Imposing independence constraints in the CP model *

Maarten De Vos[1], Lieven De Lathauwer[2] and Sabine Van Huffel[1]

[1] ESAT, Katholieke Universiteit Leuven, Leuven, Belgium
[2] CNRS - ETIS, Cergy-Pontoise, France
maarten.devos@esat.kuleuven.be

Abstract. We propose a new algorithm to impose independence constraints in one mode of the CP model, and show with simulations that it outperforms the existing algorithm.

1 Introduction

One of the most fruitful tools in linear algebra-based signal processing is the Singular Value Decomposition (SVD) [4]. Most other important algebraic concepts use the SVD as building block, generalise or refine this concept for analysing quantities that are characterised by only two variables. When the data has an intrinsically higher dimensionality, higher-order generalizations of the SVD can be used. An example of a multi-way decomposition method is the **CP** model (also known as **C**anonical Decomposition (CANDECOMP) [3] or **P**arallel Factor Model (PARAFAC) [5]). Recently, a new interesting concept arose in the biomedical field. In [1], the idea of combining Independent Component Analysis (ICA) and the CP model was introduced. However, the multi-way structure was imposed after the computation of the independent components. In this paper, we propose an algorithm to impose the CP structure during the ICA computation. We also performed some numerical experiments to compare our algorithms to the algorithm proposed in [1].

* This research is funded by a PhD grant 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; **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), research communities (ICCoS, ANMMM); IWT: PhD Grants; **Belgian Federal Science Policy Office** IUAP P6/04 ('Dynamical systems, control and optimization', 2007-2011); **EU**: BIOPATTERN (FP6-2002-IST 508803), ETUMOUR (FP6-2002-LIFESCIHEALTH 503094), Healthagents (IST-2004-27214), FAST (FP6-MC-RTN-035801); **ESA**: Cardiovascular Control (Prodex-8 C90242)

1.1 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 outerproduct 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.
Definition 3. The *Kruskal rank* or *k-rank* of a matrix is the maximal number r such that any set of r columns of the matrix is linearly independent.
Definition 4. The Frobenius norm of a tensor $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$ is defined as

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

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.

1.2 Independent Component Analysis

Assume the basic linear statistical model

$$Y = \mathbf{M} \cdot X + N \tag{2}$$

where $Y \in \mathbb{R}^I$ is called the observation vector, $X \in \mathbb{R}^J$ the source vector and $N \in \mathbb{R}^I$ additive noise. $\mathbf{M} \in \mathbb{R}^{I \times J}$ is the mixing matrix.

The goal of Independent Component Analysis is to estimate the mixing matrix $\mathbf{M}$, and/or the source vector X, given only realizations of Y. In this study, we assume that $I \geqslant J$.
Blind identification of $\mathbf{M}$ in (2) is only possible when some assumptions about the sources are made. One assumption is that the sources are mutually statistically independent, as well as independent from the noise components and that at most one source is gaussian [2].

For more details, we refer to [9, 6].

1.3 The CP Model

The model The CP model [5, 3, 15] 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:

$$\mathcal{T} = \sum_{r=1}^{R} \lambda_r \, A_r \circ B_r \circ C_r \, (+\mathcal{E}) \tag{3}$$

A pictorial representation of the CP model for third-order tensors is given in figure 1.

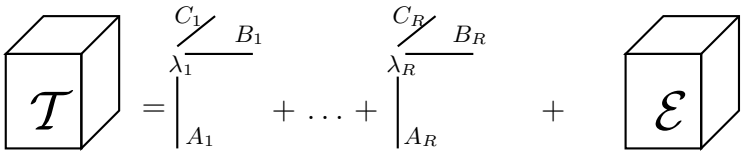

Fig. 1. Pictorial representation of the CP model

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

$$t_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr}, \qquad \forall i, j, k \tag{4}$$

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 is given with 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:

$$(\mathbf{T})_{(i-1)J+j,k} = \mathcal{T}_{ijk}. \tag{5}$$

This matrix has following structure:

$$\mathbf{T} = (\mathbf{A} \odot \mathbf{B}) \cdot \mathbf{C}^T. \tag{6}$$

Comparing the number of free parameters of a generic tensor and the CP model, it can be seen that this model is very restricted. The advantage of this model is its uniqueness under mild conditions [7, 14]:

$$\text{rank}_k(\mathbf{A}) + \text{rank}_k(\mathbf{B}) + \text{rank}_k(\mathbf{C}) \geqslant 2R + 2 \tag{7}$$

with $\text{rank}_k(\mathbf{A})$ the k-rank of matrix $\mathbf{A}$ and R the rank of the tensor.

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:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} ||\mathbf{T} - \mathbf{A}(\mathbf{B} \odot \mathbf{C})^T||^2. \tag{8}$$

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:

$$\min_{\mathbf{A}} ||\mathbf{X} - \mathbf{A}\mathbf{Z}^T||^2 \tag{9}$$

where $\mathbf{Z}$ equals $\mathbf{B} \odot \mathbf{C}$. This has to be repeated until convergence while matrices in other modes are kept fixed in order to compute all factors of the decomposition.
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 restricting 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. Substitution of (5) in (4) shows that any vector in the range of $\mathbf{T}$, can be represented by an $I \times J$ matrix that can be decomposed as:

$$\mathbf{V} = \mathbf{A} \cdot \mathbf{D} \cdot \mathbf{B}^T \tag{10}$$

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 $\{\mathbf{V}_k\}_{(1 \leqslant k \leqslant K)}$.

$$\mathbf{V}_1 = \mathbf{A} \cdot \mathbf{D}_1 \cdot \mathbf{B}^T \tag{11}$$
$$\mathbf{V}_2 = \mathbf{A} \cdot \mathbf{D}_2 \cdot \mathbf{B}^T \tag{12}$$
$$\vdots$$
$$\mathbf{V}_K = \mathbf{A} \cdot \mathbf{D}_K \cdot \mathbf{B}^T \tag{13}$$

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 (5) [10]. In order to deal with these equations in a numerical proper way, the problem can be formulated in terms of orthogonal unknowns [17, 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:

$$\mathbf{Q}\mathbf{V}_1\mathbf{Z} = \mathbf{R} \cdot \mathbf{D}_1 \cdot \tilde{\mathbf{R}} \tag{14}$$
$$\mathbf{Q}\mathbf{V}_2\mathbf{Z} = \mathbf{R} \cdot \mathbf{D}_2 \cdot \tilde{\mathbf{R}} \tag{15}$$
$$\vdots$$
$$\mathbf{Q}\mathbf{V}_K\mathbf{Z} = \mathbf{R} \cdot \mathbf{D}_K \cdot \tilde{\mathbf{R}}. \tag{16}$$

This simultaneous generalized Schur decomposition can be computed by an extended QZ-iteration [17].
Recently, in [8] it is shown that the canonical components can be obtained from a simultaneous matrix diagonalization with a much less severe restriction on R.

2 Combination of ICA and CP model

In this section, we review the tensorial extension of ICA (§2.1), called 'tensor pICA', as it was introduced by Beckmann [1]. Then we present a new algorithm to compute the CP decomposition of a tensor $\mathcal{T}$ where independence is imposed to the factors in one mode (§2.2). 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 third mode are independent. Due to the symmetric structure of the PARAFAC model, equivalent equations can be derived for the other two modes. In formulas, we consider the matricized version of the real tensor $\mathcal{T}$, given by equation (6) where matrix $\mathbf{C}$ contains the independent source values, and the mixing matrix $\mathbf{M}$ equals $(\mathbf{A} \odot \mathbf{B})$.

2.1 Tensor pICA

In [1], a generalization of the standard bilinear (two-way) exploratory analysis to higher dimensions was derived as follows.

1. Perform an iteration step for the decomposition of the full data using the twodimensional probabilistic ICA approach for the decomposition into a compound mixing matrix $\mathbf{M}^{IJ\times R}$ and the associated source signals $\mathbf{C}^{K\times R}$: $\mathbf{X}^{IJ\times K} = \mathbf{M}\mathbf{C}^T + \tilde{\mathbf{E}}_1$.
2. Decompose the estimated mixing matrix $\mathbf{M}$ such that $\mathbf{M} = (\mathbf{A} \odot \mathbf{B}) + \tilde{\mathbf{E}}_2$ via a column-wise rank-1 eigenvalue decomposition: each column in $(\mathbf{A} \odot \mathbf{B})$ is formed by K scaled repetitions of a single column from $\mathbf{A}$. In order to obtain $\mathbf{A}$ and $\mathbf{B}$, the matrices $\mathbf{G_1}, \ldots, \mathbf{G_R} \in \mathbb{R}^{I\times J}$ can be introduced as
$$(\mathbf{G_r})_{ij} = m_{(i-1)J+j,r} \qquad \forall i, j, r \tag{17}$$
$\mathbf{A}_r$ and $\mathbf{B}_r$ can then be computed as the dominant left and right singular vector of $\mathbf{G_r}, 1 \leqslant r \leqslant R$.
3. iterate decomposition of $\mathbf{X}^{IJ\times K}$ and $\mathbf{M}$ untill convergence, i.e. when $||\mathbf{A}^{\text{new}} - \mathbf{A}^{\text{old}}||_F + ||\mathbf{B}^{\text{new}} - \mathbf{B}^{\text{old}}||_F + ||\mathbf{C}^{\text{new}} - \mathbf{C}^{\text{old}}||_F < \epsilon$.

2.2 ICA-CP

The ordinary ICA problem is solved by diagonalising the fourth-order cumulant[9]. This cumulant can be written as following CP decomposition:

$$\mathcal{C}_y^{(4)} = \sum_{r=1}^{R} \kappa_{x_r} M_r \circ M_r \circ M_r \circ M_r \tag{18}$$

With a mixing matrix $\mathbf{M} = \mathbf{A} \odot \mathbf{B}$, this fourth-order cumulant can be expressed as an eighth-order tensor with CP structure:

$$\mathcal{C}_y^{(8)} = \sum_{r=1}^{R} \kappa_{x_r} A_r \circ B_r \circ A_r \circ B_r \circ A_r \circ B_r \circ A_r \circ B_r \tag{19}$$

This can be seen as follows.
Define matrices $\mathbf{E_1}, \ldots, \mathbf{E_R} \in \mathbb{R}^{I \times J}$ as

$$(\mathbf{E_r})_{ij} = m_{(i-1)J+j,r} \qquad \forall i, j, r \tag{20}$$

When the model in (6) is exactly satisfied, $\mathbf{E}_r$ can be decomposed as

$$(\mathbf{E_r}) = A_r B_r^T \qquad r = 1, \ldots, R \tag{21}$$

which explains the CP decomposition in equation (19).

This CP decomposition can be computed in different ways, depending on the rank R of the tensor. It is even not necessary to compute the full decomposition. Once $\mathbf{A}$ and $\mathbf{B}$ are known, the mixing matrix (matrix $\mathbf{A} \odot \mathbf{B}$) can be computed and the independent sources can be estimated from equation (6).

Rank R restricted by $R \leqslant \min\{I, J\}$ In order to compute the mixing matrix $(\mathbf{A} \odot \mathbf{B})$ from (19), associate a matrix $\mathbf{H} \in \mathbb{R}^{IJ \times I^3 J^3}$ with $\mathcal{C}_y^{(8)}$ as follows:

$$\mathbf{H}_{(i-1)J+j,(k-1)I^2J^3+(l-1)I^2J^2+(m-1)IJ^2+(n-1)IJ+(o-1)J+p} = (\mathcal{C}_y^{(8)})_{ijklmnop} \tag{22}$$

This mapping can be represented by a matrix $\mathbf{H} \in \mathbb{R}^{IJ \times I^3 J^3}$:

$$\mathbf{H} = (\mathbf{A} \odot \mathbf{B}) \cdot \mathbf{\Lambda} \cdot (\mathbf{A} \odot \mathbf{B} \odot \mathbf{A} \odot \mathbf{B} \odot \mathbf{A} \odot \mathbf{B})^T. \tag{23}$$

with $\mathbf{\Lambda} = diag\{\kappa_1, \ldots, \kappa_R\}$. Substituting (22) in (19) shows that any vector in the range of $\mathcal{C}_y^{(8)}$ can be represented by an $(I \times J)$ matrix that can be decomposed as:

$$\mathbf{V} = \mathbf{A} \cdot \mathbf{D} \cdot \mathbf{B}^T \tag{24}$$

with $\mathbf{D}$ diagonal. 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 8th index. An optimal approach would be to estimate the R dominant left singular values of (23). The joint decomposition of the matrices $\{\mathbf{V}_k\}_{(1 \leqslant k \leqslant K)}$ will give a set of equations similar to equations (11) - (13). We have explained in §1.3 how to solve these equations simultaneously.

3 Numerical experiments

In this section we illustrate the performance of our algorithm by means of numerical experiments and compare it to 'tensor pICA'.
Rank-R tensors $\tilde{\mathcal{T}} \in \mathbb{R}^{5 \times 3 \times 100}$, of which the components in the different modes will be estimated afterwards, are generated in the following way:

$$\tilde{\mathcal{T}} = \frac{\mathcal{T}}{||\mathcal{T}||_F} + \sigma_N \frac{\mathcal{N}}{||\mathcal{N}||_F}, \tag{25}$$

in which $\mathcal{T}$ exactly satisfies the CP model with $R = 3$ independent sources in the third mode ($\mathbf{C}$) and $\mathcal{N}$ represents gaussian noise. All the source distributions are binary (1 or -1), with an equal probability of both values. The sources are zero mean and have unit variance. The entries of the two other modes ($\mathbf{A}$ and $\mathbf{B}$) are drawn from a zero-mean unit-variance gaussian distribution.
We conduct Monte Carlo simulations consisting of 500 runs. We evaluate the performance of the different algorithms by means of the normalized Frobenius norm of the difference between the estimated and the real sources:

$$error_C = \frac{||\mathbf{C} - \hat{\mathbf{C}}||_F}{||\mathbf{C}||_F} \tag{26}$$

In figure 2, we plot the mean value of the 500 simulations. The previously

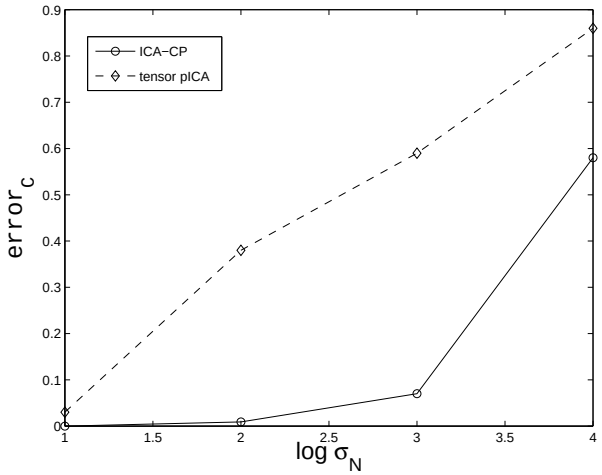

Fig. 2. The mean value of $error_C$ as a function of the noise level σ_N for the algorithms ICA-CP (solid) and tensor pICA (dash-dash).

proposed method tensor pICA is clearly outperformed by the new algorithm.

4 Conclusion

We proposed a new algorithm to impose the CP structure already during the ICA computation for the case that the rank R was restricted by $R \leqslant \min\{I, J\}$. We showed with simulations that by taking this structure into account, the algorithm outperformed tensor pICA. A follow-up paper will discuss an algorithm for the case the rank $R \geqslant \min\{I, J\}$. For a detailed comparison between CP and the combination of ICA-CP, we refer to [16].

Bibliography

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

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

[3] 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.

[4] G.H. Golub and C.F. Van Loan. *Matrix computations*. Third Edition The Johns Hopkins University Press, 1996.

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

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

[7] J.B. Kruskal. Three-way arrays: rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics. *Psychometrika*, 18:95–138, 1977.

[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] N.D. Siridopoulos and R. Bro. On the uniqueness of multilinear decomposition of n-way arrays. *J Chemometrics*, 14:229–239, 2000.

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

[16] A. Stegeman. Comparing independent component analysis and the parafac model for artificial multi-subject fmri data. *Technical Report, Heymans Institute, University of Groningen, the Netherlands*, 2007.

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