Low-Rank Tucker-2 Model for Multi-Subject fMRI Data Decomposition With Spatial Sparsity Constraint

Yue Han, Qiu-Hua Lin, *Member, IEEE*, Li-Dan Kuang, Xiao-Feng Gong, *Member, IEEE*, Fengyu Cong, *Senior Member, IEEE*, Yu-Ping Wang, *Senior Member, IEEE*, and Vince D. Calhoun, *Fellow, IEEE*

***Abstract*— Tucker decomposition can provide an intuitive summary to understand brain function by decomposing multi-subject fMRI data into a core tensor and multiple factor matrices, and was mostly used to extract functional connectivity patterns across time/subjects using orthogonality constraints. However, these algorithms are unsuitable for extracting common spatial and temporal patterns across subjects due to distinct characteristics such as high-level noise. Motivated by a successful application of Tucker decomposition to image denoising and the intrinsic sparsity of spatial activations in fMRI, we propose a low-rank Tucker-2 model with spatial sparsity constraint to analyze multi-subject fMRI data. More precisely, we propose to impose a sparsity constraint on spatial maps by using an ℓ_p norm ($0 < p \leq 1$), in addition to adding low-rank constraints on factor matrices via the Frobenius norm. We solve the constrained Tucker-2 model using alternating direction method of multipliers, and propose to update both sparsity and low-rank constrained spatial maps using half quadratic splitting. Moreover, we extract new spatial and temporal features in addition to subject-specific intensities from the core tensor, and use these features to classify multiple subjects. The results from both simulated and experimental fMRI data verify the improvement of the proposed method, compared with four related algorithms including robust Kronecker component analysis, Tucker decomposition with orthogonality constraints, canonical polyadic decomposition, and block term decomposition in extracting common spatial and temporal components across subjects. The spatial and temporal features extracted from the core tensor show promise for characterizing subjects within the same group of patients or healthy controls as well.**

***Index Terms*— Tucker decomposition, multi-subject fMRI data, sparsity constraint, low-rank, core tensor.**

I. Introduction

TENSOR decompositions have attracted increasing attention in blind source separation (BSS) of multi-subject functional magnetic resonance imaging (fMRI) data, since it can be readily represented by a three-way/mode tensor (voxel × time × subject). Canonical polyadic decomposition (CPD), Tucker decomposition (TKD), and block term decomposition (BTD) are commonly adopted forms for tensor decomposition. CPD decomposes a three-way tensor into a linear combination of a series of rank-one tensors, and each rank-one tensor is the outer product of three loading vectors [1], [2]. TKD decomposes a three-way tensor into a core tensor and three factor matrices. It can be seen as a complete form of CPD with polyadic but not necessarily canonical expansion in rank-1 terms [3], and CPD is a special case of TKD model with a diagonal core tensor [4], [5]. BTD decomposes a tensor into a sum of low multilinear rank tensors [6]–[8], thus is a complete form of TKD [6], [7] or a constrained TKD with a block diagonal core tensor [8]. Among the three forms, BTD was used to decompose 4-way fMRI tensor (e.g., 2D slices × time × subject) [9], while CPD and TKD were suitable to decompose three-way fMRI data [1], [2], [10], [11]. CPD has found its effectiveness in fMRI analyses by extracting spatial maps (SMs), time courses (TCs), and subject-specific intensities from multi-subject fMRI data [1], [2], which provide an intuitive decomposition for

Manuscript received June 18, 2021; revised September 10, 2021; accepted October 14, 2021. Date of publication October 22, 2021; date of current version March 2, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61871067, Grant 61379012, Grant 61901061, Grant 61671106, Grant 61331019, and Grant 81471742; in part by the National Science Foundation (NSF) under Grant 1539067, Grant 0840895, Grant 1539067, and Grant 0715022; in part by the National Institutes of Health (NIH) under Grant R01MH104680, Grant R01MH107354, Grant R01EB005846, Grant 5P20GM103472, and Grant R56MH124925; in part by the Fundamental Research Funds for the Central Universities, China, under Grant DUT20ZD220; and in part by the Supercomputing Center of Dalian University of Technology. *(Corresponding author: Qiu-Hua Lin.)*

Yue Han, Qiu-Hua Lin, and Xiao-Feng Gong are with the School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China (e-mail: h_yue@mail.dlut.edu.cn; qhlin@dlut.edu.cn; xfgong@dlut.edu.cn).

Li-Dan Kuang is with the School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha 410114, China (e-mail: kuangld@csust.edu.cn).

Fengyu Cong is with the School of Biomedical Engineering, Dalian University of Technology, Dalian 116024, China, and also with the Faculty of Information Technology, University of Jyväskylä, 40014 Jyväskylä, Finland (e-mail: cong@dlut.edu.cn).

Yu-Ping Wang is with the Department of Biomedical Engineering, Tulane University, New Orleans, LA 70118 USA (e-mail: wyp@tulane.edu).

Vince D. Calhoun is with the Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State University, Georgia Institute of Technology, Emory University, Atlanta, GA 30303 USA (e-mail: vcalhoun@gsu.edu).

Digital Object Identifier 10.1109/TMI.2021.3122226

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

understanding brain function and mental disorders [12]–[16]. For TKD, the factor matrices reflect the affiliation of orthogonal bases along each mode [4], [13]. The model is named as Tucker-2 when one of the factor matrices (commonly the third mode) is an identity matrix, and Tucker-1 is a TKD model with two identity matrices [4]. As a complete form of CPD [5], Tucker-2 model can similarly obtain shared SMs and TCs from the two factor matrices and subject-specific intensities from the core tensor. But the biggest difference with CPD is that the core tensor of Tucker-2 model contains additional interactions between different components [4], [13] and highly compressed information for subjects [3], [17]. Some researchers studied the dynamic functional connectivity networks of fMRI using the core tensor of TKD [10], [11], [18], [19] and the core tensor was proven to carry features of original data in image processing [20] and analyses of electroencephalography (EEG) data [3], [17].

However, the decomposition of the TKD model without constraints is not unique; constraints should be imposed on factor matrices and/or the core tensor [4], [13], [17], [21]. Among other constraints, higher-order SVD (HOSVD) and higher-order orthogonal iteration (HOOI) are two TKD models with orthogonality constraints on the core tensor and factor matrices [13]. HOSVD decompose a tensor into orthogonal matrices and all-orthogonal core tensor, while HOOI iteratively updates an orthonormal basis of the dominant subspace [4]. HOSVD has been utilized to identify dynamic functional connectivity (FC) patterns of multi-subject fMRI data from a third-order [10] or a fourth-order tensor [18] (1D/2D connection $\times$ time $\times$ subject), or to extract connectivity networks from a fourth-order fMRI tensor computed from a single-subject fMRI data [19]. Al-Sharoa *et al.* proposed to extract FC patterns across time and subjects using HOOI based on the tensor formed by adjacency matrices [11]. HOOI was also applied to decompose four-way EEG data to obtain a core tensor and factor matrices in each mode [5], [21]. The core tensor can be sequentially used to train a standard classifier [5] or seen as the links among the components from different modes [21], and the factor matrices can be utilized to extract the components of frequency/time/channel/trial [21].

In addition to the orthogonality constraint, other constraints were also used in TKD-based image processing. Lu *et al.* proposed to recover clear images based on TKD, which obtains underlying tensor (clear image) with a low-tubal-rank using nuclear norm constraint, and a sparse residual tensor using sparsity constraint [22]. Besides the nuclear norm, Frobenius norm (ℓ_{F} norm) regularization can also improve low-rankness. Grussler *et al.* verified that the rank-constrained Frobenius norm can outperform the nuclear norm [23]. Traoré *et al.* proposed to enforce a sparsity constraint on the core tensor and low-rank (ℓ_{F} norm) constraints on factor matrices, and learnt reconstructed patch for inpainting images using the factor matrices and the core tensor [24]. The approach was also flexible to incorporate non-negativity and orthogonality constraints [24]. Bahri *et al.* proposed Kronecker component analysis (RKCA) method for image denoising. They justified that ℓ_{F} norm and nuclear norm penalties can yield the same optimal solutions of low-rank factor matrices. The model was established by imposing ℓ_{F} norm constraints on the separated dictionaries (Kronecker product of two factor matrices) and sparsity constraints on the core and residual tensors [25].

As mentioned above, existing TKD algorithms for fMRI analyses mostly utilize orthogonality constraints to extract FC patterns across time and subjects. These algorithms, however, are unsuitable for extracting common spatial and temporal patterns across subjects due to distinct characteristics such as high-level noise [26]. Motivated by the success of TKD-based image denoising algorithms such as RKCA, we propose to enforce fMRI-specific constraints on Tucker-2 model to analyze three-way multi-subject fMRI data (voxel $\times$ time $\times$ subject). First, the spatial activity of fMRI is proven to be sparse. Xu *et al.* investigated that only a small percentage ($<$10%) of the entire neuronal population can be fired in any brain region at any moment [27]. Lennie found that the number of activated neurons accounted for less than 1% in human cortex [28]. The sparsity characteristic of fMRI signal is more promising [29]. As such, we propose to impose a spatial sparsity constraint on the shared SMs, which to our best knowledge has not been used in TKD-based fMRI analyses. In addition, the factor matrices representing the shared SMs and TCs should be low-rank in order to extract principal spatial and temporal components, and the core tensor should be sparse to improve the uniqueness of the TKD solution [4]. Therefore, imposing low-rank constraints on factor matrices and adding sparsity constraint on the core tensor are essential in TKD of fMRI data.

In this study, we build a Tucker-2 model with both sparsity and low-rank constraints to represent the characteristics of three-way fMRI data (voxel $\times$ time $\times$ subject). More precisely, we propose to impose sparsity constraint on SMs via an ℓ_p norm regularization ($0< p \leq 1$), which is often used in place of ℓ_0 norm [30]. Besides, since previous researches have justified that low-rankness can be promoted by Frobenius norm [22]–[25], we impose low-rank (ℓ_{F} norm) constraints on factor matrices to detect principal SM and TC components shared by all subjects, and impose sparsity constraints via an ℓ_1 norm regularization on the core tensor and the residual tensor to improve the uniqueness and performance. We estimate the core tensor and factor matrices by alternating direction method of multipliers (ADMM) and half quadratic splitting (HQS). In addition to the shared SMs and TCs estimates, we propose to explore novel information from the core tensor, and extract spatial and temporal features as well as individual intensities. These spatial and temporal features extracted from the core tensor are sequentially used to cluster multiple subjects into different groups. Simulated and experimental fMRI data are used to evaluate the proposed method.

To summarize, our contributions are as follows:

(1) We propose a low-rank Tucker-2 model with a spatial sparsity constraint to decompose multi-subject fMRI data into shared SMs, shared TCs, and a core tensor

TABLE I
DEFINITIONS OF NOTATIONS

Notation	Meaning
$\times_n$	mode-n product
$\langle\cdot\rangle$	inner product of matrices
$\lvert\cdot\rvert$	element-wise absolute value
$sgn(\cdot)$	element-wise sign function
$\circ$	Hadamard product
$./$	element-wise division
$\dagger$	pseudo-inverse of matrix

including novel information in addition to subject-specific intensities.

(2) We propose to enforce the sparsity constraint on SMs via an ℓ_p $(0 < p \leq 1)$ norm regularization, and adopt low-rank constraints on the two factor matrices via the Frobenius norm regularization and impose sparsity constraints on the core and residual tensors via the ℓ_1 norm regularization. We solve the proposed Tucker-2 model using ADMM, and propose to update SMs with both sparsity and low-rank constraints using HQS.

(3) We propose to explore the novel information embedded in the core tensor. In addition to the recovery of subject-specific intensities, we extract spatial and temporal features, which are then used to classify multiple subjects.

The rest of this paper is organized as follows. Section 2 describes the proposed method in detail. Section 3 presents the fMRI data used in our experiments and performance metrics adopted. Section 4 has the results from both simulated and experimental fMRI data. Section 5 includes the discussion of this study.

II. METHODS

In this section, we first describe the proposed method, named as slcTKD (sparsity-low rank-constraints TKD). Afterwards, we introduce the update rules of factor matrices, core and residual tensors, as well as multipliers and penalty parameters. We then present a detailed implementation of the proposed slcTKD algorithm. Finally, we present the method of extracting the subject-specific intensities as well as spatial and temporal features.

A. Notations

In the proposed method, tensors, matrices, and vectors are denoted by underlined bold capital letters (e.g., $\underline{\mathbf{X}}$), bold uppercase letters (e.g., $\mathbf{B}$), and bold lowercase letters (e.g., $\mathbf{b}$), respectively. Frontal (mode-3) slices of tensor $\underline{\mathbf{X}}$ are defined as $\underline{\mathbf{X}}_k = \underline{\mathbf{X}}(:, :, k)$, where ":" means that all samples are used. We denote the norm of tensor as $||\underline{\mathbf{X}}|| = \sum_k ||\mathbf{X}_k||$, where "$||\cdot||$" represents an ℓ_p, ℓ_1, or ℓ_F norm regularization. Table I includes the definitions of other notations used in this study.

B. The Proposed slcTKD Model

Given a three-way (voxel × time × subject) multi-subject fMRI data $\underline{\mathbf{X}} \in \mathbb{R}^{V\times T\times K}$, where V is the number of in-brain voxels, T is the number of time points, and K is the number of subjects. The Tucker-2 model of $\underline{\mathbf{X}}$ is built as follows:

$$\underline{\mathbf{X}} = \underline{\mathbf{G}} \times_1 \mathbf{S} \times_2 \mathbf{B} + \underline{\mathbf{E}} \tag{1}$$

where $\mathbf{S} \in \mathbb{R}^{V\times N}$ and $\mathbf{B} \in \mathbb{R}^{T\times N}$ are two factor matrices corresponding to the shared SMs and TCs, and N is the number of components (i.e., the model order); $\underline{\mathbf{G}} \in \mathbb{R}^{N\times N\times K}$ is the core tensor including subject-specific intensities, and spatial and temporal features; and $\underline{\mathbf{E}} \in \mathbb{R}^{V\times T\times K}$ is the residual tensor.

Taking into account all sparsity and low-rank constraints imposed on each of factor matrices and tensors in the model (1), we propose slcTKD model for analyzing multi-subject fMRI data as follows:

$$\min_{\mathbf{S},\mathbf{B},\underline{\mathbf{G}},\underline{\mathbf{E}}} \left\|\underline{\mathbf{X}} - \underline{\mathbf{G}} \times_1 \mathbf{S} \times_2 \mathbf{B} - \underline{\mathbf{E}}\right\|_F^2 + \|\mathbf{S}\|_F^2 + \|\mathbf{B}\|_F^2 + \delta\|\mathbf{S}\|_p + \lambda\left\|\underline{\mathbf{G}}\right\|_1 + \gamma\left\|\underline{\mathbf{E}}\right\|_1 \tag{2}$$

where the sparsity constraint on $\mathbf{S}$ is imposed via the ℓ_p norm $(0 < p \leq 1)$; sparsity constraints on the core and residual tensors are imposed by the ℓ_1 norm. Positive parameters δ, λ, and γ control the effects of the sparsity constraints, and the use of both p and δ makes the spatial sparsity level closer to that of actual SMs. The Frobenius norm of factor matrices $\mathbf{S}$ and $\mathbf{B}$ promote the low-rankness of shared SMs and TCs [25].

The problem in (2) is a convex optimization problem for each component individually. The work in [22], [31], and [32] exhibited the efficiency of ADMM in solving similar problems. We also take advantage of ADMM to update the factor matrices and tensors. By introducing a split variable $\underline{\mathbf{R}}$ for the core tensor [31], the slcTKD model in (2) is rewritten as follows:

$$\begin{aligned}&\min_{\mathbf{S},\mathbf{B},\underline{\mathbf{G}},\underline{\mathbf{E}},\underline{\mathbf{R}}} \|\mathbf{S}\|_F^2 + \|\mathbf{B}\|_F^2 + \delta\|\mathbf{S}\|_p + \lambda\left\|\underline{\mathbf{G}}\right\|_1 + \gamma\left\|\underline{\mathbf{E}}\right\|_1\\ &s.t.\ \underline{\mathbf{X}} = \underline{\mathbf{R}} \times_1 \mathbf{S} \times_2 \mathbf{B} + \underline{\mathbf{E}}\\ &s.t.\ \underline{\mathbf{G}} = \underline{\mathbf{R}}\end{aligned} \tag{3}$$

The mode-1 and mode-2 product of factor matrices with the core tensor can be expressed in terms of matrices multiplication:

$$\underline{\mathbf{R}} \times_1 \mathbf{S} \times_2 \mathbf{B} = \sum\nolimits_k \mathbf{S}\underline{\mathbf{R}}_k\mathbf{B}^T \tag{4}$$

As such, the augmented Lagrangian method [25], [31] derived from (3) is formulated as follows:

$$\begin{aligned}&L\left(\mathbf{S},\mathbf{B},\underline{\mathbf{G}},\underline{\mathbf{E}},\underline{\mathbf{R}},\underline{\mathbf{U}},\underline{\mathbf{W}},\alpha,\beta\right)\\ &= \|\mathbf{S}\|_F^2 + \|\mathbf{B}\|_F^2 + \delta\|\mathbf{S}\|_p + \lambda\left\|\underline{\mathbf{G}}\right\|_1\\ &\quad + \gamma\left\|\underline{\mathbf{E}}\right\|_1 + \alpha\sum\nolimits_k \left\|\underline{\mathbf{X}}_k - \mathbf{S}\underline{\mathbf{R}}_k\mathbf{B}^T - \underline{\mathbf{E}}_k\right\|_F^2 + \beta\left\|\underline{\mathbf{G}} - \underline{\mathbf{R}}\right\|_F^2\\ &\quad + \sum\nolimits_k \left\langle\underline{\mathbf{X}}_k - \mathbf{S}\underline{\mathbf{R}}_k\mathbf{B}^T - \underline{\mathbf{E}}_k, \underline{\mathbf{U}}_k\right\rangle + \sum\nolimits_k \left\langle\underline{\mathbf{G}}_k - \underline{\mathbf{R}}_k, \underline{\mathbf{W}}_k\right\rangle\end{aligned} \tag{5}$$

where $\underline{\mathbf{U}}$ and $\underline{\mathbf{W}}$ are Lagrangian multipliers, and α, β are penalty parameters. By minimizing the augmented Lagrangian problem in (5), the factor matrices, tensors, multipliers, and penalty parameters are iteratively updated until convergence. We present update rules in detail in the following sub-sections.

C. Update Rules of the Proposed slcTKD Model

*1) Updates of Two Factor Matrices **B** and **S**:* The update of factor matrix $\mathbf{B}$ can be derived from (5) by matrix algebra as follows:

$$\mathbf{B} = \frac{\sum_k \left(\alpha\left(\underline{\mathbf{X}}_k - \underline{\mathbf{E}}_k\right) + \underline{\mathbf{U}}_k/2\right)^T \mathbf{S}\underline{\mathbf{R}}_k}{\mathbf{I} + \alpha \sum_k \underline{\mathbf{R}}_k^T \mathbf{S}^T \mathbf{S}\underline{\mathbf{R}}_k} \tag{6}$$

Regarding the factor matrix $\mathbf{S}$ imposed both ℓ_{F} and ℓ_p constraints, it is difficult to be directly determined by matrix algebra. As such, we propose to utilize HQS algorithm to simplify the update of $\mathbf{S}$. We add a variable $\mathbf{Y}$ into (5) and replace the ℓ_p norm regularization on $\mathbf{S}$ with a function of $\mathbf{Y}$. The augmented Lagrangian approach for updating $\mathbf{S}$ and $\mathbf{Y}$ are in the following:

$$\begin{aligned} L\left(\mathbf{S}, \mathbf{B}, \underline{\mathbf{G}}, \underline{\mathbf{E}}, \underline{\mathbf{R}}, \underline{\mathbf{U}}, \underline{\mathbf{W}}, \mathbf{Y}, \mathbf{Q}, \alpha, \beta\right) \\ = \|\mathbf{S}\|_{\mathrm{F}}^2 + \delta\|\mathbf{S} - \mathbf{Y}\|_{\mathrm{F}}^2 + \xi\|\mathbf{Y}\|_p \\ + \alpha \sum_k \left\|\underline{\mathbf{X}}_k - \mathbf{S}\underline{\mathbf{R}}_k\mathbf{B}^T - \underline{\mathbf{E}}_k\right\|_{\mathrm{F}}^2 \\ + \sum_k \left\langle \underline{\mathbf{X}}_k - \mathbf{S}\underline{\mathbf{R}}_k\mathbf{B}^T - \underline{\mathbf{E}}_k, \underline{\mathbf{U}}_k \right\rangle \\ + \langle \mathbf{S} - \mathbf{Y}, \mathbf{Q}\rangle + L_1\left(\mathbf{B}, \underline{\mathbf{G}}, \underline{\mathbf{E}}, \underline{\mathbf{R}}, \underline{\mathbf{U}}, \underline{\mathbf{W}}, \alpha, \beta\right) \end{aligned} \tag{7}$$

where $\xi > 0$, $\mathbf{Q}$ is a Lagrangian multiplier, and L_1 is a part of L in (5) without $\mathbf{S}$. Then we update $\mathbf{S}$ by an algebraic calculation as follows:

$$\mathbf{S} = \frac{\sum_k \left(\alpha\left(\underline{\mathbf{X}}_k - \underline{\mathbf{E}}_k\right) + \underline{\mathbf{U}}_k/2\right)\mathbf{B}\underline{\mathbf{R}}_k^T + \delta\mathbf{Y} + \mathbf{Q}}{(\delta+1)\mathbf{I} + \alpha\sum_k \underline{\mathbf{R}}_k\mathbf{B}^T\mathbf{B}\underline{\mathbf{R}}_k^T} \tag{8}$$

The variable $\mathbf{Y}$ can be obtained using Newton-Raphson method [33], which is carried out by several number of iterations. We calculate $\mathbf{Y}$ in each iteration as follows:

$$\mathbf{Y}_d = \xi p \cdot sgn(\mathbf{Y}) \circ |\mathbf{Y}|^{p-1} + \delta(\mathbf{S} - \mathbf{Y}) - \mathbf{Q} \tag{9}$$

$$\mathbf{Y}_{dd} = \xi p(p-1)|\mathbf{Y}|^{p-2} - \delta\mathbf{1} \tag{10}$$

$$\mathbf{Y} = \mathbf{Y} - \mathbf{Y}_d ./ \mathbf{Y}_{dd} \tag{11}$$

where p is the value of p in ℓ_p norm, and $\mathbf{1}$ is an unit matrix which has the same size with $\mathbf{Y}$, and the elements are all 1.

*2) Updates of Core Tensor **G** and Residual Tensor **E**:* We update the core tensor $\underline{\mathbf{G}}$ using the soft-shrinkage method as follows:

$$\underline{\mathbf{G}} = \sum\nolimits_k \left(sgn(\tilde{\underline{\mathbf{G}}}_k) \circ max\left\{|\tilde{\underline{\mathbf{G}}}_k| - \lambda/2\beta, 0\right\}\right) \tag{12}$$

where $\tilde{\underline{\mathbf{G}}}_k = \underline{\mathbf{R}}_k - \underline{\mathbf{W}}_k/2\beta$.

The split variable $\underline{\mathbf{R}}$ in (5) satisfies discrete-time Sylvester equation [25] in the following form:

$$\alpha\mathbf{S}^T\mathbf{S}\underline{\mathbf{R}}_k\mathbf{B}^T\mathbf{B} + \beta\underline{\mathbf{R}}_k + \beta\underline{\mathbf{G}}_k + \underline{\mathbf{W}}_k/2 - \tilde{\underline{\mathbf{R}}}_k = 0 \tag{13}$$

where $\tilde{\underline{\mathbf{R}}}_k = \mathbf{S}^T\left(\alpha\underline{\mathbf{X}}_k - \alpha\underline{\mathbf{E}}_k + \underline{\mathbf{U}}_k/2\right)\mathbf{B}$. In the proposed algorithm, equation (13) can be solved by solvers for discrete-time Sylvester equations [25], such as Hessenberg-Schur method [34] or Bartds-Stewart algorithm [35].

Based on (5), we can also update the residual tensor $\underline{\mathbf{E}}$ by soft-shrinkage method via the following:

$$\underline{\mathbf{E}} = \sum\nolimits_k \left(sgn(\tilde{\underline{\mathbf{E}}}_k) \circ max\left\{|\tilde{\underline{\mathbf{E}}}_k| - \gamma/2\alpha, 0\right\}\right) \tag{14}$$

where $\tilde{\underline{\mathbf{E}}}_k = \underline{\mathbf{X}}_k - \mathbf{S}\underline{\mathbf{R}}_k\mathbf{B}^T + \underline{\mathbf{U}}_k/2\alpha$.

TABLE II
IMPLEMENTATION OF THE PROPOSED SLCTKD ALGORITHM

Input: multi-subject fMRI data $\underline{\mathbf{X}} \in \mathbf{R}^{I\times T\times K}$; the model order N; the maximum number of iteration $iter_{\max}$; the iteration number of Newton-Raphson method *iter_y;* termination thresholds of the error $\varepsilon_{\min}$ and the relative error change $\Delta\varepsilon_{\min}$; parameters δ, p, γ, λ, η, and ξ

Initialize factor matrices $\mathbf{S}$ and $\mathbf{B}$, core tensor $\underline{\mathbf{G}}$, multipliers $\underline{\mathbf{U}}$, $\underline{\mathbf{W}}$, and $\mathbf{Q}$, penalty parameters α, β; $\mathbf{Y} = \mathbf{S}$, $\underline{\mathbf{R}} = \underline{\mathbf{G}}$, calculate $\underline{\mathbf{E}}$ using $\underline{\mathbf{X}} - \underline{\mathbf{G}} \times_1 \mathbf{S} \times_2 \mathbf{B}$ based on (1); $\varepsilon^0 = 1$; $\Delta\varepsilon^0 = 1$; $iter$=0;
While $iter < iter_{\max}$, $\varepsilon^{iter} > \varepsilon_{\min}$, or $\Delta\varepsilon^{iter} > \Delta\varepsilon_{\min}$ do
  $iter$=$iter$+1;
  Update $\mathbf{B}$ using (6);
  Update $\mathbf{S}$ using (8);
  for $iter_1$=1: $iter_y$
    Calculate $\mathbf{Y}$ using (11);
  end
  Update $\underline{\mathbf{G}}$ using (12);
  Update $\underline{\mathbf{R}}$ from discrete-time Sylvester equation in (13);
  Update $\underline{\mathbf{E}}$ using (14);
  Update multipliers and parameters:
    Update $\underline{\mathbf{U}}$, $\underline{\mathbf{W}}$, and $\mathbf{Q}$ using (15), (16), and (17);
    Update α and β using (18) and (19);
    $\varepsilon^{iter} = \|\underline{\mathbf{X}} - \underline{\mathbf{G}} \times_1 \mathbf{S} \times_2 \mathbf{B} - \underline{\mathbf{E}}\|_{\mathrm{F}} / \|\underline{\mathbf{X}}\|_{\mathrm{F}}$;
    $\Delta\varepsilon^{iter} = |\varepsilon^{iter-1} - \varepsilon^{iter}| / \varepsilon^{iter-1}$;
End while
Output: $\mathbf{S}$, $\mathbf{B}$, $\underline{\mathbf{G}}$;

3) Updates of Multipliers and Penalty Parameters: In the proposed slcTKD algorithm, we update the multipliers and parameters after updates of all factor matrices and tensors in each iteration. Specifically, the multipliers $\underline{\mathbf{U}}$, $\underline{\mathbf{W}}$, and $\mathbf{Q}$ are updated in each iteration of ADMM method as follows:

$$\underline{\mathbf{U}} \leftarrow \underline{\mathbf{U}} + \alpha\left(\underline{\mathbf{X}} - \underline{\mathbf{R}} \times_1 \mathbf{S} \times_2 \mathbf{B} - \underline{\mathbf{E}}\right) \tag{15}$$

$$\underline{\mathbf{W}} \leftarrow \underline{\mathbf{W}} + \beta\left(\underline{\mathbf{G}} - \underline{\mathbf{R}}\right) \tag{16}$$

$$\mathbf{Q} \leftarrow \mathbf{Q} + \delta(\mathbf{Y} - \mathbf{S}) \tag{17}$$

The penalty parameters α and β are updated using the parameter tuning method proposed in [25] as follows:

$$\alpha \leftarrow \eta\alpha \tag{18}$$

$$\beta \leftarrow \eta\beta \tag{19}$$

where $\eta > 1$ is a fixed parameter representing the increase rates of α and β.

D. Implementation of the Proposed slcTKD Algorithm

We use the result from HOSVD, which better matches Tucker-2 model than CPD-related algorithm such as CPD-GEVD [36], as the initialization values of $\mathbf{S}$, $\mathbf{B}$, and $\underline{\mathbf{G}}$ to accelerate the convergence, and let $\underline{\mathbf{U}} = \underline{\mathbf{W}} = \mathbf{0}$, $\mathbf{Q} = \mathbf{0}$, $\alpha_0 = K/||\underline{\mathbf{X}}||_{\mathrm{F}}$, and $\beta_0 = K/||\underline{\mathbf{R}}||_{\mathrm{F}}$. We fix the parameters δ, p, λ and γ in the algorithm, and discuss the selection of four parameters in Section V. The implementation of slcTKD algorithm is summarized in Table II.

E. Extraction of Subject-Specific Intensities and Spatial and Temporal Features From the Core Tensor

The slcTKD model provides a sparse, non-diagonal core tensor, which consists of spatial-temporal-subject information

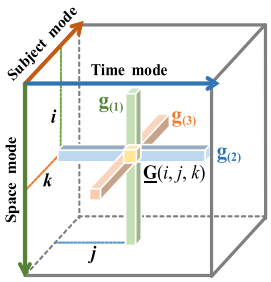


Fig. 1. Fibers of the core tensor.

of the multi-subject fMRI data. When fixing the index of two modes, the three-way core tensor can be represented by fibers [4]. As Fig. 1 shows, we abbreviate the mode-1 (space mode), mode-2 (time mode) and mode-3 (subject mode) fibers as $\mathbf{g}_{(1)} = \underline{\mathbf{G}}(:, j, k)$, $\mathbf{g}_{(2)} = \underline{\mathbf{G}}(i, :, k)$, and $\mathbf{g}_{(3)} = \underline{\mathbf{G}}(i, j, :)$, respectively, where $i, j = 1, 2, \ldots, N$, and $k = 1, 2, \ldots, K$. Based on the fibers in different modes, we next extract the subject- specific intensities and the spatial and temporal features.

1) Subject-Specific Intensities: We first extract subject-specific intensities by analyzing the core tensor. We rewrite the model in (1) by the form of frontal slices as follows:

$$\underline{\mathbf{G}}_k = \mathbf{S}^{\dagger}(\underline{\mathbf{X}}_k - \underline{\mathbf{E}}_k)(\mathbf{B}^T)^{\dagger} \tag{20}$$

where $\underline{\mathbf{G}}_k \in \mathbb{R}^{N\times N}$, $\underline{\mathbf{X}}_k \in \mathbb{R}^{V\times T}$, $\underline{\mathbf{E}}_k \in \mathbb{R}^{V\times T}$. We denote $\tilde{\mathbf{S}} = \mathbf{S}^{\dagger}$, $\tilde{\mathbf{B}} = (\mathbf{B}^T)^{\dagger}$, then we can extract the *kth* component of mode-3 fibers of the core tensor as follows:

$$\mathbf{g}_{(3)k} = \tilde{\mathbf{s}}_i(\underline{\mathbf{X}}_k - \underline{\mathbf{E}}_k)\tilde{\mathbf{b}}_j = \underline{\mathbf{G}}_k(i, j) \tag{21}$$

where $\mathbf{g}_{(3)k} = \mathbf{g}_{(3)}(k)$, $\tilde{\mathbf{s}}_i = \tilde{\mathbf{S}}(i, :)$, $\tilde{\mathbf{b}}_j = \tilde{\mathbf{B}}(:, j)$. We assume $\mathbf{s}_i = \mathbf{S}(:, i)$ and obtain $\tilde{\mathbf{s}}_i \cdot \mathbf{s}_i = 1$, where $\mathbf{s}_i$ is a spatial component in $\mathbf{S}$, thus, $\tilde{\mathbf{s}}_i$ is a spatial component related vector. Similarly, since $\mathbf{b}_j = \mathbf{B}(:, j)$ and $\tilde{\mathbf{b}}_j^T \cdot \mathbf{b}_j = 1$, $\tilde{\mathbf{b}}_j$ is a temporal component related vector. Therefore, $\mathbf{g}_{(3)k}$ defined in (21) reflects the intensity of subject k under specific spatial (i) and temporal (j) components, and mode-3 fibers of the core tensor consist of subject-specific intensities:

$$\mathbf{c} = [\mathbf{g}_{(3)1}, \mathbf{g}_{(3)2}, \ldots, \mathbf{g}_{(3)K}] \tag{22}$$

2) Spatial and Temporal Features: We exploit the relationship of core tensor $\underline{\mathbf{G}}$ and spatial components $\mathbf{S}$ by rewriting (20) as follows:

$$\underline{\mathbf{G}}_k = \tilde{\mathbf{S}} \cdot \underline{\mathbf{Z}}_k \tag{23}$$

where $\underline{\mathbf{Z}}_k \in \mathbb{R}^{V\times N}$ is given by $\underline{\mathbf{Z}}_k = (\underline{\mathbf{X}}_k - \underline{\mathbf{E}}_k)(\mathbf{B}^T)^{\dagger}$, representing temporal component related information for subject k. For a specific temporal component $\underline{\mathbf{Z}}_k^j = \underline{\mathbf{Z}}_k(:, j)$, the spatial features can be extracted based on (23):

$$\mathbf{g}_{(1)i} = \tilde{\mathbf{s}}_i \cdot \underline{\mathbf{Z}}_k^j \tag{24}$$

where $\mathbf{g}_{(1)i} = \mathbf{g}_{(1)}(i)$, $i = 1, 2, \ldots, N$, so spatial features can be represented by mode-1 fibers of the core tensor. Furthermore, the spatial feature matrix $\mathbf{G}_{\mathbf{S}} \in \mathbb{R}^{N\times K}$ is obtained as follows:

$$\mathbf{G}_{\mathbf{S}}(i, :) = [\tilde{\mathbf{s}}_i \cdot \underline{\mathbf{Z}}_1^j, \tilde{\mathbf{s}}_i \cdot \underline{\mathbf{Z}}_2^j, \ldots, \tilde{\mathbf{s}}_i \cdot \underline{\mathbf{Z}}_K^j] \tag{25}$$

Since $\mathbf{G}_{\mathbf{S}}$ consists of spatial features of all subjects, its straightforward application is to classify multiple subjects within the same group to emphasize the intra-group difference.

When extracting temporal features from the core tensor, we rewrite (20) to highlight temporal components $\mathbf{B}$ as follows:

$$\underline{\mathbf{G}}_k = \underline{\mathbf{M}}_k(\mathbf{B}^T)^{\dagger} \tag{26}$$

where $\underline{\mathbf{M}}_k \in \mathbb{R}^{N\times T}$ is represented by $\underline{\mathbf{M}}_k = (\mathbf{S})^{\dagger}(\underline{\mathbf{X}}_k - \underline{\mathbf{E}}_k)$ and includes spatial component related information for subject k. For a specific spatial component $\underline{\mathbf{M}}_k^i = \underline{\mathbf{M}}_k(i, :)$, temporal features can be represented by mode-2 fibers of the core tensor:

$$\mathbf{g}_{(2)j} = \underline{\mathbf{M}}_k^i \cdot \tilde{\mathbf{b}}_j \tag{27}$$

where $\mathbf{g}_{(2)j} = \mathbf{g}_{(2)}(j)$. Similar to the spatial features defined in (25), we have the temporal features of all subjects as follows:

$$\mathbf{G}_{\mathbf{B}}(j, :) = [\underline{\mathbf{M}}_1^i \cdot \tilde{\mathbf{b}}_j, \underline{\mathbf{M}}_2^i \cdot \tilde{\mathbf{b}}_j, \ldots, \underline{\mathbf{M}}_K^i \cdot \tilde{\mathbf{b}}_j]^T \tag{28}$$

where $\mathbf{G}_{\mathbf{B}} \in \mathbb{R}^{N\times K}$ is the temporal feature matrix. This can also be utilized to cluster multiple subjects into different groups.

III. Experimental Methods

To evaluate the efficiency of the proposed slcTKD model, we carry out experiments on both simulated and experimental fMRI data. We compare the slcTKD algorithm with four related algorithms including RKCA [25], [31], TKD (HOOI) [37], CPD determined using nonlinear least squares [38], and the multilinear rank-$(L, L, 1, 1)$ BTD proposed in [9] for decomposing 4-way multi-subject fMRI data $\underline{\mathbf{X}} \in \mathbb{R}^{I_x\times I_{yz}\times T\times K}$ ($I_x \times I_y \times I_z$ denotes spatial dimensions). By formulating the matrix representation of each shared SM component into a low-rank matrix (of rank L), the multilinear rank-$(L, L, 1, 1)$ BTD gives more flexibility in the spatial domain, compared to 3rd-order CPD.

We compute the multilinear rank-$(L, L, 1, 1)$ BTD as follows: first, we combine the first and second dimensions of the 4-way fMRI data tensor to obtain a 3-way tensor and perform 3rd-order CPD for this tensor. Second, we matricize each column vector of the SM matrix, obtained as the first CPD factor matrix, into a matrix in an inverse manner to how the first and second dimensions of the 4-way fMRI data tensor are combined, and perform rank-L approximation to this matrix to obtain each rank-L component. Finally, we combine the two rank-L components to obtain the shared SM, and have the shared TC and the subject-specific intensity from the two rank-one modes. We determine L via testing all possible choices, and select $L = 6$ (from $2 \sim 30$) for the simulated fMRI data and $L = 10$ (from $2 \sim 45$) for the experimental fMRI data.

Each algorithm is run 10 times for each case. For CPD and BTD, subject-specific intensity is directly obtained. For slcTKD, RKCA, and TKD, we extract subject-specific intensities from the core tensor using (22). The initialization values of λ and γ of RKCA are the same as the proposed slcTKD method for a fair comparison. In the implementation

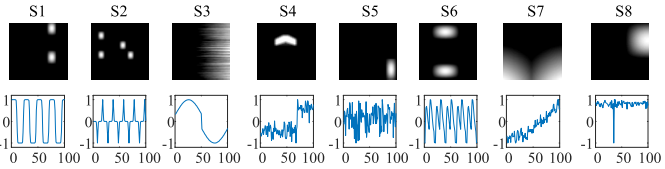


Fig. 2. Eight simulated fMRI sources.

of slcTKD, we use $p = 0.3$, $\gamma = 0.6$, $\lambda = 0.4$ for both simulated and experimental fMRI data. In addition, we select $\delta = 2.5$ for simulated data and $\delta = 0.4$ for experimental data. For other parameters, we set $\eta = 1.3$, the iteration number of Newton-Raphson method $iter_y = 10$, $\xi = 0.4$, and the stop criterion $iter_{\max} = 300$, $\varepsilon_{\min} = 10^{-7}$, $\Delta\varepsilon_{\min} = 10^{-4}$. Note ε^{iter} is calculated using $\varepsilon^{iter} = ||\underline{\mathbf{X}} - \underline{\mathbf{G}} \times_1 \mathbf{S} \times_2 \mathbf{B} - \underline{\mathbf{E}}||_{\mathrm{F}}/||\underline{\mathbf{X}}||_{\mathrm{F}}$ for slcTKD and RKCA, $\varepsilon^{iter} = ||\underline{\mathbf{X}} - \underline{\mathbf{G}} \times_1 \mathbf{S} \times_2 \mathbf{B}||_{\mathrm{F}}/||\underline{\mathbf{X}}||_{\mathrm{F}}$ for TKD and CPD, and $\varepsilon^{iter} = ||\underline{\mathbf{X}} - \sum_{n=1}^{N} (\mathbf{S}_{1_n}\mathbf{S}_{2_n}^T) \circ b_n \circ c_n||_{\mathrm{F}}/||\underline{\mathbf{X}}||_{\mathrm{F}}$ for BTD ($\mathbf{S}_1$ and $\mathbf{S}_2$ are two spatial factors of low-rank L). The same stop criterion is used for all the algorithms, i.e., $iter \geq iter_{\max}$, $\varepsilon^{iter} \leq \varepsilon_{\min}$, or $\Delta\varepsilon^{iter} \leq \Delta\varepsilon_{\min}$.

A. Simulated fMRI Data

We generated 10 simulated fMRI subjects with each subject including eight fMRI-like sources, based on the benchmark data at http://mlsp.umbc.edu/resources.html [39], as shown in Fig. 2. 8 sources include task-related component (S1), transiently task-related components (S2 and S6), and artificial related components (S3, S4, S5, S7 and S8). Each spatial source contains 60×60 voxels with 100-point time courses. Each spatial map is reshaped into a one-dimensional vector (8×3600) and mixed with time courses (100×8). We obtain the real-valued simulated fMRI datasets ($3600 \times 100 \times 10$), each of which has randomly different activated SM voxels, i.e., randomly decrease of the activated SM voxels in the range of 10% [26]. In order to test the noise robustness of the slcTKD algorithm, we add Gaussian noise to the simulated fMRI data with different noise level. The signal-to-noise ratio (SNR) levels range from -10 dB to 10 dB with a 2.5 dB interval. The SNR is defined as $20\log(\sigma_s/\sigma_n)$, where σ_s and σ_n are the temporal standard deviations of the source signal and Gaussian noise, respectively. Since components of interest cannot be well extracted by CPD and TKD when the model order is the same as the true number of components ($N = 8$), we choose a larger model order $N = 20$ for a fair comparison. Moreover, we test the effect of the model order on all algorithms by changing the model order in a range of 20 to 80 with an interval of 10.

B. Experimental fMRI Data

The experimental fMRI dataset was collected from 10 subjects performing a finger-tapping motor task while receiving auditory instructions [16], [26]. All participants signed IRB-approved informed consent at the University of New Mexico. The experimental paradigm is a block design with alternating periods of 30 seconds on (finger tapping) and 30 seconds off (rest). The experiments were performed with a 3T Siemens TIM Trio system with a 12-channel receive coil. The fMRI experiment used a standard Siemens gradient-echo EPI sequence modified to store real and imaginary data separately, and we used the magnitude data in our experiment. The following parameters were used: field-of-view = 24 cm, slice thickness = 3.5 mm, slice gap = 1 mm, number of slices = 32, matrix size = 64×64, TE = 29 ms, TR = 2 s, flip angle = 70 degrees. The preprocessing of the data is carried out using Statistical Parametric Mapping (SPM) software package. The magnitude datasets were co-registered to compensate for movements in the fMRI time series images using INRIAlign. Images were then spatially normalized into the standard Montreal Neurological Institute space. Following spatial normalization, the data (originally acquired at 3.75 mm $\times$ 3.75 mm $\times$ 4.5 mm) were slightly sub-sampled to 3 mm $\times$ 3 mm $\times$ 3 mm, resulting in $53 \times 63 \times 46$ voxels. Then, the images were spatially smoothed with a 10 mm $\times$ 10 mm $\times$ 10 mm full width at half-maximum Gaussian kernel.

After removing the voxels out of the brain and flattening the volume image data of 165 time points for each subject, we construct the three-way actual fMRI data of size $59610 \times 165 \times 10$ (voxel $\times$ time $\times$ subject). The final model order N of each algorithm is set to 50, which is consistent with [16], [26].

C. Performance Measures

For the simulated fMRI data, we choose the task-related component (S1) and two transiently task-related components (S2 and S6) as three components of interest. We calculate (1) the absolute Pearson correlation coefficients $|\rho_c|$ between the SM and TC estimates and the ground truths; (2) the mean of $|\rho_c|$ across 10 runs, denoted as $|\bar{\rho}_c|$, and the standard deviation.

For the experimental fMRI data, we select the task-related component and the default mode network (DMN) as two components of interest for evaluating the proposed method. For the task-related component, we utilize a group general linear model (GLM) map [16], [26] as the SM reference and the model TC as the TC reference. For DMN, we use the DMN component provided by Smith *et al.* [40] as the SM reference, and also use the model TC as the TC reference since DMN has the opposite trend to the task-related component. We calculate (1) the absolute Pearson correlation coefficient $|\rho_c|$ between the shared SM and TC estimates and their references; (2) the mean and standard deviation of $|\rho_c|$ values across 10 runs; (3) the total number of activated voxels as well as the voxels inside the spatial reference.

In addition, we test the performance of the subject-specific intensities as well as the spatial and temporal features. We first compare the subject-specific intensities among all five algorithms and calculate $|\bar{\rho}_c|$ values between each pair of the algorithms. Then, we test the performance of spatial and temporal features via k-means clustering [41]. The clustering results are verified using individual SM and TC references obtained by independent vector analysis [42].

IV. Results

A. Simulated fMRI Data

We first test the noise effect on the proposed method. Fig. 3 shows the shared SMs and TCs estimated at SNR = 10dB

and SNR $= -5$dB under the model order $N = 20$, since the differences of the algorithms are easy to be observed. For each algorithm, we select a run with the closest $|\rho_c|$ value to $|\bar{\rho}_c|$, i.e., the average of all runs, for both shared SMs and TCs. We detrend the TCs by baseline correction [43] and make normalization to the range of -1 to 1. From Fig. 3, we see that slcTKD yields the best performance among the five algorithms for both SMs and TCs; RKCA ranks second on average with much better SM estimates for S1 and S2, followed by BTD, CPD, and TKD.

Fig. 4 shows the means and standard deviations of the $|\rho_c|$ values for the shared SMs and TCs across 10 runs at different noise levels when $N = 20$. We observe similar increasing trends with the increase of SNRs for the five approaches, but $|\bar{\rho}_c|$ values of the proposed slcTKD are the highest for both SMs and TCs from all three components. By contrast, the other four algorithms illustrate varying performance when estimating SMs and TCs for different components and under different SNRs. Taking S1 as example, RKCA performs much better than CPD when estimating SMs under higher SNRs, but it may become worse than CPD when estimating TCs under lower SNRs. The reason for this is that the low-rank constraint helps RKCA to obtain a better SM estimate for a sparse component (e.g., S1) at higher SNR levels, but cannot capture the SM characteristic at lower SNR levels, as a noisy SM is not sparse anymore. As a result, SM estimation degrades and so also do the TC estimates. CPD has no requirement on spatial sparsity; its degradation in SM estimation is slighter than RKCA, thus may provide better SM and TC estimates.

Similarly, we test the model order effects on the proposed algorithm. Fig. 5 shows an exampling comparison of the means and standard deviations of $|\rho_c|$ values across 10 runs for SMs and TCs of the three components estimated by all five algorithms under different model orders for SNR $= 10$dB. We see that the proposed slcTKD algorithm generally achieves the highest $|\bar{\rho}_c|$ values and exhibits the least $|\bar{\rho}_c|$ changes with model order, while the other four algorithms cannot keep consistent performance when examining SMs and TCs for the three components.

B. Experimental fMRI Data

Fig. 6 shows shared SMs ($|Z|>1.5$) and shared TCs for the task-related and DMN components estimated by slcTKD, RKCA, TKD, CPD, and BTD, and their $|\rho_c|$ values with the spatial and temporal references. We show a single run with the closest $|\rho_c|$ value to the $|\bar{\rho}_c|$ value, i.e., the average of all runs, for both shared SMs and TCs. The TCs are baseline corrected [43] and normalized to the range of -1 to 1. The proposed slcTKD yields the best performance among the five algorithms when estimating both SMs and TCs for the two components, in terms of not only quantitative $|\rho_c|$ values but also qualitative observation (more similar to the SM and TC references than the other methods, though the estimated TCs of DMN are not so good as the task-related TC components in general). Unlike the simulated results, BTD and CPD are generally better than RKCA for both components and their SM and TC estimates, in terms of both quantitative $|\rho_c|$ values and qualitative observation. The results reflect that BTD and CPD

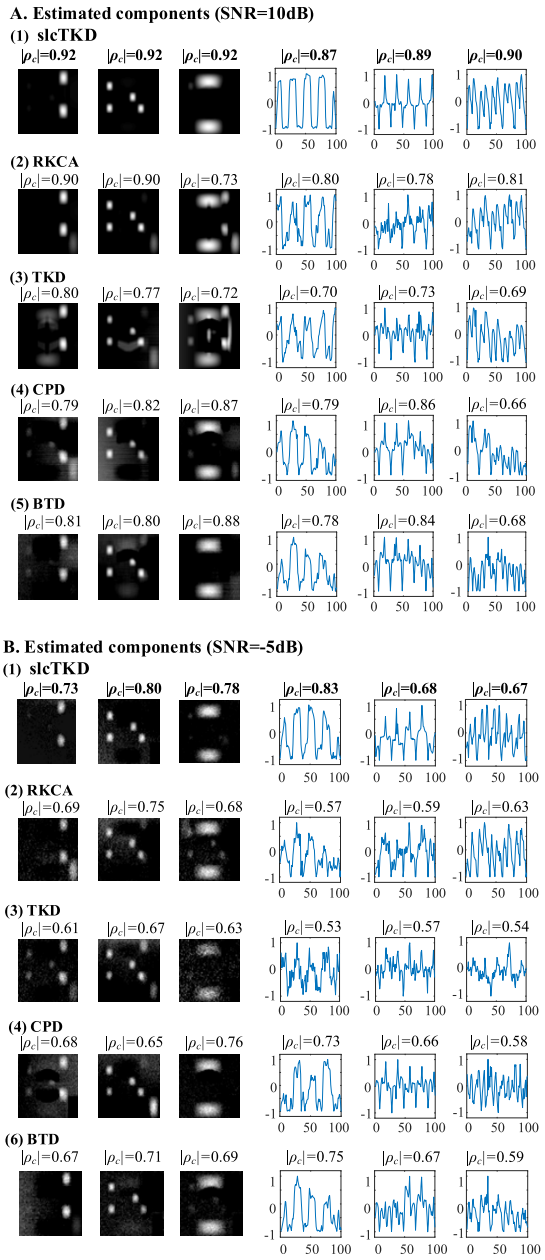


Fig. 3. A comparison of shared SMs and TCs of S1, S2 and S6 estimated at (A) SNR=10dB and (B) SNR=−5dB by (1) slcTKD, (2) RKCA, (3) TKD, (4) CPD, and (5) BTD when $N = 20$. The maximum $|\rho_c|$ values are shown in bold.

are more suitable for modelling experimental fMRI data than RKCA without a spatial sparsity constraint, and the multilinear rank-(L, L, 1, 1) BTD improves SM estimation (e.g., DMN) due to low-rank formulating. This also verifies the lower SNR level of experimental fMRI data and the importance of sparsity constraints on SMs in Tucker-2 model. TKD obtains unsatisfactory results, suggesting its inadequate modeling for highly noisy experimental fMRI data.

Table III has results on the number of activated voxels for the two shared SMs estimated by the five algorithms displayed in Fig. 6. We compare the total number of activated voxel and voxel number inside the spatial reference. Although the proposed slcTKD does not detect the largest number of voxels in total, it does extract the largest number of voxels

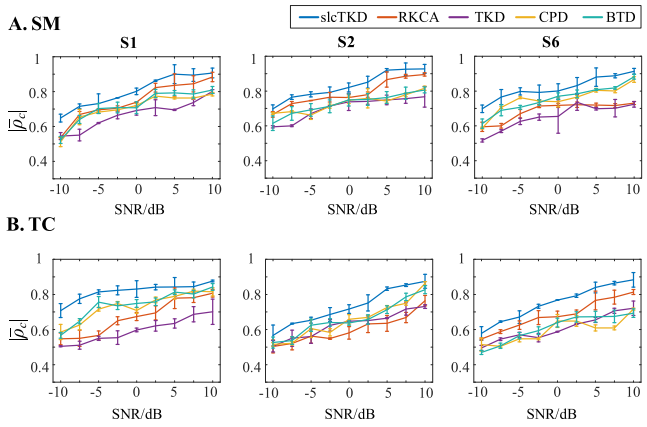


Fig. 4. The comparison of $|\bar{\rho}_C|$ values and standard deviations for (A) shared SMs and (B) shared TCs of S1, S2 and S6 estimated by slcTKD, RKCA, TKD, CPD, and BTD at different noise levels when $N = 20$.

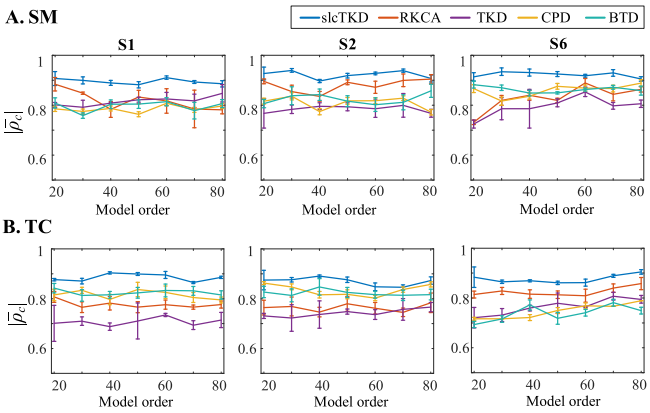


Fig. 5. The comparison of $|\bar{\rho}_C|$ values and standard deviations for (A) shared SMs and (B) shared TCs of S1, S2 and S6 estimated by slcTKD, RKCA, TKD, CPD, and BTD in different model orders when SNR = 10dB.

inside the spatial reference for both the task-related and DMN components. More precisely, slcTKD detects 31.1% ∼ 135.6% more task-related voxels and 9.4% ∼ 49.8% more DMN-related voxels, as compared to the other algorithms, providing larger and more contiguous sub-regions such as the supplementary motor area (SMA) of the task-related component and the posterior cingulate cortex (PCC) and inferior parietal lobule (IPL) for DMN.

Fig. 7 shows the mean and standard deviations of $|\rho_c|$ values for the shared SMs and TCs of the task-related and DMN components estimated by slcTKD, RKCA, TKD, CPD, and BTD, respectively, over all 10 runs for each algorithm. We can see that the proposed slcTKD algorithm achieves higher $|\bar{\rho}_c|$ values with a smaller standard deviation than the other four algorithms for both SMs and TCs of the task-related and DMN components, while TKD ranks the last. BTD and CPD performs better than RKCA for both SM and TC estimates of the two components in terms of both $|\bar{\rho}_c|$ values and standard deviations.

C. Subject-Specific Intensities and Spatial and Temporal Features Extracted From the Core Tensor

Fig. 8 shows results of subject-specific intensities extracted from the core tensor for the task-related and DMN compo-

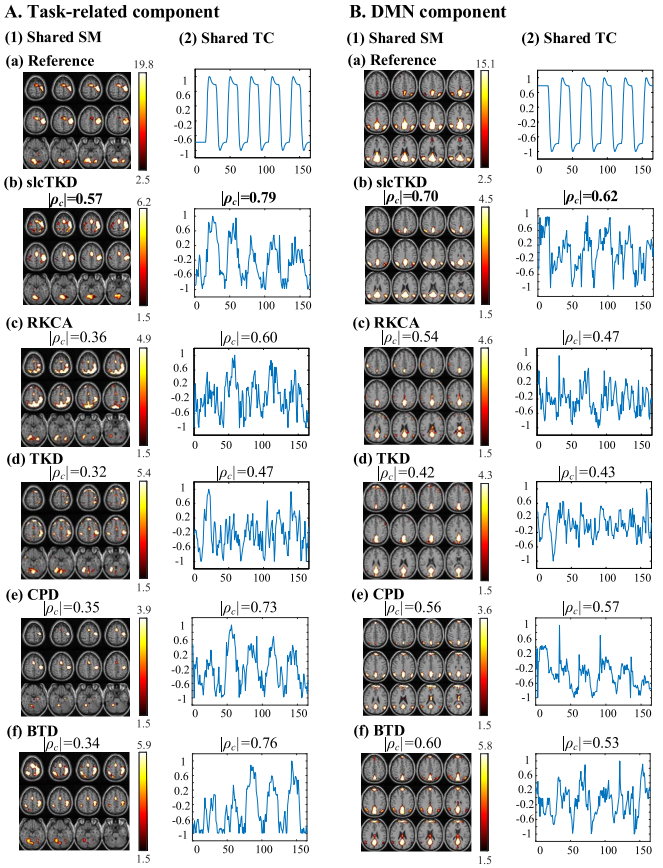


Fig. 6. The comparison of (A) Task-related component and (B) DMN component estimated by slcTKD (b), RKCA (c), TKD (d), CPD (e), BTD (f). The estimates of shared SMs (1) and shared TCs (2) and their references (a) are shown. The maximum $|\rho_C|$ values are also shown in bold.

TABLE III
A COMPARISON OF THE PROPOSED SLCTKD, RKCA, TKD, CPD, AND BTD IN TERMS OF THE TOTAL NUMBER OF ACTIVATED VOXELS AND THE VOXEL NUMBER INSIDE THE SPATIAL REFERENCE FOR THE TWO SHARED SM ESTIMATES DISPLAYED IN FIG. 6. THE MAXIMUM VALUES ARE SHOWN IN BOLD

		slcTKD	RKCA	TKD	CPD	BTD
Task	Total voxels	3376	**5266**	4970	1856	3549
	Inside reference	**2045**	1353	1560	868	1509
DMN	Total voxels	3607	4548	**5124**	2952	4563
	Inside reference	**2534**	2011	1692	1727	2317

nents, which are selected from the same run as that used in Fig. 6. They are normalized to the same range [−1, 1] for comparisons. These subject-specific intensities show mostly similar trends for both task-related component and DMN, especially for those from slcTKD, RKCA and CPD. Fig. 8C includes the $|\rho_c|$ values between the subject-specific intensities extracted by each pair of algorithms. It can be found that the $|\rho_c|$ values between slcTKD and RKCA are the largest (task-related: 0.91; DMN: 0.93) among 10 pairs of algorithms, since slcTKD and RKCA are based on a similar Tucker-2 model. The $|\rho_c|$ values between slcTKD and CPD rank the second

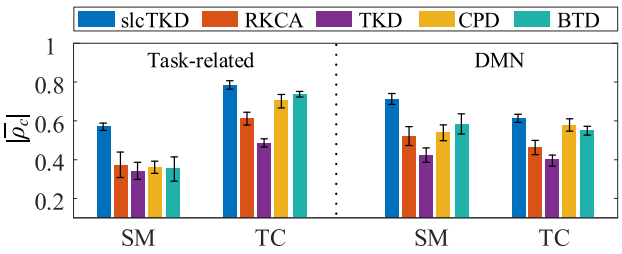


Fig. 7. The comparison of $|\rho_c|$ values and standard deviations for shared SMs and shared TCs of task-related component and DMN component estimated by slcTKD, RKCA, TKD, CPD, and BTD.

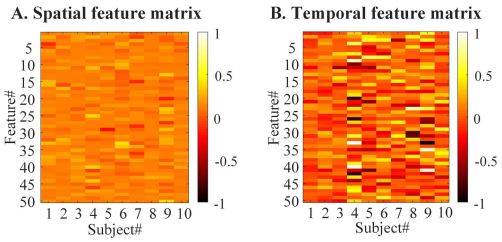
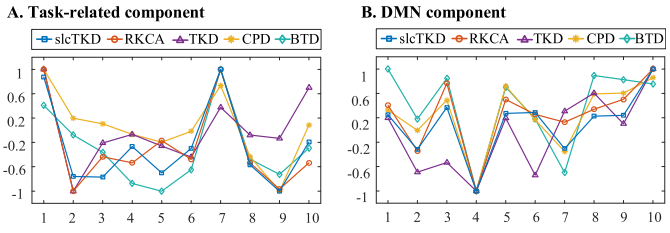


Fig. 9. The spatial (A) and temporal (B) feature matrices extracted from the core tensor for the task-related component from 10 subjects.



C. $|\rho_c|$ values between subject intensities extracted by each pair of algorithms

	slcTKD	RKCA	TKD	CPD	BTD
slcTKD	-	**0.93**	0.74	0.92	0.81
RKCA	**0.91**	-	0.68	0.90	0.78
TKD	0.65	0.69	-	0.64	0.48
CPD	0.83	0.79	0.47	-	0.92
BTD	0.81	0.73	0.45	0.77	-

Fig. 8. The comparison of subject-specific intensities extracted from the core tensor for (A) task-related component and (B) DMN component estimated by slcTKD, RKCA, TKD, CPD, and BTD, and $|\rho_c|$ values between subject-specific intensities extracted by each pair of algorithms (C).

largest, which are 0.83 and 0.92 for the task-related and DMN components, respectively. Note there can be large $|\rho_c|$ values between BTD and CPD (e.g., 0.92 for DMN) since we use CPD to compute the subject-specific intensities for BTD.

Fig. 9 shows spatial and temporal feature matrices extracted from the core tensor for the task-related component from the 10 subjects. Considering that intra-group spatial and temporal differences exist among a group of either patients or healthy subjects, due to individual anatomo-functional differences, individual cognitive strategies and specific subject variability (e.g., sex, age) [44]–[47], we utilize the spatial and temporal feature matrices to cluster multiple subjects into different subgroups. For comparisons, we generate individual SM and TC references by performing a group analysis on the experimental fMRI data using IVA-GL algorithm [42]. We run IVA-GL 10 times and extract the individual SM and TC references from the best run, as shown in Fig. 10. The $|\rho_c|$ values of these SM and TC references with the GLM reference and the model TC are also shown.

We basically cluster the 10 subjects into two groups by k-means clustering [41] based on the spatial and temporal feature matrices of the task-related component, respectively. When clustering using the spatial features, we obtain Group 1 = {subjects #1, #4, #7, #9, #10} and Group 2 = {subjects #2, #3, #5, #6, #8}. We can see that subject #3 and subject #4 are not in the same group though the $|\rho_c|$ values are the same (0.50), as shown in Fig. 10A. In fact, we can observe different task-related activations for subject #3 and subject #4 from Fig. 10A. This difference agrees with the inter-subject spatial variability reported in [44] and [45] and can be well captured by the spatial features involved in the core tensor.

When clustering subjects using the temporal features, we have Group 1 = {subjects #4} and Group 2 = {subjects #1, #2, #3, #5, #6, #7, #8, #9, #10}. This clustering is also reasonable by examining the individual TC references shown in Fig. 10B. The TC of subject #4 is significantly different from the others in that its $|\rho_c|$ value (0.54) is largely lower than those for the others (0.70 $\sim$ 0.91). This result also agrees with the inter-subject temporal variability published in [46] and [47].

We also perform similar clustering based on the spatial features of DMN extracted from the core tensor. Fig. 11 shows the individual SM references separated from the experimental multi-subject fMRI data by the IVA-GL algorithm [42]. As a result, we obtain Group 1 = {subjects #1, #4, #5, #7} and Group 2 = {subjects #2, #3, #6, #8, #9, #10}. We see that subjects #1 & #2 and subjects #7 & #10 are not classified into the same group based on spatial features though their $|\rho_c|$ values are the same (0.55 and 0.58, as shown in Fig. 11). The spatial features extracted from the core tensor can capture spatial differences across subjects, e.g., IPL activations in Group 1.

V. Discussion

This study proposes a constrained Tucker-2 model for analyzing multi-subject fMRI data. We impose the sparsity constraint on SMs to incorporate the intrinsic characteristics of fMRI data, in addition to the low-rank constraint on factor matrices and the sparsity constraint on the core tensor. Compared with the mostly used CPD, the proposed method can provide not only shared SMs, shared TCs, and subject-specific intensities, but also the novel spatial and temporal features capturing spatial and temporal intra-group differences. The results from both simulated and experimental fMRI data verify the advantage of the proposed slcTKD algorithm over BTD, CPD, RKCA, and TKD.

While we can extract shared SMs, shared TCs, and subject-specific intensities agreeing with previous results of CPD-related methods [2], [12], [16], [26], we especially treasure the rich and unique spatial-temporal-subject information involved in the core tensor, which is missing in CPD of multi-subject fMRI data. Previously, researchers have utilized Tucker core

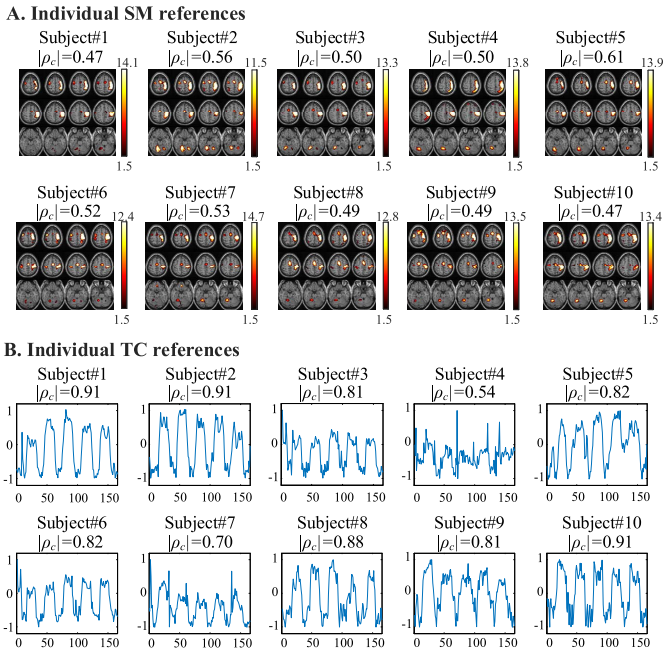


Fig. 10. The individual SM (A) and TC (B) references for the task-related component. The $|\rho_c|$ values of these SM and TC references with the GLM reference and the model TC are also shown.

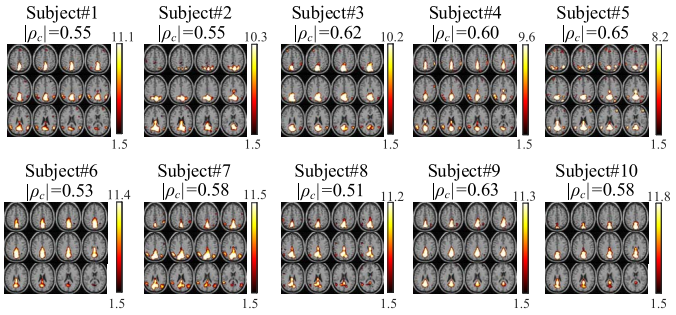


Fig. 11. The individual SM references for the DMN component.

tensor to build connectivity maps based on correlation matrices of multi-subject fMRI data, but they rarely used the spatially and temporally compressed information embedded in the core tensor for individual subjects. We extract novel spatial and temporal features apart from the subject-specific intensities from the core tensor, and apply these spatial and temporal features to make an intra-group clustering. The results show that these features can identify detailed differences in spatial activations or temporal responses across subjects, which extensively existed [44]–[47] but is hard to be evaluated by correlations with the references. Therefore, the proposed method show promise for providing new and reasonable spatial/temporal features to tell intra-group difference, as well as the shared SMs and TCs to provide common spatial/temporal features within a group.

A. Effects of Constraints

The sparsity and low-rank constraints on the spatial factors are both essential for the proposed model to successfully extract principal spatial components, and the spatial

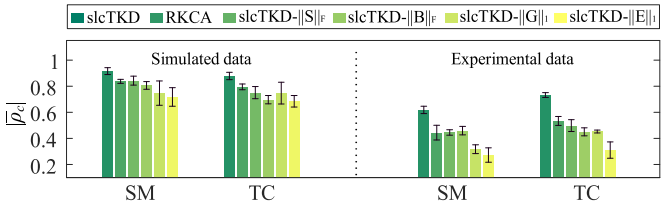


Fig. 12. Comparison of the $|\bar{\rho}_c|$ values and standard deviations of shared SM and TC estimates over 10 runs of slcTKD-$||\mathbf{S}||_F$, slcTKD-$||\mathbf{B}||_F$, slcTKD-$||\mathbf{G}||_1$, and slcTKD-$||\mathbf{E}||_1$, which remove $||\mathbf{S}||_F^2$, $||\mathbf{B}||_F^2$, $||\underline{\mathbf{G}}||_1$, and $||\underline{\mathbf{E}}||_1$ from slcTKD, respectively. The results of slcTKD and RKCA are also shown for comparison.

sparsity constraint is vital for matching the spatial activation characteristics from fMRI data. This can be found by comparing the proposed slcTKD algorithm and RKCA using only a low-rank constraint on the spatial factors. RKCA exhibited appealing properties in image processing such as background subtraction and image denoising [25], [31] but degraded in the fMRI analysis. The main reason is that spatial maps of fMRI are much sparser than optical images. Therefore, the sparsity constraint contributes more to yield better SM estimates in the proposed algorithm. As a result, the proposed slcTKD algorithm removes a larger number of noisy voxels and recovers a larger number of voxels inside the spatial reference with more contiguous activations, as shown in Fig. 6. The high-quality SM estimates would further improve the estimates of spatial-temporal link and the shared TCs, which can be seen from their relationship in (20).

In addition, the low-rankness and sparsity constraints imposed on the factor matrices and the core and residual tensors are also necessary for the proposed slcTKD model to get good performance. The low-rank constraints $||\mathbf{S}||_F^2$ and $||\mathbf{B}||_F^2$ ensure the extraction of principal spatial and temporal components shared by all subjects, whereas sparsity constraints $||\underline{\mathbf{G}}||_1$ and $||\underline{\mathbf{E}}||_1$ improve the uniqueness and performance of the proposed method. To test effects of each constraint on the proposed algorithm, we respectively remove constraints $||\mathbf{S}||_F^2$, $||\mathbf{B}||_F^2$, $||\underline{\mathbf{G}}||_1$ and $||\underline{\mathbf{E}}||_1$ from slcTKD to have four algorithms, named as slcTKD-$||\mathrm{S}||_F$, slcTKD-$||\mathrm{B}||_F$, slcTKD-$||\mathrm{G}||_1$, and slcTKD-$||\mathrm{E}||_1$. We run each algorithm 10 times. Fig. 12 shows the mean and standard deviations of $|\rho_c|$ values for the shared SMs and TCs estimated by each algorithm, with comparison to the results of slcTKD and RKCA. Fig. 13 illustrates examples of shared SMs for DMN selected from a run with the closest $|\rho_c|$ value to $|\bar{\rho}_c|$ shown in Fig. 12 for the experimental fMRI data. We see that slcTKD has much higher $|\bar{\rho}_c|$ values than the other five methods. The two methods removing the sparsity constraints $||\underline{\mathbf{G}}||_1$ and $||\underline{\mathbf{E}}||_1$ from slcTKD result in lower $|\bar{\rho}_c|$ and noisier SMs than those removing the low-rankness constraints $||\mathbf{S}||_F^2$ and $||\mathbf{B}||_F^2$, while RKCA removing the sparsity constraint $||\mathbf{S}||_p$ from slcTKD obtains a noisier SM than slcTKD-$||\mathrm{S}||_F$ and slcTKD-$||\mathrm{B}||_F$. These results verify that each constraint is necessary to the proposed method, and the sparsity constraints are especially essential for decomposing noisy multi-subject fMRI data using a Tucker-2 model.

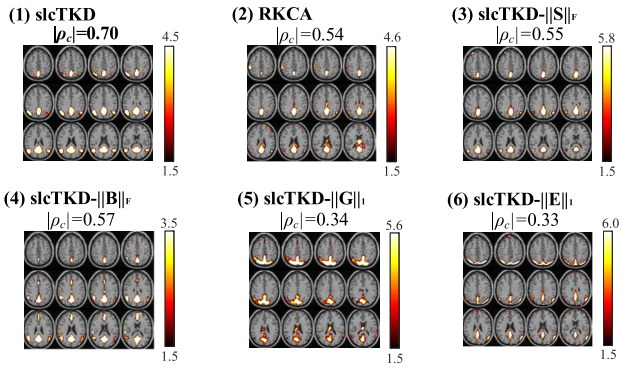


Fig. 13. Comparison of the shared SMs of DMN estimated by (1) slcTKD, (2) RKCA, (3) slcTKD-$||\mathbf{S}||_F$, (4) slcTKD-$||\mathbf{B}||_F$, (5) slcTKD-$||\mathbf{G}||_1$, (6) slcTKD-$||\mathbf{E}||_1$. The maximum $|\rho_c|$ value is shown in bold.

B. Effects of Parameters

The proposed method is mainly affected by parameters from the model constraints (sparsity and low-rankness) and the estimation algorithms (ADMM and HQS).

1) Sparsity Constraint Parameters: The sparsity constraint parameters include δ, p, λ and γ, where δ and p are related to the spatial sparsity, and λ and γ are related to the sparsity of the core and residual tensors. Thus, we classify the four parameters into two groups, i.e., group 1 = {δ and p} and group 2 = {λ and γ}. We test the parameter effects in one group by fixing the parameters in the other by using the same initial values of $\mathbf{B}$, $\mathbf{S}$ and $\underline{\mathbf{G}}$. We first test the parameter effects in group 1. The parameter p is changed from 0.1 to 1 at an interval of 0.1 for both simulated and experimental data. Considering the sparsity difference between two datasets, we change δ from 0.5 to 5 at an interval of 0.5 for the simulated data, while from 0.1 to 1 at an interval of 0.1 for the experimental data. Fig. 14A shows the $|\rho_c|$ values of the task-related SMs with $\lambda = 0.4$ and $\gamma = 0.6$, which are utilized in the experiments. We can see that the $|\rho_c|$ values are generally large and effects are slight. When testing the parameter effects in group 2, we change both λ and γ from 0.1 to 1 at an interval of 0.1 for both simulated and experimental data. The results in Fig. 14B shows that the parameters λ and γ have larger effects on the experimental data than simulated data, which is reasonable because of the different sparsity level of SMs from different subjects.

In summary, we use the same values of p, λ and γ for the simulated and experimental data but different value of δ to match their spatial sparsity difference. The sparser the spatial activations of the dataset are, the larger the δ value is. A larger δ is used for simulated fMRI data than for experimental fMRI data ($\delta = 2.5$ vs. $\delta = 0.4$) because the simulated data is much sparser than experimental fMRI data in spatial activations. Note the same sparsity parameters are effective to extract all the sources for a specific (simulated/experimental) fMRI dataset [48]. Thus, our choices ($p = 0.3$, $\delta = 0.4$, $\gamma = 0.6$, $\lambda = 0.4$) can be a good start to fine tune the four parameters for analyzing other experimental fMRI data.

2) Low-Rank Constraint Parameters: The shared SMs and TCs are extracted in pairs by the proposed method to represent

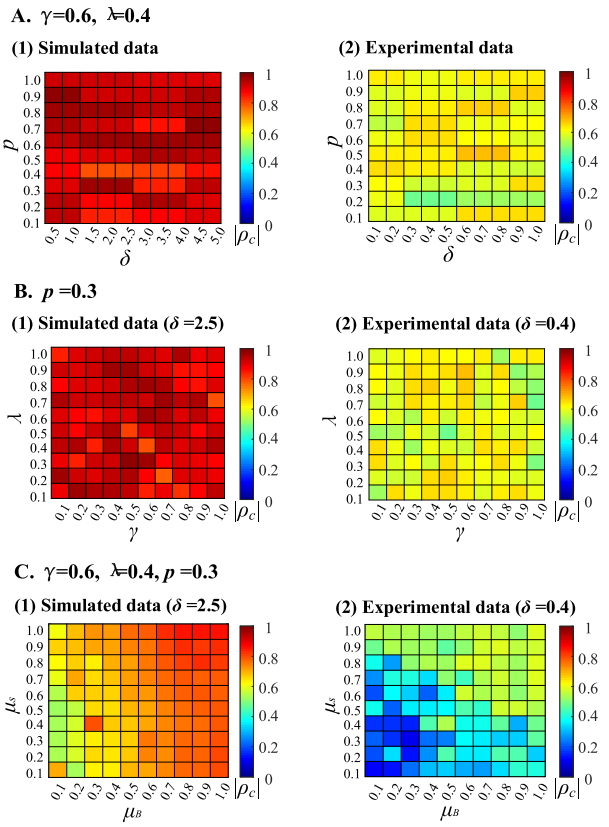


Fig. 14. Parameters δ and p (A), λ and γ (B), and μ_S and μ_B (C) effects on the slcTKD model for simulated (1) and experimental (2) fMRI data.

associated spatial and temporal brain activity. Thus, we select the same low-rankness parameters for the two factor matrices $\mathbf{S}$ and $\mathbf{B}$. We verify the effectiveness of doing this by rewriting the two terms of low-rank constraints in (2) as $\mu_S||\mathbf{S}||_F^2$ and $\mu_B||\mathbf{B}||_F^2$, and by changing μ_S and μ_B from 0.1 to 1 at an interval of 0.1 to test their effects. The $|\rho_c|$ values are shown in Fig. 14C. For both simulated and experimental fMRI data, the larger the parameters μ_S and μ_B are, the larger the $|\rho_c|$ values are. Therefore, $\mu_S = \mu_B = 1$ is a good and simple choice.

3) ADMM and HQS Parameters: The parameters from estimation algorithms consist of η for ADMM in (18) and (19), and ξ for HQS in (7). A previous η setting for RKCA [25], [31] ($\eta = 1.3$) is recommended. Comparing (7) with (2), the role of ξ is similar to that of δ, but its effects on the proposed method is much slighter than that of δ, thus, we select $\xi = 0.4$ for both simulated and experimental fMRI data ($\delta = 0.4$).

Note we have retuned the parameters for the four algorithms obtained by removing each of constraints from the proposed slcTKD, as shown in Fig. 12, which is based on the above-mentioned parameter selection. We do not change the sparsity constraint parameters δ and p representing the spatial sparsity levels of the simulated ($\delta = 2.5$) and experimental ($\delta = 0.4$) data (both: $p = 0.3$), and use the same ADMM and HQS parameters ($\eta = 1.3$ and $\xi = 0.4$) as used by slcTKD. For the sparsity constraint parameters λ and γ, previous choices ($\gamma = 0.6$, $\lambda = 0.4$) also achieve better performance. Finally,

TABLE IV
TIME AND SPACE COMPLEXITY OF MATRICES AND TENSORS PER ITERATION OF THE PROPOSED METHOD. THOSE CAUSED BY $||\mathbf{S}||_p$ ARE SHOWN IN BOLD

	Time complexity	Space complexity
S	$O(K(N^3+TN^2+\boldsymbol{VN}+VT+VTN))$	$O(TN+KN^2+\boldsymbol{VN}+VN+KVT)$
B	$O(K(N^3+TN^2+VT+VTN))$	$O(TN+KN^2+VN+KVT)$
G	$O(KN^2)$	$O(N^2+KN^2)$
E	$O(K(VT+VN^2+VTN))$	$O(TN+KN^2+VN+KVT)$

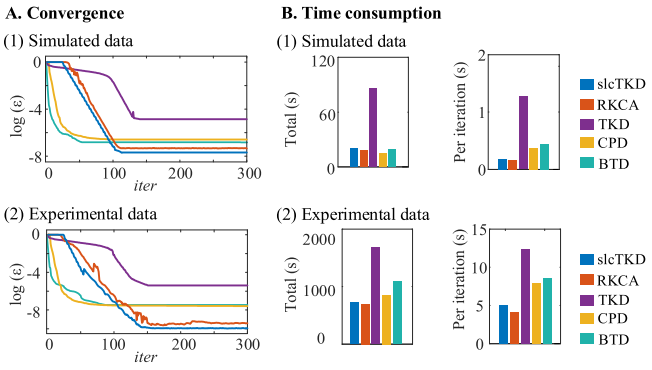


Fig. 15. Convergence (A) and time consumption (B) of slcTKD, RKCA, TKD, CPD, and BTD for simulated data (1) and experimental data (2).

we change to select $\mu_S = 0.1$ for slcTKD-$||\mathrm{S}||_\mathrm{F}$, and $\mu_B = 0.1$ for slcTKD-$||\mathrm{B}||_\mathrm{F}$ to achieve higher $|\bar{\rho}_c|$ values for both shared SMs and TCs from both simulated and experimental data.

C. Computational Complexity and Convergence

Based on the proposed slcTKD model in (2) and its algorithm implementation in Table II, we derive the total time and space complexity per iteration. They are $O(K(N^3 + TN^2 + VN + VT + VN^2 + VTN))$ and $O(N^2 + TN + KN^2 + VN + KVT)$ (sorted from smallest to largest), respectively. Table IV shows details for the two factor matrices and the core and residual tensors. The spatial sparsity constraint $||\mathbf{S}||_p$ causes VN additional time and space complexity per iteration when estimating the shared SMs, resulting in slight increases of time complexity (approximately $1/T$; simulated data: 1%; experimental data: 0.6%) and space complexity (approximately N/KT; simulated data: 2%; experimental data: 3%).

We carry out experiments on Intel 6226R CPU, 256G RAM, windows 10 (64-bit) using MATLAB 2020. Fig. 15 shows the convergence ($iter = 1 \sim 300$) of each algorithm in terms of the residual ε and the time consumption (total and per iteration) for both simulated and experimental fMRI data. We normalize ε to the range of 0 to 1 and show its logarithm value to have a clear visualization. CPD and BTD show faster convergence, followed by the proposed slcTKD, RKCA and TKD. However, slcTKD and RKCA consume similarly smaller running time than TKD, BTD and CPD for the experimental fMRI data, due to smaller time consumption per iteration. The proposed slcTKD algorithm yields the smallest residual among the five algorithms by incorporating the spatial sparsity constraint conforming to the intrinsic characteristic of fMRI data.

D. Future Work

Our future work includes the following. First, we will further improve the Tucker-2 model to represent the multi-subject fMRI data and to extract more information such as individual time delays. Second, we will extend the proposed method to complex-valued fMRI data since the fMRI data is actually complex-valued and recent researches including ours [26] decompose complex-valued fMRI data, yielding promising results [26], [49]. Third, an adaptive sparsity constraint parameter selection method will be explored to discover the links between sparsity parameters and the experimental fMRI data. Finally, we will apply the proposed method to resting-state fMRI data or fMRI data based on regions of interest (ROIs) to analyze patients with neurological disorders.

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, Jun. 2004.
[2] A. Stegeman, "Comparing independent component analysis and the Parafac model for artificial multi-subject fMRI data," Univ. Groningen, Groningen, The Netherlands, Tech. Rep., 2007, vol. 455. [Online]. Available: http://www.alwinstegeman.nl/docs/Stegeman%20-%20parafac%20&%20ica.pdf and https://scholar.google.com/citations?view_op=view_citation&hl=zh-CN&user=7EdP-IUAAAAJ&citation_for_view=7EdP-IUAAAAJ:LkGwnXOMwfcC
[3] A. Cichocki *et al.*, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Mar. 2015.
[4] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
[5] A. Cichocki, "Tensor decompositions: A new concept in brain data analysis?" 2013, *arXiv:1305.0395*. [Online]. Available: http://arxiv.org/abs/1305.0395
[6] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms—Part II: Definitions and uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1033–1066, Jan. 2008.
[7] L. Sorber, M. V. Barel, and L. D. Lathauwer, "Structured data fusion," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 4, pp. 586–600, Jun. 2015.
[8] A. A. Rontogiannis, E. Kofidis, and P. V. Giampouras, "Block-term tensor decomposition: Model selection and computation," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 3, pp. 464–475, Apr. 2021.
[9] C. Chatzichristos, E. Kofidis, M. Morante, and S. Theodoridis, "Blind fMRI source unmixing via higher-order tensor decompositions," *J. Neurosci. Methods*, vol. 315, pp. 17–47, Mar. 2019.
[10] N. Leonardi and D. Van De Ville, "Identifying network correlates of brain states using tensor decompositions of whole-brain dynamic functional connectivity," in *Proc. Int. Workshop Pattern Recognit. Neuroimaging*, Philadelphia, PA, USA, Jun. 2013, pp. 74–77.
[11] E. Al-Sharoa, M. Al-Khassaweneh, and S. Aviyente, "Tensor based temporal and multilayer community detection for studying brain dynamics during resting state fMRI," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 3, pp. 695–709, Mar. 2019.
[12] M. Mørup, L. K. Hansen, S. M. Arnfred, L.-H. Lim, and K. H. Madsen, "Shift-invariant multilinear decomposition of neuroimaging data," *NeuroImage*, vol. 42, no. 4, pp. 1439–1450, Oct. 2008.
[13] G. Zhou *et al.*, "Linked component analysis from matrices to high-order tensors: Applications to biomedical data," *Proc. IEEE*, vol. 104, no. 2, pp. 310–331, Feb. 2016.
[14] S. M. Hamdi, Y. Wu, S. F. Boubrahimi, R. Angryk, L. C. Krishnamurthy, and R. Morris, "Tensor decomposition for neurodevelopmental disorder prediction," in *Proc. Int. Conf. Brain Informat. (BI)*, Arlington, TX, USA, Dec. 2018, pp. 339–348.

[15] B. Hunyadi, P. Dupont, W. Van Paesschen, and S. Van Huffel, "Tensor decompositions and data fusion in epileptic electroencephalography and functional magnetic resonance imaging data," *Wiley Interdiscipl. Rev.-Data Mining Knowl. Discovery*, vol. 7, no. 1, p. e1197, 2017.

[16] L. D. Kuang, Q.-H. Lin, X.-F. Gong, F. Cong, J. Sui, and V. D. Calhoun, "Multi-subject fMRI analysis via combined independent component analysis and shift-invariant canonical polyadic decomposition," *J. Neurosci. Methods*, vol. 256, pp. 127–140, Dec. 2015.

[17] F. Cong, Q.-H. Lin, L.-D. Kuang, X.-F. Gong, P. Astikainen, and T. Ristaniemi, "Tensor decomposition of EEG signals: A brief review," *J. Neurosci. Methods*, vol. 248, pp. 59–69, Jun. 2015.

[18] R. Mayhugh, C. Hugenschmidt, J. Rejeski, P. Laurienti, and F. Mokhtari, "Tensor-based vs. matrix-based rank reduction in dynamic brain connectivity," in *Proc. SPIE Conf. Med. Imag.-Image Process. (SPIE-MI)*, Houston, TX, USA, vol. 10574, Feb. 2018, pp. 1–10.

[19] A. Ozdemir, M. Villafane-Delgado, D. C. Zhu, M. A. Iwen, and S. Aviyente, "Multi-scale higher order singular value decomposition (MS-HoSVD) for resting-state fMRI compression and analysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 6299–6303.

[20] C. M. Crainiceanu, B. S. Caffo, S. Luo, V. M. Zipunnikov, and N. M. Punjabi, "Population value decomposition, a framework for the analysis of image populations," *J. Amer. Statist. Assoc.*, vol. 106, no. 495, pp. 775–790, 2011.

[21] G. Zhou and A. Cichocki, "Fast and unique Tucker decompositions via multiway blind source separation," *Bull. Polish Acad. Sci., Tech. Sci.*, vol. 60, pp. 389–405, Jan. 2012.

[22] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2016, pp. 5249–5257.

[23] C. Grussler, A. Rantzer, and P. Giselsson, "Low-rank optimization with convex constraints," *IEEE Trans. Autom. Control*, vol. 63, no. 11, pp. 4000–4007, Nov. 2018.

[24] A. Traoré, M. Berar, and A. Rakotomamonjy, "Online multimodal dictionary learning," *Neurocomputing*, vol. 368, pp. 163–179, Nov. 2019.

[25] M. Bahri, Y. Panagakis, and S. Zafeiriou, "Robust Kronecker component analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 10, pp. 2365–2379, Oct. 2019.

[26] L.-D. Kuang, Q.-H. Lin, X.-F. Gong, F. Cong, Y.-P. Wang, and V. D. Calhoun, "Shift-invariant canonical polyadic decomposition of complex-valued multi-subject fMRI data with a phase sparsity constraint," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 844–853, Apr. 2020.

[27] J. Xu *et al.*, "Large-scale functional network overlap is a general property of brain functional organization: Reconciling inconsistent fMRI findings from general-linear-model-based analyses," *Neurosci. Biobehav. Rev.*, vol. 71, pp. 83–100, Dec. 2016.

[28] P. Lennie, "The cost of cortical computation," *Current Biol.*, vol. 13, pp. 654–658, Mar. 2003.

[29] I. Daubechies *et al.*, "Independent component analysis for brain fMRI does not select for independence," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 26, pp. 10415–10422, 2009.

[30] N. Hurley and S. Rickard, "Comparing measures of sparsity," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4723–4741, Oct. 2009.

[31] M. Bahri, Y. Panagakis, and S. Zafeiriou, "Robust kronecker-decomposable component analysis for low-rank modeling," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, vol. 3352, no. 1, pp. 3352–3361.

[32] Q. Xie, Q. Zhao, D. Meng, and Z. Xu, "Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 8, pp. 1888–1902, Aug. 2018.

[33] L. El Alaoui, A. Ern, and M. Vohralík, "Guaranteed and robust a posteriori error estimates and balancing discretization and linearization errors for monotone nonlinear problems," *Comput. Methods Appl. Mech. Eng.*, vol. 200, pp. 2782–2795, Sep. 2011.

[34] S. J. Hammarling, "Numerical solution of the stable, non-negative definite Lyapunov equation," *IMA J. Numer. Anal.*, vol. 2, no. 3, pp. 303–323, Jul. 1982.

[35] G. H. Golub, S. Nash, and C. Van Loan, "A Hessenberg-Schur method for the problem AX+XB=C," *IEEE Trans. Autom. Control*, vol. AC-24, no. 6, pp. 909–913, Dec. 1979.

[36] E. Sanchez and B. R. Kowalski, "Tensorial resolution: A direct trilinear decomposition," *J. Chemometrics*, vol. 4, no. 1, pp. 29–49, 1990.

[37] P. A. Regalia, "Monotonically convergent algorithms for symmetric tensor approximation," *Linear Algebra Appl.*, vol. 438, no. 2, pp. 875–890, Jan. 2013.

[38] L. Sorber, M. Van Barel, and L. De Lathauwer, "Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-(L_r, L_r, 1) terms, and a new generalization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 695–720, 2013.

[39] N. Correa, T. Adali, Y.-O. Li, and V. D. Calhoun, "Comparison of blind source separation algorithms for fMRI using a new MATLAB toolbox: GIFT," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Philadelphia, PA, USA, Mar. 2005, pp. 401–404.

[40] S. M. Smith *et al.*, "Correspondence of the brain's functional architecture during activation and rest," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 31, pp. 13040–13045, 2009.

[41] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient K-means clustering algorithm: Analysis and implementation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 881–892, Jul. 2002.

[42] M. Anderson, T. Adalı, and X.-L. Li, "Joint blind source separation with multivariate Gaussian model: Algorithms and performance analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1672–1683, Apr. 2012.

[43] F. Gan, G. Ruan, and J. Mo, "Baseline correction by improved iterative polynomial fitting with automatic threshold," *Chemometric Intell. Lab. Syst.*, vol. 82, nos. 1–2, pp. 59–65, May 2006.

[44] M. L. Seghier *et al.*, "Variability of fMRI activation during a phonological and semantic language task in healthy subjects," *Hum. Brain Mapping*, vol. 23, no. 3, pp. 140–155, 2004.

[45] B. B. Biswal *et al.*, "Toward discovery science of human brain function," *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 10, pp. 4734–4739, 2010.

[46] D. A. Handwerker, J. M. Ollinger, and M. D'Esposito, "Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses," *NeuroImage*, vol. 21, no. 4, pp. 1639–1651, Apr. 2004.

[47] S. G. Vij, J. S. Nomi, D. R. Dajani, and L. Q. Uddin, "Evolution of spatial and temporal features of functional brain networks across the lifespan," *NeuroImage*, vol. 173, pp. 498–508, Jun. 2018.

[48] J. Lv *et al.*, "Task fMRI data analysis based on supervised stochastic coordinate coding," *Med. Image Anal.*, vol. 38, pp. 1–16, May 2017.

[49] A. Iqbal, M. Nait-Meziane, A. K. Seghouane, and K. Abed-Meraim, "Adaptive complex-valued dictionary learning: Application to fMRI data analysis," *Signal Process.*, vol. 166, pp. 1–14, Jan. 2020.