Applying fully tensorial ICA to fMRI data

Joni Virta
Dept. of Mathematics and Statistics
University of Turku
Turku, FIN-20014, Finland
Email: joni.virta@utu.fi

Sara Taskinen
Dept. of Mathematics and Statistics
University of Jyväskylä
Jyväskylä, FIN-40014, Finland
Email: sara.l.taskinen@jyu.fi

Klaus Nordhausen
Dept. of Mathematics and Statistics
University of Turku
Turku, FIN-20014, Finland
Email: klaus.nordhausen@utu.fi

***Abstract*—There are two aspects in functional magnetic resonance imaging (fMRI) data that make them awkward to analyse with traditional multivariate methods — high order and high dimension. The first of these refers to the tensorial nature of observations as array-valued elements instead of vectors. Although this can be circumvented by vectorizing the array, doing so simultaneously loses all the structural information in the original observations. The second aspect refers to the high dimensionality along each dimension making the concept of dimension reduction a valuable tool in the processing of fMRI data. Different methods of tensor dimension reduction are currently gaining popularity in literature, and in this paper we apply two recently proposed methods of tensorial independent component analysis to simulated task-based fMRI data. Additionally, as a preprocessing step we introduce a novel extension of PCA for tensors. The simulations show that when extracting a sufficiently large number of principal components, the tensor methods find the task signals very reliably, something the standard temporal independent component analysis (tICA) fails in.**

I. Introduction

A. fMRI data as tensors

Functional magnetic resonance imaging (fMRI) data are well-known for their large volume, both in size and in number of dimensions. Together these result in an enormous number of variables to deal with. For example, assuming 1000 three-dimensional scans with 256 voxels per dimension means observing over 16 million variables and leading into a severe case of high dimension, low sample size data. Thus a first step in the analysis of fMRI data is often some form of dimension reduction.

In order to apply standard multivariate methods in the analysis of fMRI data, the sample of observed scans is usually first vectorized. However, in doing so one also inadvertently discards all structural information in the observations. It seems natural to assume that, if almost all voxels in two layers of a scan correlate highly, then so do the rest of them. By intentionally losing the information on the spatial proximity of the voxels seems therefore counterintuitive. A more logical alternative is obtained by not vectorizing but instead keeping the data in the array, or *tensor*, form for the whole time. Various approaches for applying tensorial analysis methods to fMRI data are discussed in [1], [2], [3], for example.

B. fMRI and ICA

Independent component analysis (ICA) is a well-established analysis method for fMRI data. For some discussions on using ICA for fMRI data, see for example [4], [5]. When applied in the traditional way, ICA means that one first has to decide whether to perform the so-called spatial ICA (sICA) or temporal ICA (tICA), and then choose an ICA algorithm out of the many possibilities. To illustrate this, assume next that the vectorized tensors are contained in a $T \times p$ matrix $\boldsymbol{X}$ where T is the number of samples (time points) and p is the number of voxels. As $T \ll p$, the most common approach has been to apply ICA to the transpose of the data matrix, $\boldsymbol{X}^T$. The method, which is called spatial ICA, however makes no use of the spatial dependencies in the data. A more natural approach, temporal ICA, treats the data matrix $\boldsymbol{X}$ as such and thus accounts for the spatial aspect of the data. The drawback of tICA, however, is that the temporal dependencies in the data are ignored. Notice also that a common preprocessing step for both sICA and tICA is the reduction of the dimension by a singular value decomposition (SVD) or principal component analysis (PCA) on the vectorized tensors.

As already stated, relying on vectorization means that neither of the two analysis approaches makes use of the tensorial structure of fMRI data, and thus ignore the Kronecker covariance structure of the data. See [6] for the benefits of the exploitation of Kronecker structure in the context of EEG/MEG data. Steps for bringing ICA to a tensorial form have been suggested already in [7], [8] but a fully tensorial ICA model and two appropriate methods (tensorial FOBI, TFOBI, and tensorial JADE, TJADE) have been suggested only recently in [9], [10]. In this paper we first introduce some relevant notation and tensor terminology in Section II. Then, in Section III, we review TFOBI and TJADE and propose a tensorial scheme for extending the tICA method based on them. In Section IV we then compare the scheme to standard tICA using simulated fMRI data, and finally we conclude the paper with some prospective ideas for further research in Section V.

II. Notation

We use lower-case letters for scalar constants, a, b, c, lower-case boldface letters for vector constants, $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$, upper-case boldface letters for matrix constants, $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ and script letters for general tensor constants, $\mathcal{A}, \mathcal{B}, \mathcal{C}$. The same convention is used with random elements but instead using letters from the end of the alphabet, e.g. $x, \boldsymbol{x}, \boldsymbol{X}, \mathcal{X}$.

By random tensor $\mathcal{X}$ we mean a random element taking values in $\mathbb{R}^{p_1 \times \cdots \times p_r}$ and, although generally difficult to visualize, a mental image of a tensor can be formed by considering it as a collection of vectors. For a tensor of order r this can be done in a total of r ways giving us the concept of *m-mode vectors* or *fibers*. More formally, for a given mode m, the $\rho_m := \Pi_{s \neq m}^r p_s$ m-mode vectors are obtained by varying the mth index while holding the others fixed. In a sense an opposite construct, the p_m *m-mode faces* or *slices* of a tensor are obtained by fixing the value of the mth index and varying the others.

To manipulate tensors we introduce two forms of tensor contraction. The linear transformation $\mathcal{X} \odot_m \boldsymbol{A}_m$ of a tensor $\mathcal{X} = (x_{i_1 \cdots i_r}) \in \mathbb{R}^{p_1 \times \cdots \times p_r}$ by a matrix $\boldsymbol{A}_m = (a_{ij}^{(m)}) \in \mathbb{R}^{q_m \times p_m}$ from the mth mode produces a $p_1 \times \cdots \times q_m \times \cdots \times p_r$ tensor with the elements

$$(\mathcal{X} \odot_m \boldsymbol{A}_m)_{i_1 \cdots i_r} = \sum_{j_m=1}^{p_m} x_{i_1 \cdots i_{m-1} j_m i_{m+1} \cdots i_r} a_{i_m j_m}^{(m)}.$$

The previous operation is most easily understood as applying the linear transformation given by $\boldsymbol{A}_m$ separately to each m-mode vector of $\mathcal{X}$. For ease of reading we use in the following the shorthand notation $\mathcal{X} \odot_{m=1}^r \boldsymbol{A}_m := (\cdots (\mathcal{X} \odot_1 \boldsymbol{A}_1) \cdots \odot_r \boldsymbol{A}_r)$. Furthermore, for tensors of order $r = 1, 2$ we have the following connections to ordinary matrix multiplication: $\boldsymbol{x} \odot_1 \boldsymbol{A}_1 = \boldsymbol{A}_1 \boldsymbol{x}$, $\boldsymbol{X} \odot_1 \boldsymbol{A}_1 = \boldsymbol{A}_1 \boldsymbol{X}$ and $\boldsymbol{X} \odot_2 \boldsymbol{A}_2 = \boldsymbol{X} \boldsymbol{A}_2^T$.

The second form of tensor contraction takes a single tensor $\mathcal{X} \in \mathbb{R}^{p_1 \times \cdots \times p_r}$ and returns the symmetric matrix $\mathcal{X} \odot_{-m} \mathcal{X} \in \mathbb{R}^{p_m \times p_m}$ with the elements

$$(\mathcal{X} \odot_{-m} \mathcal{X})_{kl} = \sum x_{i_1 \cdots i_{m-1} k i_{m+1} \cdots i_r} x_{i_1 \cdots i_{m-1} l i_{m+1} \cdots i_r},$$

where the summing is over all indices expect the mth one. The above operation can be seen to be equivalent to the sum of outer products of all m-mode vectors of $\mathcal{X}$ with themselves. Thus, for tensors of order $r = 1, 2$ we again have the connections $\boldsymbol{x} \odot_{-1} \boldsymbol{x} = \boldsymbol{x}\boldsymbol{x}^T$, $\boldsymbol{X} \odot_{-1} \boldsymbol{X} = \boldsymbol{X}\boldsymbol{X}^T$ and $\boldsymbol{X} \odot_{-2} \boldsymbol{X} = \boldsymbol{X}^T \boldsymbol{X}$. For a comprehensive introduction to manipulating tensors, see e.g. [11].

Finally, denote the standard basis vectors of $\mathbb{R}^p$ by $\boldsymbol{e}_i$, $i = 1, \ldots, p$, and by $\boldsymbol{E}^{ij} := \boldsymbol{e}_i \boldsymbol{e}_j^T$ the matrix with a single one as the (i, j)th element and rest of the entries zero.

III. Methods

A. Independent component models

We begin by shortly reviewing the tensorial independent component analysis framework along with two methods, TFOBI and TJADE. For the ease of understanding, several contrasting comparisons to the vector-based methods will be made.

Recall that the traditional vector independent component model assumes that the sample of i.i.d. observations $\boldsymbol{x}_i \in \mathbb{R}^p$ satisfies

$$\boldsymbol{x}_i = \boldsymbol{m} + \boldsymbol{\Omega} \boldsymbol{z}_i, \quad i = 1, \ldots, T,$$

where $\boldsymbol{m} \in \mathbb{R}^p$, the latent i.i.d p-vectors $\boldsymbol{z}_i$ have mutually independent components and the square $p \times p$ matrix $\boldsymbol{\Omega}$ is invertible. An extension of the model for a sample of i.i.d. tensors $\mathcal{X}_i \in \mathbb{R}^{p_1 \times \cdots \times p_r}$ was constructed in [9] as

$$\mathcal{X}_i = \mathcal{M} + \mathcal{Z}_i \odot_{m=1}^r \boldsymbol{\Omega}_m, \quad i = 1, \ldots, T, \tag{1}$$

where $\mathcal{M} \in \mathbb{R}^{p_1 \times \cdots \times p_r}$, the latent i.i.d. tensors $\mathcal{Z}_i \in \mathbb{R}^{p_1 \times \cdots \times p_r}$ have independent components, and the square matrices $\boldsymbol{\Omega}_m \in \mathbb{R}^{p_m \times p_m}$, $m = 1, \ldots, r$, are invertible. To quarantee identifiability we set the following constraints for the corresponding population quantities: (i) $\mathrm{E}[\mathrm{vec}(\mathcal{X})] = \boldsymbol{0}$, (ii) $\mathrm{Cov}[\mathrm{vec}(\mathcal{X})] = \boldsymbol{I}$, and (iii) for each mode $m = 1, \ldots, r$, at most one m-mode slice of $\mathcal{Z}$ has only Gaussian components. The operation "vec" stacks the elements of a tensor into a vector, the order it is done playing no role here. The first two constraints require the components of $\mathcal{Z}$ to be marginally standardized, and the third one fixes the issue with the orthogonal invariance of the standard multivariate Gaussian distribution, see [12], [9]. Note that while in the vector case the last assumption means that at most single component of $\boldsymbol{z}$ can be Gaussian, in the general tensor case most components of $\mathcal{Z}$ can be Gaussian without violating constraint (iii) [10].

In the case of a single-subject fMRI data we thus have a sequence of 3-mode tensors and we assume that $\mathcal{Z}_i$ is a latent tensor of independent signals where each activation signal resides only in a single voxel (unlike in $\mathcal{X}_i$ where the activation signal can cover multiple voxels with varying amplitudes) other voxels containing just noise. The mixing matrices $\boldsymbol{\Omega}_m$, $m = 1, \ldots, r$ then create dependencies in $\mathcal{Z}_i$ from their respective modes/directions and produce the observed tensor $\mathcal{X}_i$ for the ith time point. The aim of ICA is to reverse this process so that we can pick the individual signals of interest from the estimated $\mathcal{Z}_i$.

As in standard ICA, the first step in tensorial ICA is standardization. Assuming in the following centered random vectors, the standardization in vector ICA is performed as

$$\boldsymbol{x} \mapsto \boldsymbol{x}_{st} := \boldsymbol{\Sigma}^{-1/2} \boldsymbol{x},$$

where $\boldsymbol{\Sigma} := \mathrm{E}[\boldsymbol{x}\boldsymbol{x}^T]$ is the covariance matrix of the random vector $\boldsymbol{x}$ and the inverse square root is chosen to be symmetric. The standardized random vector then satisfies $\boldsymbol{x}_{st} = \boldsymbol{U}\boldsymbol{z}$ for some orthogonal $\boldsymbol{U} \in \mathbb{R}^{p \times p}$, reducing the problem of inverting an unknown full-rank matrix to that of inverting an unknown orthogonal matrix, see [13].

In [9] a tensorial extension of the above standardization is formulated as

$$\mathcal{X} \mapsto \mathcal{X}_{st} := \mathcal{X} \odot_{m=1}^r \boldsymbol{\Sigma}_m^{-1/2}, \tag{2}$$

where we again (from now on) assume that the tensor $\mathcal{X}$ is centered and the *m-mode covariance matrices* are computed as $\boldsymbol{\Sigma}_m := \rho_m^{-1} \mathrm{E}[\mathcal{X} \odot_{-m} \mathcal{X}]$, $m = 1, \ldots, r$, their inverse square roots being chosen to be symmetric. [9] further show that the standardized tensor satisfies

$$\mathcal{X}_{st} \propto \mathcal{Z} \odot_{m=1}^r \boldsymbol{U}_m, \tag{3}$$

for some orthogonal matrices $\boldsymbol{U}_m \in \mathbb{R}^{p_m \times p_m}$, $m = 1, \ldots, r$. The estimation of these unknown rotations is then what both

TFOBI and TJADE aim to do. Note that having proportionality instead of equality in (3) is non-restrictive as the overall scale in (1) is not estimable due to multiple mixing matrices.

B. TFOBI

The standard FOBI [14] utilizes the matrix of fourth moments, $\boldsymbol{B}(\boldsymbol{x}) := \mathrm{E}[\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{x}\boldsymbol{x}^T]$, which can be shown to satisfy $\boldsymbol{B}(\boldsymbol{x}_{st}) = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T$ for some diagonal matrix $\boldsymbol{D}$. Assuming that the kurtoses of the components of $\boldsymbol{z}$ (that is, the diagonal elements of $\boldsymbol{D}$) are distinct, the unknown rotation $\boldsymbol{U}$ is then directly estimable from the eigendecomposition of $\boldsymbol{B}(\boldsymbol{x}_{st})$.

[9] extended FOBI method to tensors with the product operation $\odot_{-m}$ by defining the *m-mode matrices of fourth moments* as

$$\boldsymbol{B}_m(\mathcal{X}) := \rho_m^{-1}\mathrm{E}[(\mathcal{X} \odot_{-m} \mathcal{X})^2], \quad m = 1, \ldots, r, \quad (4)$$

which can be shown to satisfy $\boldsymbol{B}_m(\mathcal{X}_{st}) = \boldsymbol{U}_m\boldsymbol{D}_m\boldsymbol{U}_m^T$, for some diagonal matrix $\boldsymbol{D}_m \in \mathbb{R}^{p_m \times p_m}$, for all $m = 1, \ldots, r$. Thus, the rotation matrices $\boldsymbol{U}_m$ can be estimated from the eigendecompositions of $\boldsymbol{B}_m(\mathcal{X}_{st})$, $m = 1, \ldots, r$, and they are identifiable (up to sign and permutation of their columns) if, for all $m = 1, \ldots, r$, the average kurtoses of the m-mode slices are distinct. An alternative formulation (that is generally inferior in performance) for TFOBI also exists and can be found in [9].

C. TJADE

JADE [13] is commonly considered as an improvement over FOBI. Instead of using the information contained in a single matrix of fourth moments, $\boldsymbol{B}$, JADE uses all possible p^4 joint fourth cumulants to estimate the orthogonal matrix $\boldsymbol{U}$. These cumulants are conveniently contained in the following set of p^2 matrices indexed by two indices:

$$\boldsymbol{C}^{ij} = E\left[\boldsymbol{e}_i^T\boldsymbol{x}_{st}\boldsymbol{x}_{st}^T\boldsymbol{e}_j \cdot \boldsymbol{x}_{st}\boldsymbol{x}_{st}^T\right] - \delta_{ij}\boldsymbol{I} - \boldsymbol{E}^{ij} - \boldsymbol{E}^{ji}, \quad (5)$$

with $i, j = 1, \ldots, p$. It can be shown that the orthogonal matrix $\boldsymbol{U}$ diagonalizes the matrix $\boldsymbol{C}_{ij}$ for all $i, j = 1, \ldots, p$, but due to individual rank deficiencies we *jointly diagonalize* them to estimate the unknown rotation. The joint (approximate) diagonalization is captured by the optimization problem

$$\boldsymbol{U} = \underset{\boldsymbol{U}:\ \boldsymbol{U}^T\boldsymbol{U}=\boldsymbol{I}}{\operatorname{argmax}} \sum_{i=1}^{p}\sum_{j=1}^{p} \|\mathrm{diag}(\boldsymbol{U}\boldsymbol{C}^{ij}\boldsymbol{U}^T)\|^2. \quad (6)$$

For a technique for solving (6) using Jacobi angles, see [13]. The statistical properties of FOBI and JADE are given in [15], and JADE is in general considered as the preferred method of these two. When comparing ICA methods for fMRI data, JADE is also often included in comparisons. For recent papers, see for example [16], [17].

In [10] the matrices $\boldsymbol{C}_{ij}$ are extended separately for all modes of a random tensor using again the operation $\odot_{-m}$ to yield the following sets of matrices:

$$\boldsymbol{C}_m^{ij} = \boldsymbol{B}_m^{ij} - \boldsymbol{S}_m\left(\delta_{ij}\rho_m\boldsymbol{I} + \boldsymbol{E}^{ij} + \boldsymbol{E}^{ji}\right)\boldsymbol{S}_m^T,$$

where $i, j = 1, \ldots, p$ and $m = 1, \ldots, r$. Here $\boldsymbol{B}_m^{ij} := \rho_m^{-1}E\left[\boldsymbol{e}_i^T(\mathcal{X}_{st} \odot_{-m} \mathcal{X}_{st})\boldsymbol{e}_j \cdot (\mathcal{X}_{st} \odot_{-m} \mathcal{X}_{st})\right]$, and $\boldsymbol{S}_m := \rho_m^{-1}E\left[\mathcal{X}_{st} \odot_{-m} \mathcal{X}_{st}\right]$ is the m-mode covariance matrix of the standardized tensor which is needed to estimate the unknown constant of proportionality in (3). [10] then showed that the orthogonal matrix $\boldsymbol{U}_m$ can be estimated by replacing the matrices $\boldsymbol{C}_{ij}$ in (6) by $\boldsymbol{C}_{ij}^m$ and performing the joint diagonalization separately for all modes. For identifiability we must further assume that, for all $m = 1, \ldots, r$, at most one of the average kurtoses of the m-mode slices is zero. An alternative, computationally more intensive but similar in performance, version of TJADE also exists, see [10].

D. The proposed method

To provide a fully tensorial alternative for the temporal ICA we still need a counterpart for the singular value decomposition used to reduce the dimension of the initial random vector. This is given by *tensorPCA* (TPCA), which is based on the use of the m-mode covariance matrices in a similar fashion as regular PCA is based on the use of the covariance matrix. The TPCA transformation is given by

$$\mathcal{X} \mapsto \mathcal{X}_{PCA} := \mathcal{X} \odot_{m=1}^{r} \boldsymbol{V}_m^T, \quad (7)$$

where the orthogonal matrices $\boldsymbol{V}_m$ have the eigenvectors of the m-mode covariance matrices $\boldsymbol{\Sigma}_m = \boldsymbol{V}_m\boldsymbol{\Lambda}_m\boldsymbol{V}_m^T$ as their columns, $m = 1, \ldots, r$. It is easy to see that the diagonal matrices $\boldsymbol{\Lambda}_m$ obtained as a side product contain the sums of variances of the components of the m-mode faces of the transformed tensors, $m = 1, \ldots, r$. The diagonal elements can be used as in regular PCA to assess the importance of each column vector of $\boldsymbol{V}_m$ in capturing the information content of $\mathcal{X}$. For example, scree plots for choosing suitable numbers of components are easily generated. Denoting by $\boldsymbol{V}_m^* \in \mathbb{R}^{p_m \times d_m}$ the matrix that contains the chosen d_m columns of $\boldsymbol{V}_m$, $m = 1, \ldots, r$, a reduced $d_1 \times \cdots \times d_r$ tensor is obtained as $\mathcal{X}_{PCA}^* := \mathcal{X} \odot_{m=1}^{r} (\boldsymbol{V}_m^*)^T$ and is analogous to using regular PCA to reduce the dimension of a random vector.

Our proposed regime for processing fMRI data uses TPCA to compress the observed random tensor prior to utilizing the ICA method, and is as follows:

1) Use TPCA to obtain the transformed random tensor $\mathcal{X}_{PCA}$ and for each mode $m = 1, \ldots, r$ retain the indices corresponding to the d_m highest eigenvalues in $\boldsymbol{\Lambda}_m$, yielding the reduced tensor $\mathcal{X}_{PCA}^*$.
2) Subject the reduced random tensor $\mathcal{X}_{PCA}^*$ to either TFOBI or TJADE and based on a chosen criteria choose individual components of interest from the resulting tensor of independent components $\mathcal{Z}$.

Although our approach to tensorial PCA is original, several other formulations for extending PCA for tensor observations have been proposed. For approaches with a statistical viewpoint, see e.g. [18], [19]. For more algorithmic approaches using various tensor decompositions such as the Tucker- and CP-decomposition, see e.g. the review in [20].

IV. Simulation

The simulations were performed in R version 3.3.0 [21] using the packages *ggplot2* [22], *neuRosim* [23] *JADE* [24]

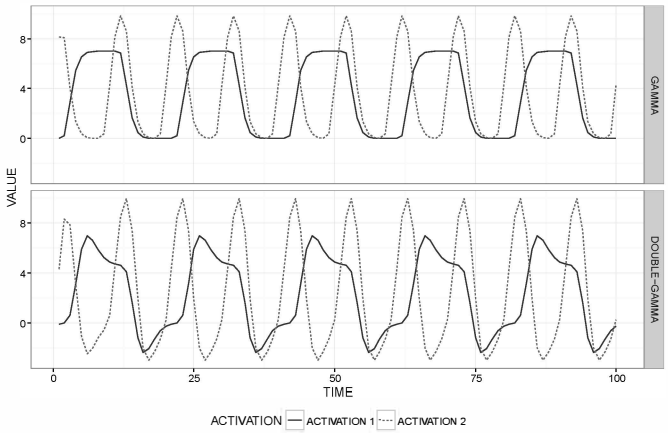

Fig. 1. The activation signals in the simulated data. The upper plot shows the signals of the two activation regions when the option "gamma" is chosen and the lower plot when "double-gamma" is chosen, see Appendix.

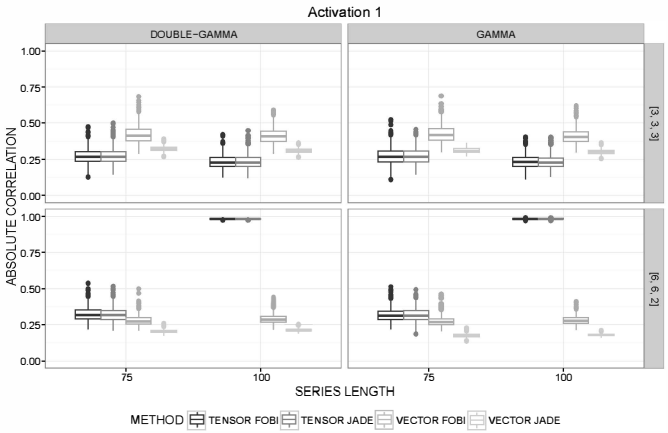

Fig. 2. The boxplots of the highest absolute correlations that any of the independent components had with the true signal of the first activation. The number of repetitions per setting is 1000.

and *tensorBSS* [25], from which the implementation of the tensor regime discussed in Section III-D can be found.

A. Simulation setting

The R-package *neuRosim* [23] provides a tool for simulating realistic resting-state and task-based fMRI data, and allows one to control the sources of noise and many other effects. In order to compare the methods of interest we use the package to simulate 4-dimensional task-based fMRI data (3 spatial and 1 temporal dimensions). The chosen dimensions for the voxels are $p_1 = p_2 = p_3 = 64$ and for the time $T = 75$ or $T = 100$. In general we use mainly the default settings of *neuRosim* and consider two haemodynamic response functions (HRF) with which the stimulus function is convoluted to model the activation produced by the repetition of the task. The two HRFs under consideration are *gamma HRF* and *double-gamma HRF*. Furthermore, several noise sources are added to the observations, comprising of Rician system noise, temporal noise, low-frequency drift, physiological noise, task-related noise and spatial noise, see [23]. The general code for simulating the data is given in the Appendix. The following results will be based on 1000 repetitions for the different settings.

The generated datasets contain two separate activation regions, see Appendix for more information and a visualization. In both of these regions a single signal is localized and the different signals that result from convoluting the stimulus with either the gamma HRF or double-gamma HRF are depicted in Figure 1. The goal in our simulation study is the accurate estimation of these signals from the observed tensors, and two different methods, a tensor-based and a vector-based, are used to carry this out.

The first of the methods is the two-step tensor regime proposed in Section III-D. In step 1 we consider two reduced tensor sizes, namely $3 \times 3 \times 3$ and $6 \times 6 \times 2$. These represent respectively the ideas that most of the information can be contained in a small sub-tensor and that the height dimension (the last one) contains little information as compared to the other two. The resulting reduced tensor is then subjected to either TFOBI or TJADE to obtain a tensor of independent components.

The second method (tICA) starts by vectorizing the observed tensors resulting in a sample of vectors of length $p = 64^3$. Then the singular value decomposition is used to reduce the size of the $T \times p$ data matrix to $T \times p^*$ where the variable dimension p^* is either 27 or 72 to correspond with the number of components retained in the tensor scheme. To obtain the vectors of p^* independent component, FOBI and JADE are used. However, regular JADE has the disadvantage that its computational complexity increases dramatically with p. To overcome this issue, [26] introduced the so-called k-JADE procedure. In k-JADE the whitening is performed using FOBI and the orthogonal matrix $\boldsymbol{U}$ is then found using only those matrices $\boldsymbol{C}^{ij}$ in (5), where $|i - j| \leq k$. The tuning parameter k can be seen as the upper limit of components with identical kurtosis values. As such we use k-JADE with $k = 5$ in place of JADE in the case $p^* = 72$.

B. Simulation results

For each of the methods, the highest absolute correlations between the estimated individual independent components and the two true activation signals given in Figure 1 were recorded. As the absolute correlations measure the success of the estimation, the value of one means that the activation signal was captured perfectly. The boxplots of the values are shown individually for both activation signals in Figures 2 and 3 under all combinations of the simulations settings.

The upper rows of the figures correspond to using the $3 \times 3 \times 3$ reduced tensors (or the vectors of length 27) and the low correlations therein reveal that the reduced tensor is not large enough to contain the information on either of the activation signals. The vector methods perform slightly better, showing median absolute correlations of almost 0.5. However, the fact that the performance does not increase with sample size is also

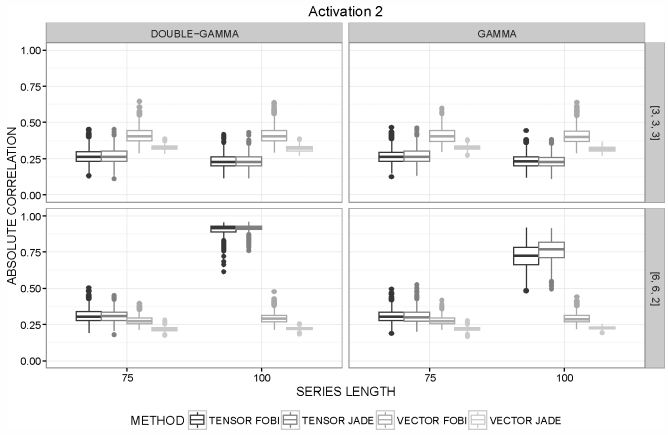

Fig. 3. The boxplots of the highest absolute correlations that any of the independent components had with the true signal of the second activation. The number of repetitions per setting is 1000.

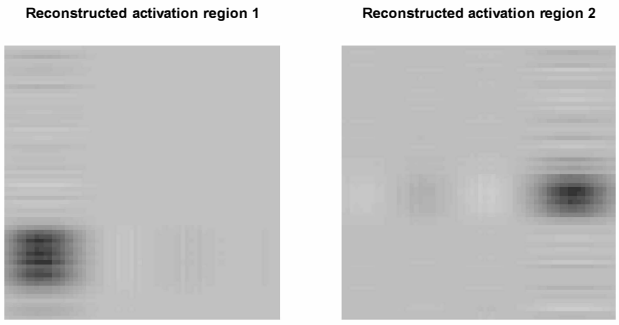

Fig. 4. Reconstructions of the first time point of the first height layer of the observations using only the estimated first signal (left-hand side) or the estimated second signal (right-hand side). The sample size was $T = 100$, the HRF double-gamma, the size of the reduced tensors $6 \times 6 \times 2$ and the used ICA method TJADE.

indicating that most of the signal information is not contained in the 27 components retained by SVD.

Turning our attention to the lower row where the algorithm used the $6 \times 6 \times 2$ reduced tensors (or the vectors of length 72), we see the tensor methods outmatching their vector counterparts. Although the sample size $T = 75$ seems not to be high enough for the estimation of the signals, for $T = 100$ the first activation signal is consistently being perfectly estimated by both TFOBI and TJADE. Also the second signal is very nicely estimated by both tensorial methods for $T = 100$. Interestingly, the functional form of the signal, *gamma HRF* or *double-gamma HRF*, had overall very little effect on the results.

Using the signal estimates one can also create reconstructions of the original observations where the activation regions are clearly visible, see Figure 4 and compare it to the true regions in Figure 5. This is analogous to using regular PCA to recreate original images using only a few principal components and can be carried out in practice e.g. by setting all other elements but the one containing the signal in the tensors $\mathcal{Z}_i$ to zero and carrying out the linear ICA and TPCA transformations backwards.

V. Discussion

The standard way of treating fMRI data, vectorizing and applying vector methods, has the drawbacks of being both computationally intensive and unaware of the tensorial structure of the observations. A more natural course of action should thus preserve the tensor form of the observations. Based on this paradigm, we proposed in this paper a fully tensorial alternative to the commonly used tICA framework and showed that the former was superior in estimating the task signals in simulated fMRI data.

Future work includes the extensions of various aspects of the current setting: The simulation setup could be made more complex, e.g. by letting multiple activation regions overlap. The selection of the number of components in the PCA-step should be investigated carefully while considering also alternative reduction methods, such as the various tensor decompositions. Other tensorial ICA methods such as the tensor versions of k-JADE and FastICA [27] will be developed and investigated. Tensorial blind source separation (BSS) methods, such as the tensorial SOBI (TSOBI) [28], that take into account both the spatial and temporal dependence will be considered. And finally, in future comparisons naturally also results for real fMRI data will be of interest as well as the extending of model (1) for the case of multisubject data.

Acknowledgment

This work was supported by the Academy of Finland grants 268703 and 251965.

Appendix

The data generating code

The following wrapper function was used to generate the simulation data of Section III-D. The function takes as arguments the length T of the series as `n` and the haemodynamic response function, either `"gamma"` or `"double-gamma"`, as `hrf`. The output is a list containing the observation tensor and the true underlying activation signals.

```
nsWrap <- function(n, hrf){

  regions <- simprepSpatial(regions = 2,
        coord = list(c(10, 15, 50),
        c(53, 29, 24)), radius = c(8, 5),
        form = "sphere")
  TR <- 2
  total <- TR * n
  os1 <- seq(1, total, 40)
  os2 <- seq(15, total, 20)
  dur <- list(20, 7)
  os <- list(os1, os2)
  effect <- list(7, 10)

  design <- simprepTemporal(
        regions = 2, onsets = os,
        durations = dur, TR = TR, hrf = hrf,
```

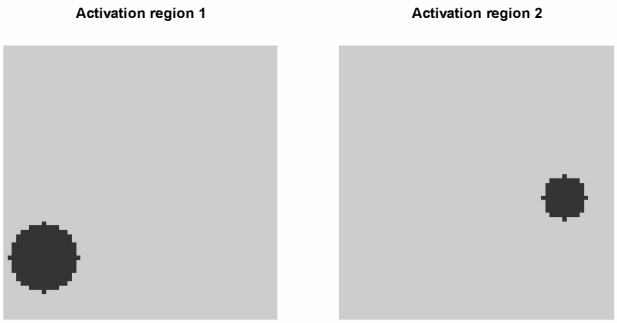

Fig. 5. The left-hand side plot shows where the first activation region resides in the 50th height layer of the observations and the right-hand side plot does the likewise for the second activation region in the 24th height layer.

```
        effectsize = effect,
        totaltime = total)

  w <- c(0.3, 0.3, 0.01, 0.09, 0.1, 0.2)
  x <- simVOLfmri(dim = c(64, 64, 64),
        base = 100, design = design,
        image = regions, SNR = 10,
        noise = "mixture", type = "rician",
        weights = w, verbose = FALSE)

  s <- specifydesign(totaltime = total,
        onsets = list(os1, os2),
        durations = dur, effectsize = effect,
        TR = TR, conv = hrf)
  return(list(x = x, s = s))
}
```

THE ACTIVATION REGIONS

The activation regions of the two signals in the simulated data are spheres with radii 8 and 5 and centers at coordinates $(10, 15, 50)$ and $(53, 29, 24)$, respectively. These two regions are depicted in Figure 5.

REFERENCES

[1] A. H. Andersen and W. S. Rayens, "Structure-seeking multilinear methods for the analysis of fMRI data," *NeuroImage*, vol. 22, no. 2, pp. 728–739, 2004.

[2] E. Martinez-Montes, P. A. Valdes-Sosa, F. Miwakeichi, R. I. Goldman, and M. S. Cohen, "Concurrent EEG/fMRI analysis by multiway partial least squares," *NeuroImage*, vol. 22, no. 3, pp. 1023–1034, 2004.

[3] L. D. Kuang, Q. H. Lin, X. F. Gong, J. Fan, F. Y. Cong, and V. D. Calhoun, "Multi-subject fMRI data analysis: Shift-invariant tensor factorization vs. group independent component analysis," in *IEEE China Summit International Conference on Signal and Information Processing (ChinaSIP), 2013*, 2013, pp. 269–272.

[4] V. Calhoun, T. Adali, L. K. Hansen, J. Larsen, and J. Pekar, "ICA of functional MRI data: An overview," in *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, Nara, Japan, 2003, pp. 281–288.

[5] V. D. Calhoun and T. Adali, "Unmixing fMRI with independent component analysis," *IEEE Engineering in Medicine and Biology Magazine*, vol. 25, no. 2, pp. 79–90, 2006.

[6] B. P. Roś, F. Bijma, M. C. de Gunst, and J. C. de Munck, "A three domain covariance framework for EEG/MEG data," *Neuroimage*, vol. 119, pp. 305–315, 2015.

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

[8] N. E. Helwig and S. Hong, "A critique of tensor probabilistic independent component analysis: Implications and recommendations for multi-subject fMRI data analysis," *Journal of Neuroscience Methods*, vol. 213, no. 2, pp. 263–273, 2013.

[9] J. Virta, B. Li, K. Nordhausen, and H. Oja, "Independent component analysis for tensor-valued data," *Preprint available as arXiv:1602.00879*, 2016.

[10] ——, "JADE for tensor-valued observations," *Preprint available as arXiv:1603.05406*, 2016.

[11] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.

[12] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York, USA: John Wiley & Sons, 2001.

[13] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," in *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, no. 6. IET, 1993, pp. 362–370.

[14] J.-F. Cardoso, "Source separation using higher order moments," in *International Conference on Acoustics, Speech, and Signal Processing, 1989. ICASSP-89.* IEEE, 1989, pp. 2109–2112.

[15] J. Miettinen, S. Taskinen, K. Nordhausen, and H. Oja, "Fourth moments and independent component analysis," *Statistical Science*, vol. 30, no. 3, pp. 372–390, 2015.

[16] N. Correa, T. Adali, and V. D. Calhoun, "Performance of blind source separation algorithms for fMRI analysis using a group ICA method," *Magnetic Resonance Imaging*, vol. 25, no. 5, pp. 684–694, 2007.

[17] B. B. Risk, D. S. Matteson, D. Ruppert, A. Eloyan, and B. S. Caffo, "An evaluation of independent component analyses with an application to resting-state fMRI," *Biometrics*, vol. 70, no. 1, pp. 224–236, 2014.

[18] S. Ding and R. D. Cook, "Dimension folding PCA and PFC for matrix-valued predictors," *Statistica Sinica*, vol. 24, pp. 463–492, 2014.

[19] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Multilinear principal component analysis of tensor objects for recognition," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 2. IEEE, 2006, pp. 776–779.

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

[21] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016. [Online]. Available: https://www.R-project.org/

[22] H. Wickham, *ggplot2: elegant graphics for data analysis*. Springer New York, 2009.

[23] M. Welvaert, J. Durnez, B. Moerkerke, G. Verdoolaege, and Y. Rosseel, "neuRosim: An R package for generating fMRI data," *Journal of Statistical Software*, vol. 44, no. 10, pp. 1–18, 2011. [Online]. Available: http://www.jstatsoft.org/v44/i10/

[24] K. Nordhausen, J.-F. Cardoso, J. Miettinen, H. Oja, E. Ollila, and S. Taskinen, *JADE: Blind Source Separation Methods Based on Joint Diagonalization and Some BSS Performance Criteria*, 2015, R package version 1.9-93. [Online]. Available: https://CRAN.R-project.org/package=JADE

[25] J. Virta, K. Nordhausen, H. Oja, and B. Li, *tensorBSS: Blind Source Separation Methods for Tensor-Valued Observations*, 2016, R package version 0.2. [Online]. Available: https://CRAN.R-project.org/package=tensorBSS

[26] J. Miettinen, K. Nordhausen, H. Oja, and S. Taskinen, "Fast equivariant JADE," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2013*, May 2013, pp. 6153–6157.

[27] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, pp. 626–634, 1999.

[28] J. Virta and K. Nordhausen, "Blind source separation of tensor-valued time series," *Submitted*, 2016.