# Tensor Decompositions: New Concepts in Brain Data Analysis ?

Andrzej CICHOCKI[†a)],

**SUMMARY**   Matrix factorizations and their extensions to tensor factorizations and decompositions have become prominent techniques for linear and multilinear blind source separation (BSS), especially multiway Independent Component Analysis (ICA), Nonnegative Matrix and Tensor Factorization (NMF/NTF), Smooth Component Analysis (SmoCA) and Sparse Component Analysis (SCA). Moreover, tensor decompositions have many other potential applications beyond multilinear BSS, especially feature extraction, classification, dimensionality reduction and multiway clustering. In this paper, we briefly overview new and emerging models and approaches for tensor decompositions in applications to group and linked multiway BSS/ICA, feature extraction, classification and Multiway Partial Least Squares (MPLS) regression problems.
*key words:  Multilinear BSS, linked multiway BSS/ICA, tensor factorizations and decompositions, constrained Tucker and CP models, Penalized Tensor Decompositions (PTD), feature extraction, classification, multiway PLS and CCA.*

## 1.   Introduction

Although the basic models for tensor (i.e., multiway array) decompositions and factorizations such as Tucker and (Canonical-decomposition/Parafac) CP models were proposed long time ago, they are recently emerging as promising tools for exploratory analysis of multidimensional data in diverse applications, especially, in multiway blind source separation (BSS), feature extraction, classification, prediction, and multiway clustering [1]–[5]. The recent advances in neuroimage technologies (e.g., high density array EEG/MEG, fMRI, NIRS) have generated massive amounts of brain data exhibiting high dimensionality, multiple modality and multiple couplings, functional connectivity. By virtue of their multiway nature, tensors provide a powerful tools for analysis and fusion of such massive data together with a mathematical backbone for the discovery of underlying hidden complex data structures [1], [6], [7]. Constrained Matrix Factorizations, called sometimes penalized matrix decompositions (e.g., ICA, SCA, NMF, SVD/PCA, CCA) and their extensions or penalized tensor decompositions - multidimensional constrained models (Tucker, CP, NTF, NTD models [1], [6]–[8]), with some constraints such as statistical independence, decorrelation, orthogonality, sparseness, nonnegativity, and smoothness - have been recently proposed as meaningful and efficient representations of signals, images and in general natural multidimensional data [1]. From signal processing and data anal-

ysis point of view, tensor decompositions are very attractive because they take into account spatial, temporal and spectral information and provide links among various data and extracted factors or hidden (latent variables) components with physical or physiological meaning and interpretations [1], [8], [10]–[12]. In fact, tensor decompositions are emerging techniques for data fusion, dimensionality reduction, pattern recognition, object detection, classification, multiway clustering, sparse representation and coding, and multilinear blind source separation (MBSS) [1], [2], [13]–[15].

**Basic notations.** Tensors (i.e., multiway arrays) are denoted by underlined capital boldface letters, e.g., $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. The order of a tensor is the number of modes, also known as ways or dimensions (e.g., space, time, frequency, subjects, trials, classes, groups, conditions). In contrast, matrices (two-way tensors) are denoted by boldface capital letters, e.g., $\mathbf{Y}$; vectors (one-way tensors) are denoted by boldface lowercase letters, e.g., columns of the matrix $\mathbf{U}$ by $\mathbf{u}_j$ and scalars are denoted by lowercase letters, e.g., $u_{ij}$.

The mode-$n$ product $\underline{\mathbf{Y}} = \underline{\mathbf{G}} \times_n \mathbf{U}$ of a tensor $\underline{\mathbf{G}} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$ and a matrix $\mathbf{U} \in \mathbb{R}^{I \times J_n}$ is a tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{J_1 \times \cdots \times J_{n-1} \times I \times J_{n+1} \times \cdots \times J_N}$, with elements $y_{j_1, j_2, \ldots, j_{n-1}, i_n, j_{n+1}, \ldots, j_N} = \sum_{j_n=1}^{J_n} g_{j_1, j_2, \ldots, j_N} u_{i_n, j_n}$. Unfolding (matricization, flattening) of a tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ in $n$-mode is denoted as $\mathbf{Y}_{(n)} \in \mathbb{R}^{I_n \times (I_1 \cdots I_{n-1} I_{n+1}, \cdots I_N)}$, which consists of arranging all possible $n$-mode tubes in as columns of a matrix. [2]. Throughout this paper, standard notations and basic tensor operations are used [1].

## 2.   Basic Models for Multilinear BSS/ICA

Most of the linear blind source separation (BSS) models can be represented as constrained matrix factorization problems, with suitable constraints imposed on factor matrices (or their columns -referred to as components)

$$\mathbf{Y} = \mathbf{A}\mathbf{B}^T + \mathbf{E} = \sum_{j=1}^{J} \mathbf{a}_j \mathbf{b}_j^T + \mathbf{E} = \sum_{j=1}^{J} \mathbf{a}_j \circ \mathbf{b}_j + \mathbf{E}, \quad (1)$$

where $\circ$ denotes outer product[*], $\mathbf{Y} = [y_{it}] \in \mathbb{R}^{I \times T}$ is a known data matrix (representing observations or measurements), $\mathbf{E} = [e_{it}] \in \mathbb{R}^{I \times T}$ represents errors or noise, $\mathbf{A} =$

---

[†]RIKEN Brain Science Institute, 2-1 Hirosawa, Wako, 351-0198 Saitama, JAPAN and Warsaw University of Technology and Systems Research Institute, Polish Academy of Science, POLAND.
   a) E-mail: a.cichocki@riken.jp

[*]The outer product of two vectors $\mathbf{a} \in \mathbb{R}^I$, $\mathbf{b} \in \mathbb{R}^T$ builds up a rank-one matrix $\mathbf{Y} = \mathbf{a} \circ \mathbf{b} = \mathbf{a}\mathbf{b}^T \in \mathbb{R}^{I \times T}$ and the outer product of three vectors: $\mathbf{a} \in \mathbb{R}^I$, $\mathbf{b} \in \mathbb{R}^T$, $\mathbf{c} \in \mathbb{R}^Q$ builds up a third-order rank-one tensor: $\underline{\mathbf{Y}} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \in \mathbb{R}^{I \times T \times Q}$, with entries defined as $y_{itq} = a_i b_t c_q$.

$[a_{ij}] = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_J] \in \mathbb{R}^{I \times J}$ is an unknown basis (mixing) matrix, with basis vectors $\mathbf{a}_j \in \mathbb{R}^I$ and the $J$ columns of a matrix $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_J] \in \mathbb{R}^{T \times J}$ represent unknown components. latent variables or sources $\mathbf{b}_j$.

**Remark**:We notice that we have symmetry in the factorization: For Eq (1) we could just as easily write $\mathbf{Y}^T \cong \mathbf{B}\mathbf{A}^T$, so the meaning of sources and mixture are often somewhat arbitrary.

Our primary objective in the BSS is to estimate uniquely (neglecting unavoidable scaling and permutation ambiguities) factor matrices $\mathbf{A}$ and $\mathbf{B}$, subject to various specific constraints imposed on the vectors $\mathbf{b}_j$ and/or $\mathbf{a}_j$, such as mutual statistical independence (ICA), sparseness (SCA), smoothness (SmoCA), nonnegativity (NMF) or orthogonality (PCA/SVD), uncorrelation, etc.

In some applications the data matrix $\mathbf{Y}$ is factorized into three or more factors [1]. In the special case, of a singular value decomposition (SVD) of the data matrix $\mathbf{Y} \in \mathbb{R}^{I \times T}$, we have the following factorization:

$$\mathbf{Y} = \mathbf{A}\mathbf{D}\mathbf{B}^T = \mathbf{D} \times_1 \mathbf{A} \times_2 \mathbf{B} = \sum_j d_{jj} \mathbf{a}_j \mathbf{b}_j^T, \qquad (2)$$

where $\mathbf{A} \in \mathbb{R}^{I \times I}$ and $\mathbf{B} \in \mathbb{R}^{T \times T}$ are orthogonal matrices and $\mathbf{D}$ is a diagonal matrix containing only nonnegative singular values. The SVD and its generalizations play key roles in signal processing and data analysis [15].

Often multiple subject, multiple task data sets can be represented by a set of data matrices $\mathbf{Y}_n$ and it is necessary to perform simultaneous constrained matrix factorizations[†]:

$$\mathbf{Y}_n \cong \mathbf{A}_n \mathbf{B}_n^T \qquad (n = 1, 2, \ldots, N) \qquad (3)$$

or in preprocessed form with a dimensionally reduction

$$\tilde{\mathbf{Y}}_n \cong \mathbf{Q}_n \tilde{\mathbf{A}}_n \mathbf{P}_n \tilde{\mathbf{B}}_n^T, (n = 1, 2, \ldots, N), \qquad (4)$$

subject to various additional constraints (e.g., $\mathbf{B}_n = \mathbf{B}$, $\forall n$ and their columns are mutually independent and/or sparse). This problem is related to various models of group ICA, with suitable pre-processing, dimensionality reduction and post-processing procedures [16]–[18]. In this paper, we introduce the group multiway BSS concept, which is more general and flexible than the group ICA, since various constraints can be imposed on factor matrices in different modes (i.e., not only mutual independence but also nonnegativity, sparseness, smoothness or orthogonality). There is neither a theoretical nor an experimental basis that statistical independence (ICA) is the unique right concept to extract brain sources or isolate brain networks [6]. In real world scenarios, latent (hidden) components (e.g., brain sources) have various complex properties and features. In other words, true unknown involved components are seldom all statistically independent. Therefore, if we apply only one single criterion like ICA, we may fail to extract all desired components with physical interpretation. We need rather to apply

an ensemble of strategies by employing several suitably chosen criteria and associated learning algorithms to extract all desired components [1], [6]. For these reasons, we have developed multiway BSS methods, which are based not only on the statistical independence of components, but rather exploits multiple criteria or diversities of factor matrices in various modes, in order to extract physiologically meaningful components, with specific features and statistical properties [1], [6]. By diversity, we mean different characteristics, features or morphology of source signals or hidden latent variables [7]. Since multi-array data can be always interpreted in many different ways, some *a priori* knowledge is needed to determine, which diversities, characteristics, features or properties represent true latent (hidden) components with physical meaning.

It should be noted, although standard 2D BSS (constrained matrix factorizations) approaches, such as ICA, NMF, SCA, SmoCA, PCA/SVD, and their variants, are invaluable tools for feature extraction and selection, dimensionality reduction, noise reduction, and data mining, they have only two modes or 2-way representations (typically, space and time), and their use is therefore limited. In many neuroscience applications the data structures often contain higher-order ways (modes) such as subjects, groups, trials, classes and conditions, together with the intrinsic dimensions of space, time, and frequency. For example, studies in neuroscience often involve multiple subjects (people or animals) and trials leading to experimental data structures conveniently represented by multiway arrays or blocks of multiway data.

The simple linear BSS models (1)-(3) can be naturally extended for multidimensional data to the multiway BSS models using constrained tensor decompositions. In this paper, we consider a general and flexible approach based on the Tucker decomposition model - called also Tucker-N model (see Fig. 1 (a)) [1], [2], [5]:

$$\begin{aligned}
\underline{\mathbf{Y}} &= \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \cdots \sum_{j_N=1}^{J_N} g_{j_1 j_2 \cdots j_N} \left( \mathbf{u}_{j_1}^{(1)} \circ \mathbf{u}_{j_2}^{(2)} \circ \cdots \circ \mathbf{u}_{j_N}^{(N)} \right) + \underline{\mathbf{E}} \\
&= \underline{\mathbf{G}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \cdots \times_N \mathbf{U}^{(N)} + \underline{\mathbf{E}} \\
&= \underline{\mathbf{G}} \times \{\mathbf{U}\} + \underline{\mathbf{E}} = \widehat{\underline{\mathbf{Y}}} + \underline{\mathbf{E}}, \qquad (5)
\end{aligned}$$

where $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_N}$ is the given data tensor, $\underline{\mathbf{G}} \in \mathbb{R}^{J_1 \times J_2 \cdots \times J_N}$ is a core tensor of reduced dimension, $\mathbf{U}^{(n)} = [\mathbf{u}_1^{(n)}, \mathbf{u}_2^{(n)}, \ldots, \mathbf{u}_{J_n}^{(n)}] \in \mathbb{R}^{I_n \times J_n}$ ($n = 1, 2, \ldots, N$) are factors (component matrices) representing components, latent variables, common factors or loadings, $\widehat{\underline{\mathbf{Y}}}$ is an approximation of the measurement $\underline{\mathbf{Y}}$, and $\underline{\mathbf{E}}$ denotes the approximation error or noise depending on the context [1], [2], [5]. The objective is to estimate factor matrices: $\mathbf{U}^{(n)}$, with components (vectors) $\mathbf{u}_{j_n}^{(n)}$, ($n = 1, 2, \ldots, N$, $j_n = 1, 2, \ldots, J_N$) and the core tensor $\underline{\mathbf{G}} \in \mathbb{R}^{J_1 \times J_2 \cdots \times J_N}$ assuming that the number of factors in each mode $J_n$ are known or can be estimated [1].

If the factor matrices and a core tensor are orthogonal the Tucker model can be considered as extension of the SVD model (2), known as the High Order SVD (HOSVD)

---

[†]This form of factorizations are typical with EEG/MEG related data for multi-subjects, multi-tasks, while the factorization for the transposed of $\mathbf{Y}$ is typical for fMRI data [16]–[18].
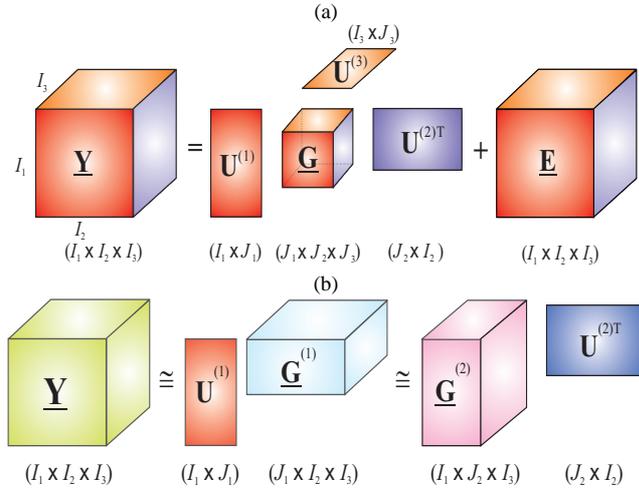
**Fig. 1** Illustration of a 3-way tensor decomposition using a constrained Tucker-3 model; Objective is to estimate factor matrices $\mathbf{U}^{(n)} = [\mathbf{u}_1^{(n)}, \mathbf{u}_2^{(n)}, \ldots, \mathbf{u}_{J_n}^{(n)}] \in \mathbb{R}^{I_n \times J_n}$ (with desired diversities or statistical properties) and a possibly sparse core tensor $\underline{\mathbf{G}} \in \mathbb{R}^{J_1 \times J_2 \times J_3}$, typically with $J_n \ll I_n$ ($n =, 1, 2, 3$). (b) Preselected factor matrices in Tucker-3 can be absorbed by a core tensor $\underline{\mathbf{G}}$, and this leads to Tucker-1 models: $\underline{\mathbf{Y}} = \underline{\mathbf{G}}^{(1)} \times_1 \mathbf{U}^{(1)} + \underline{\mathbf{E}}^{(1)} = \underline{\mathbf{G}}^{(2)} \times_2 \mathbf{U}^{(2)} + \underline{\mathbf{E}}^{(2)} = \underline{\mathbf{G}}^{(3)} \times_3 \mathbf{U}^{(3)} + \underline{\mathbf{E}}^{(3)}$. Instead of applying the standard Alternating Least Squares (ALS) algorithms for the Tucker-3 model, we can apply unfolding of data tensor according to the Tucker-1 models and then perform constrained factorizations of the unfolded matrices (multiway BSS) by imposing desired constraints (independence, sparseness, nonnegativity, smoothness or uncorrelation, etc.)

[15]. While the optimal approximation of a matrix can be obtained by truncation of its SVD, the optimal tensor approximation (with a minimal norm of the error) cannot in general be obtained by truncation of the Tucker decomposition. However, it was shown that the truncation of the particular constrained version usually yields a quite good approximation [15]. In general, in the Tucker model the orthogonality constraints are usually not imposed. Instead, we consider alternative constraints, such as sparseness, nonnegativity, smoothness or statistical mutual independence. It should be noted that the Tucker model is not unique in general, even if we impose some weak constraints. However, this model is unique if we impose suitable constraints (e.g., sparseness or independence).

This leads to a concept of group multiway BSS. There are two possible interpretations or concepts of employing Tucker decomposition as multiway BSS. In the first concept the columns of factor matrices $\mathbf{U}^{(n)}$ represent desired components or latent variables and the core tensor represent some "mixing process" or more precisely the core tensor shows links among components in different modes, while data tensor $\underline{\mathbf{Y}}$ represents collection of 1-D or 2-D mixing signals. In the second concepts, the core tensor represents desired but unknown (hidden) $N$-dimensional signal (e.g., 3D MRI image or 4D video) and factor matrices represent various transformations, e.g., time frequency transformation or wavelets dictionaries (mixing or filtering processes), while a data tensor $\underline{\mathbf{Y}}$ represents observed $N$-dimensional signal which is distorted, transformed compressed or mixed depending on applications. In this paper, we will consider only

the first interpretation.

The Tucker-$N$ model (5) can be represented by $n$ approximative matrix factorizations with three factors:

$$\mathbf{Y}_{(n)} \cong \mathbf{U}^{(n)} \mathbf{G}_{(n)} \mathbf{Z}^{(n)}, \qquad (n = 1, 2, \ldots, N), \qquad (6)$$

where $\mathbf{Z}^{(n)} = \left[ \mathbf{U}^{(N)} \otimes \cdots \otimes \mathbf{U}^{(n+1)} \otimes \mathbf{U}^{(n-1)} \cdots \otimes \mathbf{U}^{(1)} \right]^T$.

Moreover, the Tucker-$N$ model (5) can be compressed to $N$ Tucker-1 models, with only one factor matrix $\mathbf{U}^{(n)}$, in each mode $n$ (see Fig 1 (b)):

$$\underline{\mathbf{Y}} \cong \underline{\mathbf{G}}^{(n)} \times_n \mathbf{U}^{(n)} \text{ or in matrix form } \mathbf{Y}_{(n)} \cong \mathbf{U}^{(n)} \mathbf{G}_{(n)}^{(n)}, \quad (7)$$

where $\underline{\mathbf{G}}^{(n)} = \underline{\mathbf{G}} \times_1 \mathbf{U}^{(1)} \times_2 \cdots \times_{n-1} \mathbf{U}^{(n-1)} \times_{n+1} \mathbf{U}^{(n+1)} \cdots \times_N \mathbf{U}^{(N)}$, $(n = 1, 2, \ldots, N)$. Note that the above models correspond to group BSS/ICA models (3), with $\mathbf{Y}_n = \mathbf{Y}_{(n)}^T$, $\mathbf{A}_n = [\mathbf{G}_{(n)}^{(n)}]^T$ and $\mathbf{B}_n = \mathbf{U}^{(n)}$, $\forall n$, under the assumption that we impose desired constraints for factor matrices $\mathbf{U}^{(n)}$.

Most of the existing approaches exploit only the CP model and impose only statistical independence constraints This leads to tensor probabilistic ICA [19] or ICA-CPA models [22], [23]. However, our approaches are quite different, because we use the Tucker models and we are not restricting only to ICA assumptions but exploit multiple criteria and allows for diversities of components. The advantage of the Tucker model over the CP model is that the Tucker model is more general and the number of components in each mode can be different and furthermore the components are linked via a core tensor, and hence allows us to model more complex hidden data structures. Note that the Tucker decomposition can be simplified to the CP model in the special case, where the "hyper-cube" core tensor (with $J = J_1 = J_2 = \cdots = J_N$) has nonzero elements only on the super-diagonal. The CP model has unique factorization without any constraints under some mild conditions.

For the Tucker and the CP models there exist many efficient algorithms [1], [2]. Most of them are based on the ALS (Alternating Least Squares) and HALS (Hierarchical ALS) [1], [8], [11], [24] and CUR (Column-Row) decompositions [25]. However, description of these algorithms is out of the scope of this paper.

We can implement Multilinear BSS algorithms in several ways. First of all, we can minimize a global cost function, with suitable penalty and/or regularization terms to estimate desired components (see Eq. (5)):

$$D_F \left( \underline{\mathbf{Y}} \| \underline{\mathbf{G}}, \{\mathbf{U}\} \right) = \| \underline{\mathbf{Y}} - \underline{\mathbf{G}} \times \{\mathbf{U}\} \|_F^2 + \sum_n \alpha_n C_n(\mathbf{U}^{(n)}), (8)$$

where $\alpha_n \geq 0$ are penalty coefficients and $C_n(\mathbf{U}^{(n)})$ are penalty terms, which are added to achieve specific characterstic of the components. For example, if we need to impose mutual independence constraints the penalty terms can take the following form $C_n(\mathbf{U}^{(n)}) = \sum_{n=1}^{N} \sum_{j \neq p} \mathbf{u}_j^{(n)T} \psi_p(\mathbf{u}_n^{(n)})$, where $\psi_n(u)$ are suitable nonlinear functions. In principle, this method referred as penalized tensor decompositions, allows to find the factor matrices $\mathbf{U}^{(n)}$, with unique components $\mathbf{u}_{j_n}^{(n)}$ and an associated core tensor but the method involves heavy computations and it is time consuming.

Another approach is to apply standard Tucker decompositions method, without applying any desired constraints using ALS [1], CUR [25] or HOOI/HOSVD [15], [20] algorithms and in the next step apply for each factor matrix standard constrained matrix factorization algorithms (e.g., ICA, NMF or SCA) [21], [26]. Assuming that each unconstrained factor $\mathbf{U}^{(n)}$ in the model (5) can be further factorized using standard BSS algorithms as $\mathbf{U}^{(n)} \cong \tilde{\mathbf{B}}_n \tilde{\mathbf{A}}_n^T$, we can formulate the following decomposition model:

$$\underline{\mathbf{Y}} \cong \underline{\mathbf{G}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \cdots \times_N \mathbf{U}^{(N)} \tag{9}$$

$$\cong \underline{\tilde{\mathbf{G}}} \times_1 \tilde{\mathbf{B}}_1 \times_2 \tilde{\mathbf{B}}_2 \cdots \times_N \tilde{\mathbf{B}}_N, \tag{10}$$

where $\underline{\tilde{\mathbf{G}}} = \left[ \underline{\mathbf{G}} \times_1 \tilde{\mathbf{A}}_1^T \times_2 \tilde{\mathbf{A}}_2^T \cdots \times_N \tilde{\mathbf{A}}_N^T \right]$. It is worth to note, that in each mode, we can apply different criteria for matrix factorization.

Alternatively, a simpler approach is to perform the unfolding the data tensor $\underline{\mathbf{Y}}$ for each mode $n$, according to the Tucker-1 decompositions (7) and to apply directly a suitable constrained factorization of matrices (or a penalized matrix decomposition) †

$$\mathbf{Y}_{(n)} \cong \mathbf{B}_n \mathbf{A}_n^T \quad \text{or} \quad \mathbf{Y}_{(n)}^T \cong \mathbf{A}_n \mathbf{B}_n^T, \quad (n = 1, \ldots, N) \tag{11}$$

subject to desired constraints, by employing standard efficient BSS algorithms (e.g., NMF or ICA) [1], [26]. Since some matrices $\mathbf{Y}_{(n)}$ may have large dimensions, we usually need to apply some efficient methods for dimensionality reduction [26]. Finally, the core tensor, which shows the links among components in different modes can be computed as

$$\underline{\tilde{\mathbf{G}}} = \underline{\mathbf{Y}} \times_1 [\mathbf{B}_1]^+ \times_2 [\mathbf{B}_2]^+ \cdots \times_N [\mathbf{B}_N]^+, \tag{12}$$

where $[\cdot]^+$ denotes Moore-Penrose pseudo-inverse of a matrix.

Finally, we can apply the Block Tensor Decomposition (BTD) approach [3], with suitable constraints imposed on all or only in some preselected factors or components. In this particular case, the simplest scenario is to employ a constrained Block Oriented Decomposition (BOD) model (combining or averaging $N$ Tucker-1 models) [1]:

$$\underline{\mathbf{Y}} = \frac{1}{N} \sum_{n=1}^{N} \left( \underline{\mathbf{G}}^{(n)} \times_n \mathbf{U}^{(n)} \right) + \underline{\tilde{\mathbf{E}}}. \tag{13}$$

This model allows us to estimate all desired components in parallel or sequentially by minimizing the norm of the total errors ($\|\underline{\tilde{\mathbf{E}}}\|_F^2$) subject to specific constraints.

## 3. Dimensionality Reduction, Feature Extraction and Classification of Multiway Data

Dimensionality reduction, feature extraction and selection

---

† Unfortunately, this approach does not guarantee best fitness of the model to the data or minimum norm of errors $\|\mathbf{E}\|_F^2$, but usually the solutions are close to optimal. Note that in practice, for large scale problems, we do not need to perform explicitly unfolding of the full data tensor. Instead, we may apply fast sampling of tubes (columns) of data tensors [25].

are essential problems in the analysis of multidimensional datasets with large number of variables [14]. We shall first illustrate the basic concepts of dimensionality reduction and feature extraction for a set of large-scale sample matrices. Let us consider, that we have available a set of $K$ matrices (2-D samples) $\mathbf{X}^{(k)} \in \mathbb{R}^{I_1 \times I_2}, (k = 1, 2, \ldots, K)$ that represent multi subjects or multi-trials 2D data, which belong to $C$ different classes or categories (e.g., multiple mental task/state data or different mental diseases). In order to perform dimensionality reduction and to extract essential features for all the training samples, we apply simultaneous (approximative and constrained) matrix factorizations (see Fig. 2 (a)):

$$\mathbf{X}^{(k)} = \mathbf{U}^{(1)} \mathbf{F}^{(k)} \mathbf{U}^{(2)T} + \mathbf{E}^{(k)}, \qquad (k = 1, 2, \ldots, K), \tag{14}$$

where the two common factors (basis matrices) $\mathbf{U}^{(1)} \in \mathbb{R}^{I_1 \times J_1}$ and $\mathbf{U}^{(2)} \in \mathbb{R}^{I_2 \times J_2}$, $J_n \leq I_n$, $\forall n$ code (explain) each sample $\mathbf{X}^{(k)}$ simultaneously along the horizontal and vertical dimensions and the extracted features are represented by matrices $\mathbf{F}^{(k)} \in \mathbb{R}^{J_1 \times J_2}$, typically, with $J_1 << I_1$ and $J_2 << I_2$. In special cases $\mathbf{F}^{(k)}$ are squared diagonal matrices. This problem can be considered as a generalization of Joint Approximative Diagonalization (JAD) [1], [27].

The common method to solve such matrix factorizations problem is to minimize a set of cost functions $\|\mathbf{X}^{(k)} - \mathbf{U}^{(1)} \mathbf{F}^{(k)} \mathbf{U}^{(2)T}\|_F^2$, $\forall k$ sequentially or in parallel, with respect to all the factor matrices. We can solve the problem more efficiently by concatenation or tensorization of all matrices $\mathbf{X}^{(k)}$ along the third dimension to form an $I_1 \times I_2 \times K$ dimensional data tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times K}$ and perform the standard Tucker decomposition (see Fig. 2 (b)).

In order to obtain meaningful components and a unique decomposition it is often convenient to impose additional constraints [1], [14]. In the special case when the feature matrices $\mathbf{F}_k$ are diagonal the Tucker-2 model is reduced to special form of the unique CP model [1].

This approach can be naturally and quite straightforwardly extended to multidimensional data. Assume, that we have available a set of multidimensional tensors $\underline{\mathbf{X}}^{(k)} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, $(k = 1, 2, \ldots, K)$, representing training data belonging to $C$ classes or categories. Each training sample $\underline{\mathbf{X}}^{(k)}$ is given a label $c_k$ indicating the category (class) to which it belongs.

In order to preform a dimensionality reduction and to extract significant features, we need to apply simultaneous approximative tensor decompositions (see Fig. 3)

$$\underline{\mathbf{X}}^{(k)} = \underline{\mathbf{G}}^{(k)} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \cdots \times_N \mathbf{U}^{(N)} + \underline{\mathbf{E}}^{(k)}, \tag{15}$$

$(k = 1, 2, \ldots, K)$, where the compressed core tensors $\underline{\mathbf{G}}^{(k)} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$ representing features are of lower dimension than the original data tensors $\underline{\mathbf{X}}^{(k)}$, and we assume that the factors (basis matrices) $\mathbf{U}^{(n)} = [\mathbf{u}_1^{(n)}, \mathbf{u}_2^{(n)}, \ldots, \mathbf{u}_{J_n}^{(n)}] \in \mathbb{R}^{I_n \times J_n}, (n = 1, 2, \ldots, N)$ are common factors for all data tensors.

To compute the factor matrices $\mathbf{U}^{(n)}$ and the core tensors $\underline{\mathbf{G}}^{(k)}$, we concatenate all training (sample) tensors into one $N + 1$ order training data tensor $\underline{\mathbf{X}} =$
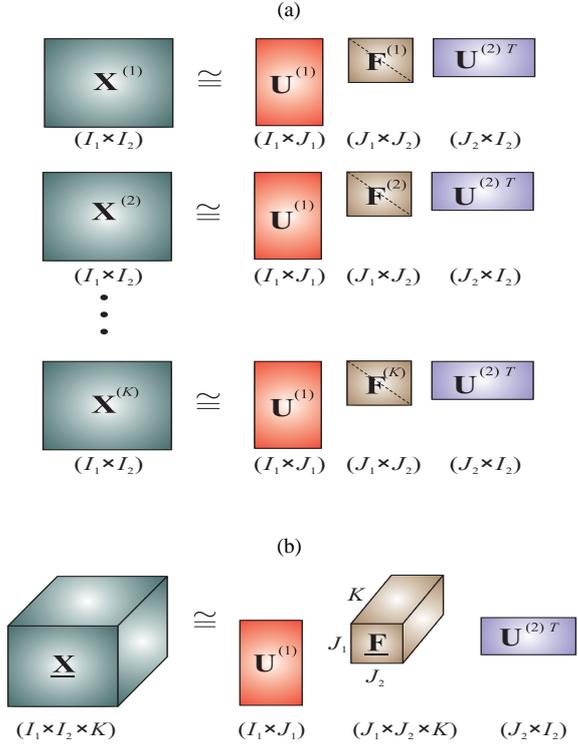
**Fig. 2** (a) Simultaneous matrix factorizations for dimensionality reduction and feature extraction. (b) This problem is equivalent to a Tucker-2 decomposition of a 3-way tensor into a constrained core tensor $\mathbf{F} \in \mathbb{R}^{J_1 \times J_2 \times K}$ (representing features) and two basis factors $\mathbf{U}^{(1)} \in \mathbb{R}^{I_1 \times J_1}$ and $\mathbf{U}^{(2)} \in \mathbb{R}^{I_2 \times J_2}$.
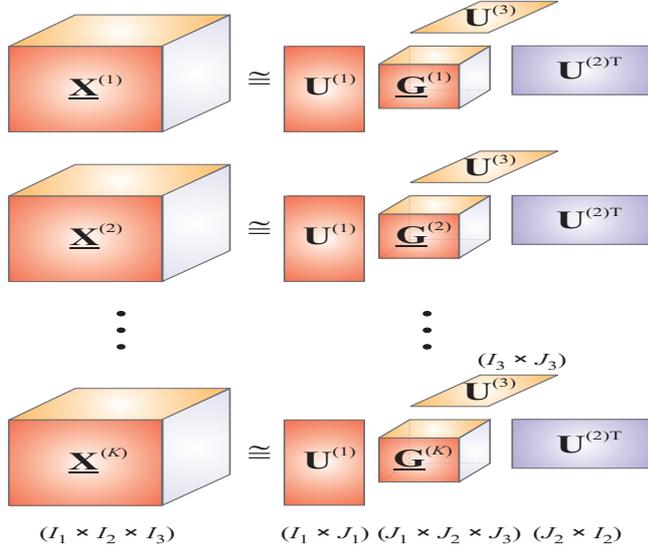


**Fig. 3** Illustration of feature extraction from a set of 3-way data tensors $\underline{\mathbf{X}}^{(k)} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, ($k = 1, 2, \ldots, K$). The objective is to estimate common factor matrices (bases) $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times J_n}$ ($n = 1, 2, 3$) and core tensors $\underline{\mathbf{G}}^{(k)} \in \mathbb{R}^{J_1 \times J_2 \times J_3}$, typically with $J_n \ll I_n$.

$\mathtt{cat}(\underline{\mathbf{X}}^{(1)}, \underline{\mathbf{X}}^{(2)}, \ldots, \underline{\mathbf{X}}^{(K)}, N + 1) \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_N \times I_{N+1}}$, with $N + 1 = K$ and perform the Tucker-$N$ decomposition [14]:

$$\underline{\mathbf{X}} = \underline{\mathbf{G}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \cdots \times_N \mathbf{U}^{(N)} + \underline{\mathbf{E}}, \qquad (16)$$
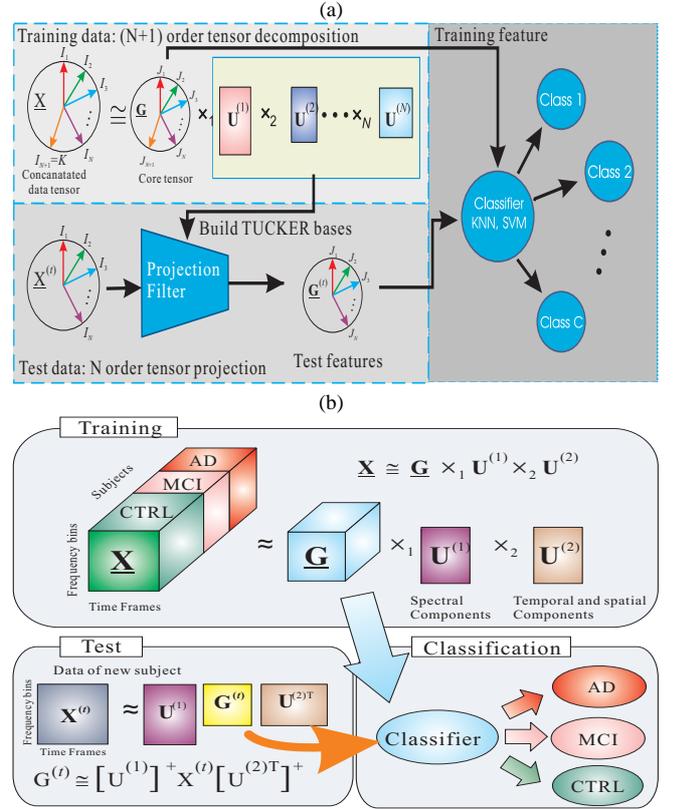


**Fig. 4** (a) Conceptual diagram illustrating a classification procedure based on the Tucker decomposition of the concatenated tensor $\underline{\mathbf{X}} \cong \underline{\mathbf{G}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \cdots \times_N \mathbf{U}^{(N)} \in \mathbb{R}^{I_1 \times I_2 \cdots I_N \times K}$ consisting of all sampling training data $\underline{\mathbf{X}}^{(k)} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_N}$. Reduced features are obtained by projecting the testing data tensor $\underline{\mathbf{X}}^{(t)}$ onto the feature subspace spanned by factors (bases) $\mathbf{U}^{(n)}$. The projection filter depends on the factors $\mathbf{U}^{(n)}$. If they are orthogonal, we can apply a simple projection filter: $\underline{\mathbf{G}}^{(t)} = \underline{\mathbf{X}}^{(t)} \times_1 \mathbf{U}^{(1) T} \times_2 \mathbf{U}^{(2) T} \cdots \times_N \mathbf{U}^{(N) T}$ (otherwise, see Eq. (12)). (b) Application of the general procedure for the classification of Alzheimer's disease (AD), Mild Cognitive Impairment (MCI) patients and age matched control subjects (CTRL).

where the sample tensors $\underline{\mathbf{X}}^{(k)}$ can be extracted back from the concatenated tensor by fixing the $(N + 1)$-th index at a value $k$, i.e., $\underline{\mathbf{X}}(\underset{1}{:}, \underset{2}{:}, \ldots, \underset{N}{:}, \underset{N+1}{k}) = \underline{\mathbf{X}}^{(k)}$ and the individual features (corresponding to different classes) are extracted from the core tensor $\underline{\mathbf{G}} \in \mathbb{R}^{J_1 \times J_2 \cdots \times J_N \times J_{N+1}}$ as $\underline{\mathbf{G}}^{(k)} = \underline{\mathbf{G}}(\underset{1}{:}, \underset{2}{:}, \ldots, \underset{N}{:}, \underset{N+1}{k})$, with $J_{N+1} = K$. In other words, the features of a specific training data $\underline{\mathbf{X}}^{(k)}$ are represented by the $k$-th row of the mode-$(N + 1)$ matricized version of the core tensor $\underline{\mathbf{G}}$.

The above procedure for the feature extraction has been applied for a wide class of classification problems, as illustrated in Fig 4 (a) [14]. In the first stage, we extracted features and estimated the basis matrices $\mathbf{U}^{(n)}$ for the concatenated tensor of all training data $\underline{\mathbf{X}}^{(k)}$ with labels and we built up a projection filter to extract the features of the test data (without labels). In the next step, we extracted features of test data using estimated common basis. The extracted features were then compared with the training features using standard classifier, e.g., LDA, KNN, SVM [14]. We applied the procedure described above to various feature extraction classification problems [14]. For example, we applied

this procedure to successfully classify three groups of human subjects: Control age matched subjects (CTRL), Mild Cognitive Impairment (MCI) and probable Alzheimer's disease (AD) patients (see Fig. 4 (b)) on the basis of spontaneous, followup EEG data [14], [28]. All the training EEG data were organized in the form of 3-way tensors using a time-frequency representation by applying Morlet wavelets. Each frontal slice represented preprocessed data of one subject. We applied the standard Tucker decomposition using HOOI and NTD algorithms to extract the basis matrices $\mathbf{U}^{(n)}$ ($n = 1, 2$) and reduced features represented by a core tensor $\underline{\mathbf{G}}$. The new test data in the form of matrices $\mathbf{X}^{(t)}$ were projected via the basis matrices to extract features. The extracted test features were compared with the training features and classification was performed using LDA, KNN and SVM. We obtained quite promising classification accuracies to predict AD ranging from 82% to 96% depending on EEG data sets.

## 4. Linked Multiway BSS/ICA

In neuroscience, we often need to perform a so called group analysis which seeks to identify mental tasks or stimuli driven brain patterns that are common in two or more subjects in a group.

Various group ICA methods have been proposed to combine the results of multi-subjects, multi-tasks only after ICA is carried out on individual subjects and usually no constraints are imposed on components, which are allowed to differ with respect to their spatial, spectral maps, as well as to their temporal patterns. However, in many scenarios some links need to be considered to analyze variability and consistency of the components across subjects. Furthermore, some components do not need to be necessarily independent, they can be instead sparse, smooth or non-negative (e.g., for spectral components). Moreover, it is often necessary to impose some constraints to be able to estimate some components, which are identical or maximally correlated across subjects with regard to their spatial distributions, spectral or temporal patterns. This leads to a new concept and model of linked multiway BSS (or linked multiway ICA, if statistical independence criteria are used).

In the linked multiway BSS, we perform approximative decompositions of a set of data tensors $\underline{\mathbf{X}}^{(s)} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_N}$, ($s = 1, 2, \ldots, S$) representing multiple subjects and/or multiple tasks (see Fig. 5):

$$\underline{\mathbf{X}}^{(s)} = \underline{\mathbf{G}}^{(s)} \times_1 \mathbf{U}^{(1,s)} \times_2 \mathbf{U}^{(2,s)} \cdots \times_N \mathbf{U}^{(N,s)} + \underline{\mathbf{E}}^{(s)}, \quad (17)$$

where each factor (basis matrix) $\mathbf{U}^{(n,s)} = [\mathbf{U}_C^{(n)}, \ \mathbf{U}_I^{(n,s)}] \in \mathbb{R}^{I_n \times J_n}$ is composed of two bases: $\mathbf{U}_C^{(n)} \in \mathbb{R}^{I_n \times R_n}$ (with $0 \le R_n \le J_n$), which are common bases for all subjects in the group and correspond to the same or maximally correlated components and $\mathbf{U}_I^{(n)} \in \mathbb{R}^{I_n \times (J_n - R_n)}$, which correspond to stimuli/tasks independent individual characteristics.

For example, $\underline{\mathbf{X}}^{(s)}$, ($s = 1, 2, \ldots S$) denotes the brain activity in space-time-frequency domains for the $s$-th sub-
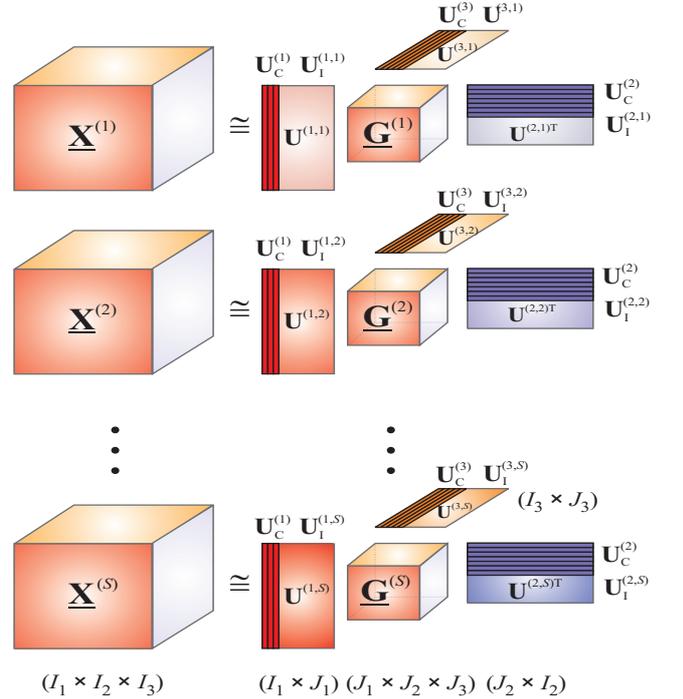


**Fig. 5** Conceptual model of tensors decomposition for linked multiway BSS (especially, Linked Multiway ICA). Objective is to find constrained factor matrices $\mathbf{U}^{(n,s)} = [\mathbf{U}_C^{(n)}, \ \mathbf{U}_I^{(n,s)}] \in \mathbb{R}^{I_n \times J_n}, \ n = 1, 2, 3)$ and core tensors $\underline{\mathbf{G}}^{(s)} \in \mathbb{R}^{J_1 \times J_2 \times J_3}$, which are partially linked or maximally correlated, i.e., they have the same common components or highly correlated components.

ject. Based on a set of such data, we can compute common factors and interactions between them. For example, linked multiway BSS approach my reveal coupling of brain regions, possibly at different time slots and/or different frequency bins.

To solve problems formulated this way, we can apply similar procedures to the one described in previous two Sections. If $\mathbf{U}^{(n,s)} = \mathbf{U}_C^{(n)} \in \mathbb{R}^{I_n \times J_n}$ for a specific mode $n$, then we can concatenate all data tensors along this mode, perform tensor decomposition and apply any standard BSS algorithm to compute desired components in each mode (e.g., to estimate independent components, we apply any standard ICA algorithm). In the more general case, when the number of common components $R_n$ are unknown, we perform an additional unfolding for each data tensor $\underline{\mathbf{X}}^{(s)}$ in each mode and then perform a set of constrained matrix factorizations (by applying standard algorithms for ICA, NMF, SCA, SmoCA, PCA/SVD, etc.):

$$\mathbf{X}_{(n)}^{(s)} = \mathbf{U}^{(n,s)} \mathbf{G}_{(n)}^{(s)} \mathbf{Z}^{(n,s)} + \mathbf{E}^{(n)} \cong \mathbf{B}_{n,s} \mathbf{A}_{n,s}^T, \quad (18)$$

for $n = 1, 2, \ldots, N$ and $s = 1, 2, \ldots, S$, where matrices $\mathbf{B}_{n,s} = \mathbf{U}^{(n,s)}$ represent individual linked components, while matrices $\mathbf{A}_{n,s}^T = \mathbf{G}_{(n)}^{(s)} \mathbf{Z}^{(n,s)}$ represent basis vectors or linked mixing processes and $\mathbf{Z}^{(n,s)} = \left[ \mathbf{U}^{(N,s)} \otimes \cdots \otimes \mathbf{U}^{(n+1,s)} \otimes \mathbf{U}^{(n-1,s)} \cdots \otimes \mathbf{U}^{(1,s)} \right]^T$.

In the next stage, we need to perform the statistical analysis and to compare the individual components ex-

tracted in each mode $n$, for all subjects $S$, by performing clustering and ordering components to identify common or similar (highly correlated) components (see e.g., [1], [17]). In the last stage, we compute core tensors, which describe the functional links between components (see Eq. (12)).

For the linked multiway BSS, we can also exploit constrained Block Tensor Decomposition (BTD) models for an averaging data tensor across subjects:

$$\underline{\bar{\mathbf{X}}} = \sum_{s=1}^{S} \left( \underline{\mathbf{G}}^{(s)} \times_1 \mathbf{U}^{(1,s)} \times_2 \mathbf{U}^{(2,s)} \cdots \times_N \mathbf{U}^{(N,s)} \right) + \underline{\bar{\mathbf{E}}}. \quad (19)$$

Such model can provide us additional information.

In group and linked multiway BSS and ICA, we usually seek stimuli driven ERPs (event related responses) and/or task related common components or common bases reflecting both intra subject and inter subject features as bases, which are independent involving individual on going brain activities [29]. In other words, we seek event/task-related components $\mathbf{U}_C^{(n)}$, that are identical in each mode or maximally correlated across subjects and event/task independent individual bases $\mathbf{U}_I^{(s,n)}$, which are independent or even as far apart as possible (anti-correlated) [29].

Note that our Linked Multiway BSS is different from the linked ICA [23] and group NMF [29], [30], since we do not limit components diversity by constraining them to be only statistically independent or only nonnegative components and our model is based on constrained Tucker models instead of rather the quite restrictive trilinear CP model.

How to select common components will depend on the validity of the underlying assumptions and *a priori* knowledge. For instance, identical spatial distributions can well be assumed for homogeneous subject groups, but may be unacceptable in studies that include patients with different ages or mental diseases or abnormalities. Temporal components may be the same for stimulus- or task-evoked responses that are related to a specific mental task or paradigm, but these will vary for the spontaneous fluctuations that occur in resting state experiments. In some experiments, responses may be assumed similar or identical within but different across subgroups [17], [18].

We conclude that the proposed Linked BSS model provides a framework that is very flexible and general and it may substantially supplement many of the currently available techniques for group ICA and feature extraction models. Moreover, the model can be extended to a multiway Canonical Correlation Analysis (MCCA), in which we impose maximal correlations among normalized factor matrices (or subset of components) and/or core tensors.

## 5. Multiway Partial Least Squares

The Partial Least Squares (PLS) methods (originally developed in chemometrics and econometrics) are particularly suitable for the analysis of relationships among multi-modal brain data (e.g., EEG, MEG, ECoG (electrocorticogram), fMRI) or relationships between measures of brain activity

and behavior data [31], [32]. The standard PLS approaches have been recently summarized by H. Abdi [31] and their suitability to model relationships between brain activity and behavior (experimental design) has been highlighted by A. Krishnan *et al.* [32].

In computational neuroscience, there are two related basic PLS methods: PLS correlation (PLSC), which analyzes correlations or associations between two or more sets of data (e.g., two modalities brain data or brain and behavior data) and PLS regression (PLSR) methods, which attempt to predict one set of data from another independent data that constitutes the predictors (e.g., experimental behavior data from brain data such multichannel ECoG or scalp EEG from ECoG, by performing simultaneous recordings for epileptic patients).

In order to predict response variables represented by matrix $\mathbf{Y}$ from the independent variables $\mathbf{X}$, the standard PLSR techniques find a set of common orthogonal latent variables (also called latent vectors, score vectors or components) by projecting both $\mathbf{X}$ and $\mathbf{Y}$ onto a new subspace, which ensures a maximal correlation between the latent variables of $\mathbf{X}$ and $\mathbf{Y}$ [31]. In other words, the prediction is achieved by simultaneous approximative decompositions training data sets: $\mathbf{X} \in \mathbb{R}^{I \times N}$ and $\mathbf{Y} \in \mathbb{R}^{I \times M}$ into components ($\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_J] \in \mathbb{R}^{I \times J}, \mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_J] \in \mathbb{R}^{N \times J}$ and $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_J] \in \mathbb{R}^{M \times J}$) :

$$\mathbf{X} \cong \mathbf{AB}^T = \sum_{j=1}^{J} \mathbf{a}_j \mathbf{b}_j^T, \quad (20)$$

$$\mathbf{Y} \cong \mathbf{ADC}^T = \sum_{j=1}^{J} d_{jj} \mathbf{a}_j \mathbf{c}_j^T, \quad (21)$$

where $\mathbf{D} \in \mathbb{R}^{J \times J}$ is a scaling diagonal matrix; with the constraint that these components explain as much as possible of the covariance between $\mathbf{X}$ and $\mathbf{Y}$ [31], [33]. The latent components $\mathbf{A}$ are defined as $\mathbf{A} = \mathbf{XW}$, where $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_J] \in \mathbb{R}^{N \times J}$ consists of $J$ direction vectors $\mathbf{w}_j$. The basic concept of the standard PLS is to find these directions vectors $\mathbf{w}_j$ from successive optimizations problems:

$$\mathbf{w}_j = \arg \max_{\mathbf{w}} \left\{ \mathrm{corr}^2(\mathbf{Xw}, \mathbf{Y}) \, \mathrm{var}(\mathbf{Xw}) \right\},$$

$$s.t. \qquad \mathbf{w}^T \mathbf{w} = 1, \quad \mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w}_k = 0$$

for $k = 1, 2, \ldots, j - 1$.

Such decompositions and relationships between components can be used to predict values of dependent variables for new situations, if the values of the corresponding independent variables are available.

Since the brain data are often multidimensional and multi-modal and can be naturally represented via tensors, attempts have been made to extend PLS methods to multiway models [33]. In this section, we briefly describe multiway PLS based on a constrained Tucker model (PTD): Given an $N$th-order independent data tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_n \times \cdots \times I_N}$ and an $M$th-order dependent data tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{K_1 \times \cdots \times K_m \times \cdots \times K_M}$, with
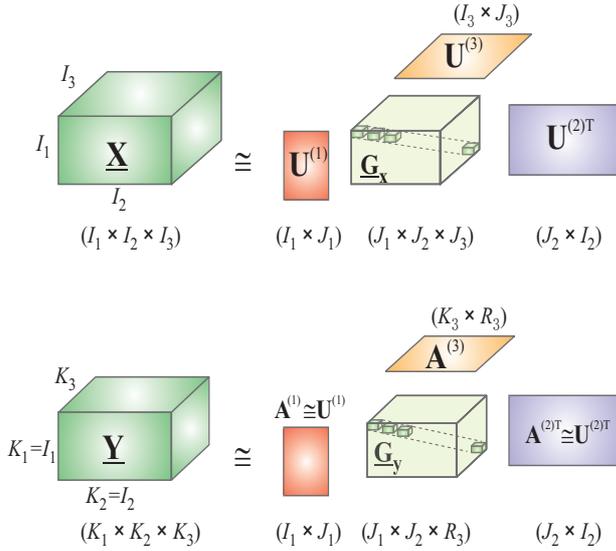
**Fig. 6**  Illustration of a generalized Tucker-based Multiway PLS model for 3-way tensors in the case when two factor matrices are common (see Eqs. (22)-(23)). Our objective is to perform approximative decompositions of independent tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_N}$ and dependent tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{K_1 \times K_2 \cdots \times K_M}$, with $I_1 = K_1$ and $I_2 = K_2$, by imposing additional constraints, especially that the two factor matrices are the same or highly correlated: $\mathbf{A}^{(1)} \cong \mathbf{U}^{(1)}$ and $\mathbf{A}^{(2)} \cong \mathbf{U}^{(2)}$, while the other factors are uncorrelated (or anti-correlated) or statistically independent. We assumed also that the core tensors are sparse and block diagonal.

the same size in at least one mode (typically, the first mode)$^{\dagger}$ $I_1 = K_1$.

Our objective is to preform simultaneous constrained Tucker decomposition, with at least one (or two) common or maximally correlated factor(s) (see Fig. 6):

$$\underline{\mathbf{X}} = \underline{\mathbf{G}}_{\mathbf{x}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \cdots \times_N \mathbf{U}^{(N)} + \underline{\mathbf{E}}_{\mathbf{x}}, \qquad (22)$$

$$\underline{\mathbf{Y}} = \underline{\mathbf{G}}_{\mathbf{y}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \cdots \times_M \mathbf{A}^{(M)} + \underline{\mathbf{E}}_{\mathbf{y}}, \qquad (23)$$

where additional constraints are imposed: $\mathbf{U}^{(1)} \cong \mathbf{A}^{(1)} \in \mathbb{R}^{I_1 \times J_1}$ (and $\mathbf{U}^{(2)} \cong \mathbf{A}^{(2)} \in \mathbb{R}^{I_2 \times J_2}$ of $I_2 = K_2$), while other factor matrices are essentially different (e.g., orthogonal or mutually independent). Note that the core tensors $\underline{\mathbf{G}}_{\mathbf{x}}$ and $\underline{\mathbf{G}}_{\mathbf{y}}$ have special block-diagonal structures (see Fig. 6) that indicate sparseness. New algorithms for constrained Tucker based multiway PLS models have been developed in [33].

Such models allow for different types of structures on $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$ and provides a general framework for solving multiway regression problems that explore complex relationships between multidimensional dependent and independent variables. For example, tensor decompositions can be applied in emerging neuroimaging genetics studies to investigate links between biological parameters measured with brain imaging and genetic variability [34].

---

$^{\dagger}$The first mode, usually, corresponds to the samples mode or the time mode. However, in some applications, for example, for simultaneous recordings of EEG and ECoG data two modes (ways) may have the same size and they represent time and frequency (temporal and spectral modes, respectively).

## 6.  Conclusions

Tensor decompositions is a fascinating emerging field of research, with many applications in multilinear blind source separation, linked multiway BSS/ICA, feature extraction, classification and prediction. In this review/tutorial paper, we have briefly discussed several new and promising models for decompositions of linked or correlated sets of tensors, by imposing various constraints on factors, components or latent variables depending on the specific applications. We have developed several promising and efficient algorithms for such models, which can be found in our recent book, publications and reports.

**References**

[1] A. Cichocki, R. Zdunek, A.H. Phan, and S. Amari, Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multiway Data Analysis and Blind Source Separation, Wiley, Chichester, 2009.

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

[3] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms – Part I and II," SIAM Journal on Matrix Analysis and Applications (SIMAX), vol.30, no.3, pp.1022–1066, 2008.

[4] F. Hitchcock, "Multiple invariants and generalized rank of a p-way matrix or tensor," Journal of Mathematics and Physics, vol.7, pp.39–79, 1927.

[5] L. Tucker, "Some mathematical notes on three-mode factor analysis," Psychometrika, vol.31, pp.279–311, 1966.

[6] A. Cichocki, "Generalized Component Analysis and Blind Source Separation Methods for Analyzing Mulitchannel Brain Signals," in Statistical and Process Models for Gognitive Neuroscience and Aging, pp.201–272, Lawrence Erlbaum Associates, 2007.

[7] A. Cichocki and S. Amari, Adaptive Blind Signal and Image Processing, John Wiley & Sons Ltd, New York, 2003.

[8] A. Cichocki and A.H. Phan, "Fast local algorithms for large scale nonnegative matrix and tensor factorizations," IEICE (invited paper), vol.EA92-A(3), pp.708–721, March.

[9] A. Cichocki, R. Zdunek, and S.I. Amari, "Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization," Springer, Lecture Notes on Computer Science, LNCS-4666, pp.169–176, 2007.

[10] M. Mørup, L.K. Hansen, C.S. Herrmann, J. Parnas, and S.M. Arnfred, "Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG," NeuroImage, vol.29, no.3, pp.938–947, 2006.

[11] E. Acar, D.M. Dunlavy, T.G. Kolda, and M. Mørup, "Scalable tensor factorizations for incomplete data," Chemometrics and Intelligent Laboratory Systems, vol.106 (1), pp.41–56, 2011.

[12] F. Miwakeichi, E. Martinez-Montes, P. Valds-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchi, "Decomposing EEG data into space−time−frequency components using parallel factor analysis," NeuroImage, vol.22, no.3, pp.1035–1045, 2004.

[13] A. Cichocki, R. Zdunek, S. Choi, R. Plemmons, and S. Amari, "Nonnegative tensor factorization using Alpha and Beta divergencies," Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP07), Honolulu, Hawaii, USA, pp.1393–1396, April 15–20 2007.

[14] A.H. Phan and A. Cichocki, "Tensor decompositions for feature extraction and classification of high dimensional datasets," Nonlinear Theory and Its Applications, NOLTA IEICE (invited paper), vol.1, pp.37–68, 2010.

[15] L. De Lathauwer, B.D. Moor, and J. Vandewalle, "On the best rank-1 and rank-(R1,R2,...,RN) approximation of higher-order tensors," SIAM Journal of Matrix Analysis and Applications, vol.21, no.4, pp.1324–1342, 2000.

[16] V. Calhoun, J. Liu, and T. Adali, "A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data," Neuroimage, vol.45, pp.163–172, 2009.

[17] Y. Guo and G. Pagnoni, "A unified framework for group independent component analysis for multi-subject fMRI data," Neuroimage, vol.42 (3), pp.1078–1093, 2008.

[18] D. Langers, "Unbiased group-level statistical assessment of independent component maps by means of automated retrospective matching," Human Brain Mapping, vol.31, pp.727–742, 2010.

[19] C. Beckmann and S. Smith, "Tensorial extensions of independent component analysis for multisubject fMRI analysis," NeuroImage, vol.25, no.1, pp.294 – 311, 2005.

[20] M. Vasilescu and D. Terzopoulos, "Multilinear independent components analysis," Computer Vision and Pattern Recognition, 2005. CVPR 2005, IEEE Computer Society Conference on, pp.547 – 553, 2005.

[21] S. Unkel, A. Hannachi, N. Trendafilov, and I. Jolliffe, "Independent component analysis for three-way data with an application from atmospheric science," Journal of Agricultural, Biological, and Environmental Statistics, p.(In press), 2011.

[22] M. De Vos, D. Nion, S. Van Huffel, and L. De Lathauwer, "A combination of parallel factor and independent component analysis," tech. rep., 2011.

[23] A. Groves, C. Beckmann, S. Smith, and M. Woolrich, "Linked independent component analysis for multimodal data fusion," NeuroImage, vol.54, no.1, pp.2198 – 21217, 2011.

[24] I. Kopriva, I. Jeric, and A. Cichocki, "Blind decomposition of infrared spectra using flexible component analysis," Chemometrics and Intelligent Laboratory Systems, vol.97, pp.170–178, 2009.

[25] C. Caiafa and A. Cichocki, "Generalizing the column-row matrix decomposition to multi-way arrays," Linear Algebra and its Applications, vol.433, no.3, pp.557–573, 2010.

[26] G. Zhou and A. Cichocki, "Fast and unique Tucker decompositions via Multiway Blind Source Separation," 2011 (submitted).

[27] L. De Lathauwer, "A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization," SIAM Journal on Matrix Analysis and Applications, vol.28, pp.642–666, 2006.

[28] A. Cichocki, S. Shishkin, T. Musha, Z. Leonowicz, T. Asada, and T. Kurachi, "EEG filtering based on blind source separation (BSS) for early detection of Alzheimer disease," vol.116, pp.729–737, 2005.

[29] H. Lee and S. Choi, "Group nonnegative matrix factorization for EEG classification," in Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS-2009), Clearwater Beach, Florida, April 16-18, 2009, 2009.

[30] H. Lee, Y.D. Kim, A. Cichocki, and S. Choi, "Nonnegative tensor factorization for continuous EEG classification," International Journal of Neural Systems, vol.17, no.4, pp.1–13, August 2007.

[31] H. Abdi, "Partial least square regression, projection on latent structure regression, PLS-regression," Wiley Interdisciplinary Reviews: Computational Statistics, vol.2, pp.97–106, 2010.

[32] A. Krishnan, L. Williams, H. McIntosh, and H. Abdi, "Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review," NeuroImage, vol.56, pp.455–475, 2011.

[33] Q. Zhao, C. Caiafa, Z. Chao, Y. Nagasaka, D. Mandic, N. Fujii, L. Zhang, and A. Cichocki, "Higher-order partial least squares (HOPLS): A generalized multi-linear regression method," PAMI, vol.(submitted), 2012.

[34] J. Poline, C. Lalanne, A. Tenenhaus, E. Duchesnay, B. Thirion, and V. Frouin, "Imaging genetics: bio-informatics and bio-statistics challenges," in Proceedings of the 19th International Conference on Computational Statistics, Paris, France, 2010.
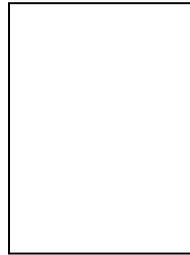
**Andrzej Cichocki** was born in Poland. He received his M.Sc. (with honors), Ph.D. and Habilitate Doctorate (Dr.Sc.) degrees, all in electrical engineering, from the Warsaw University of Technology (Poland). He is the co-author of four international books and monographs (two of them translated to Chinese): Nonnegative Matrix and Tensor Factorizations, J. Wiley 2009, Adaptive Blind Signal and Image Processing, J. Wiley 2002, MOS Switched-Capacitor and Continuous-Time Integrated Circuits and Systems (Springer-Verlag, 1989) and Neural Networks for Optimization and Signal Processing (J. Wiley and Teubner Verlag, 1993/94) and author or co-author of more than 200 papers. Currently, he is senior team leader and head of the laboratory for Advanced Brain Signal Processing in the RIKEN Brain Science Institute, JAPAN.