors in the same mode are highly correlated (> 0.99). We conduct Monte Carlo simulations consisting of 150 runs. We evaluate the performance of the algorithms by means of the normalized Frobenius norm of the difference between the estimated and the real sources, optimally ordered and scaled:

$$e = \frac{||\mathbf{C} - \hat{\mathbf{C}}||_F}{||\mathbf{C}||_F} \qquad (25)$$

In figures 2 to 4, we plot the mean value of the error for the different simulations. These figures clearly show that the new method outperformed classical CP analysis.

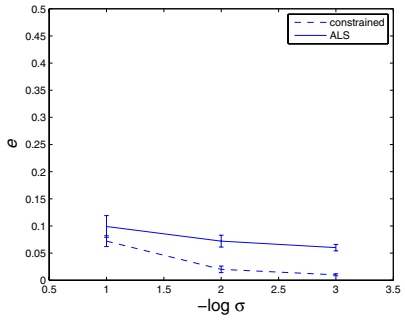

Fig. 2. The mean value of e as a function of the noise level σ for the algorithm with imposed independence (solid) and ALS (dashed) for R=3.

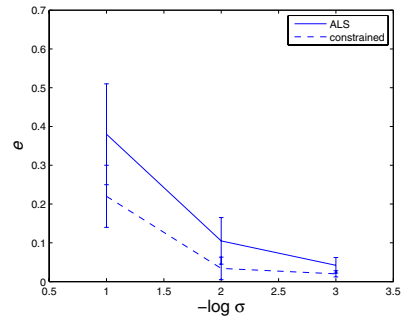

Fig. 3. The mean value of e as a function of the noise level σ for the algorithm with imposed independence (solid) and ALS (dashed) for R=4.

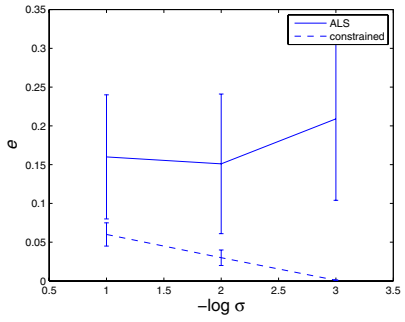

Fig. 4. The mean value of e as a function of the noise level σ for the algorithm with imposed independence (solid) and ALS (dashed) when vectors in one mode are highly correlated.

IV. Discussion

In previous work, we developed an algorithm to impose independence constraints based on higher order statistics in the CP structure [16]. In this work, we propose a new algorithm based on second order statistics. Simulations illustrated the accuracy of the new algorithm and compare it to the accuracy of the standard CP model. The results clearly showed that the new method outperformed classical CP analysis. In particular, the new algorithm is more accurate than the ALS computation of the standard CP model when the rank of the tensor is overestimated and when some components of the trilinear structure are highly correlated. We did not compare with the algorithm presented in ([16]) as it makes different assumptions on the sources.

References

[1] C.F. Beckmann and S.M. Smith. Tensorial extensions of independent component analysis for multisubject fmri analysis. *Neuroimage*, 25:294–311, 2005.

[2] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique using second order statistics. *IEEE Trans Signal Proc*, 45:434–444, 1997.

[3] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non-gaussian signals. *IEE Proc. F*, 140:362–370, 1994.

[4] J.D. Carroll and J. Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of 'Eckart-Young' decomposition. *Psychometrika*, 35:283–319, 1970.

[5] P. Comon. Independent component analysis, a new concept? *Signal Process*, 36:287–314, 1994.

[6] R.A. Harshman. Foundations of the parafac procedure: models and conditions for an 'explanatory' multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 16, 1970.

[7] A. Hyvärinen, J. Karhunen and E. Oja. *Independent Component Analysis*. John Wiley & Sons, 2001.

[8] L. De Lathauwer. A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization. *SIAM J Matrix Anal Appl*, 28:642–666, 2006.

[9] L. De Lathauwer, B. De Moor, and J. Vandewalle. An introduction to independent component analysis. *J Chemometrics*, 14:123–149, 2000.

[10] L. De Lathauwer, B. De Moor, and J. Vandewalle. On the best rank-1 and rank-$\{R_1, R_2, \ldots, R_N\}$ approximation of higher-order tensors. *SIAM J Matrix Anal Appl*, 21:1324–1342, 2000.

[11] L. De Lathauwer, B. De Moor, and J. Vandewalle. Computation of the canonical decomposition by means of a simultaneous generalized schur decomposition. *SIAM J Matrix Anal Appl*, 26:295–327, 2004.

[12] S.E. Leurgans, R.T. Ross, and R.B. Abel. A decomposition for three-way arrays. *SIAM J Matrix Anal Appl*, 14:1064–1083, 1993.

[13] E. Sanchez and B.R. Kowalski. Tensorial resolution: a direct trilinear decomposition. *J Chemometrics*, 4:29–45, 1990.

[14] A. Smilde, R. Bro and P. Geladi. *Multi-way Analysis with applications in the Chemical Sciences*. John Wiley & Sons, 2004.

[15] A.-J. Van Der Veen and A. Paulraj. An analytical constant modulus algorithm. *IEEE Trans Signal Proc*, 44:1136–1155, 1996.

[16] M. De Vos, L. De Lathauwer, and S. Van Huffel. Imposing independence constraints in the CP model. In *Proc. of 7th international conference of independent component analysis*, London, U.K., 2007.