

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Kuang, Li-Dan; Lin, Qiu-Hua; Gong, Xiao-Feng; Cong, Fengyu; Wang, Yu-Ping; Calhoun, Vince D.

**Title:** Shift-Invariant Canonical Polyadic Decomposition of Complex-Valued Multi-Subject fMRI Data with a Phase Sparsity Constraint

**Year:** 2020

**Version:** Accepted version (Final draft)

**Copyright:** © 2020 The Authors

**Rights:** CC BY 4.0

**Rights url:** <http://rightsstatements.org/page/InC/1.0/?language=en>; <https://creativecommons.org/licenses/by/4.0/>

**Please cite the original version:**

Kuang, L.-D., Lin, Q.-H., Gong, X.-F., Cong, F., Wang, Y.-P., & Calhoun, V. D. (2020). Shift-Invariant Canonical Polyadic Decomposition of Complex-Valued Multi-Subject fMRI Data with a Phase Sparsity Constraint. *IEEE Transactions on Medical Imaging*, 39(4), 844-853.  
<https://doi.org/10.1109/TMI.2019.2936046>

# Shift-Invariant Canonical Polyadic Decomposition of Complex-Valued Multi-Subject fMRI Data With a Phase Sparsity Constraint

Li-Dan Kuang, Qiu-Hua Lin<sup>1</sup>, Member, IEEE, Xiao-Feng Gong<sup>2</sup>, Member, IEEE, Fengyu Cong, Senior Member, IEEE, Yu-Ping Wang<sup>3</sup>, Senior Member, IEEE, and Vince D. Calhoun<sup>4</sup>, Fellow, IEEE

**Abstract**—Canonical polyadic decomposition (CPD) of multi-subject complex-valued fMRI data can be used to provide spatially and temporally shared components among groups with both magnitude and phase information. However, the CPD model is not well formulated due to the large subject variability in the spatial and temporal modalities, as well as the high noise level in complex-valued fMRI data. Considering that the shift-invariant CPD can model temporal variability across subjects, we propose to further impose a phase sparsity constraint on the shared spatial maps to denoise the complex-valued components and to model the inter-subject spatial variability as well. More precisely, subject-specific time delays are first estimated for the complex-valued shared time courses in the framework of real-valued shift-invariant CPD. Source phase sparsity is then imposed on the complex-valued

shared spatial maps. A smoothed  $\ell_0$  norm is specifically used to reduce voxels with large phase values after phase de-ambiguity based on the small phase characteristic of BOLD-related voxels. The results from both the simulated and experimental fMRI data demonstrate improvements of the proposed method over three complex-valued algorithms, namely, tensor-based spatial ICA, shift-invariant CPD and CPD without spatiotemporal constraints. When comparing with a real-valued algorithm combining shift-invariant CPD and ICA, the proposed method detects 178.7% more contiguous task-related activations.

**Index Terms**—Canonical polyadic decomposition (CPD), complex-valued fMRI data, source phase sparsity, shift-invariant, spatiotemporal constraints.

## I. INTRODUCTION

TENSOR decomposition applied to multi-subject functional magnetic resonance imaging (fMRI) data is of growing interest, as such approaches enable us to make use of the multiway structure of fMRI data in terms of space, time and subject/trial [1]–[3]. Canonical polyadic decomposition (CPD), also called CANDECOMP or parallel factor analysis (PARAFAC), is a common method for tensor decomposition [4], [5] because it can provide a unique decomposition under mild conditions [6], [7]. CPD has been widely applied to multi-trial/subject fMRI data since 2004 to extract shared spatial maps (SMs), shared time courses (TCs), and trial/subject-specific intensities [1], [2], [8]–[11]. CPD is a complementary method to independent vector analysis (IVA), which emphasizes the capture of inter-subject variability and the decomposition of multi-subject fMRI data into individual SMs and TCs [12].

CPD was first applied to the analysis of fMRI data in [1]. The multi-trial fMRI data from a single subject generally accords with the CPD model that decomposes a  $K$ -way tensor into a linear combination of a series of rank-one tensors, and each rank-one tensor is the out product of  $K$  loading vectors. By contrast, multi-subject fMRI data tend to violate the CPD model since the data involve larger inter-subject spatial and temporal variability than multi-trial fMRI data. For this reason, some constraints on spatial and temporal modalities were imposed to make the CPD model more rigid. Most

Manuscript received June 19, 2019; revised August 2, 2019 and August 5, 2019; accepted August 12, 2019. Date of publication August 19, 2019; date of current version April 1, 2020. 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 NSF under Grant 1539067, Grant 0840895, Grant 1539067, and Grant 0715022, in part by the NIH Grant R01MH104680, Grant R01MH107354, Grant R01EB005846, and Grant 5P20GM103472, in part by the Fundamental Research Funds for the Central Universities, China, under Grant DUT14RC(3)037, and in part by the Supercomputing Center of Dalian University of Technology. (Corresponding author: Qiu-Hua Lin.)

L.-D. Kuang is with the School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China, and also with the School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha 410114, China (e-mail: kuangld@csust.edu.cn).

Q.-H. Lin and X.-F. Gong are with the School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China (e-mail: qhlin@dlut.edu.cn; xfgong@dlut.edu.cn).

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

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

V. 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 USA (e-mail: vcalhoun@gsu.edu).

This article has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the author.

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2019.2936046

studies have focused on imposing spatial constraints on CPD. Beckmann and Smith proposed a tensor-based spatial independent component analysis (ICA) (T-sICA) algorithm, which imposed spatial independence on the SMs of multi-subject fMRI data via ICA, followed by a rank-one approximate of CPD to obtain shared TCs and subject-specific intensities [2]. Zhou and Cichocki proposed a novel CPD algorithm incorporating *a priori* knowledge about the components in one mode, thus the spatial independence can be readily constrained by extracting a factor matrix using ICA [10]. De Vos *et al.* also proposed a CPD algorithm where independence is imposed on the factors in one mode [13]. This algorithm may require heavy computation in the analysis of multi-subject fMRI data due to working with the covariance matrix and fourth-order cumulant tensor.

In addition to the spatial variability, time shifts occur naturally in multi-subject fMRI data and even in multi-trial fMRI data from an individual, due to the hemodynamic delay. As such, Mørup *et al.* extended the CPD model to incorporate time shifts, called shift-invariant CPD, and then applied the algorithm to detect variable latencies in multi-trial fMRI data from one single subject [8]. The shift-invariant CPD employs a frequency-domain method to estimate time shifts since the exhaustive searching strategy is time-consuming in the time domain when many time delays are involved [14]. The frequency-domain method estimates more accurate time delays and uses less computation time than the time-domain algorithm [8]. Under the same experimental conditions, the inter-subject temporal variability from multi-subject fMRI data tends to be larger than the inter-trial temporal variability from single-subject fMRI data, due to different response times or hemodynamic delays for different subjects. As such, we leveraged both inter-subject spatial and temporal variability and proposed an improved method combining ICA and shift-invariant CPD (named ICA-sCPD) for analyzing multi-subject fMRI data. The results showed that our method outperformed T-sICA and shift-invariant CPD [11].

To our best knowledge, CPD has not yet been applied to complex-valued multi-subject fMRI data. Motivated by the benefits of using both magnitude and phase information in fMRI data [15]–[20], we propose to exploit the phase in a CPD algorithm to take advantage of the full brain information contained in complex-valued fMRI data. Different from spatial independence, we propose the incorporation of a spatial sparsity constraint on CPD. The consideration is that the spatial sparsity is in line with the intrinsic characteristic of brain activation [21]–[23], and sparse representation utilizing the SM sparsity exhibited promising results in the analysis of magnitude-only fMRI data [24]–[29]. Because magnitudes of complex-valued SM components are highly noisy, we seek to utilize the small phase characteristic of blood oxygenation-level dependent (BOLD)-related voxels. Specifically, BOLD-related voxels show small phase changes within the range of  $[-4/\pi, 4/\pi]$ , while unwanted voxels tend to have large phase changes relative to a baseline phase image  $([-\pi, -4/\pi])$  and  $(4/\pi, \pi])$  [15]. We call the phase map of the complex-valued SM component the spatial source phase, to distinguish it from the observed phase data. In addition,

time shifts are still used to incorporate a temporal constraint. Simulated fMRI data with different SM and TC changes and noise levels as well as experimental fMRI data are used to examine the performance of our proposed method.

To summarize, our contributions are as follows:

(1) We propose to exploit the spatial source phase in the CPD algorithm with spatiotemporal constraints for decomposing complex-valued multi-subject fMRI data into shared SMs, shared TCs, and subject-specific time delays and intensities.

(2) We propose to use the sparsity of the spatial source phase to constrain the complex-valued shared SM in the CPD algorithm. After performing phase de-ambiguity for complex-valued SM estimates, the voxels with larger phase values are reduced based on the smoothed  $\ell_0$  norm [31].

(3) We derive the update rule to estimate subject-specific time delays for complex-valued shared time courses in the framework of a real-valued shift-invariant CPD algorithm [8].

The remainder of this paper is organized as follows. Section II describes the details of our proposed method. Section III introduces the simulated and experimental fMRI dataset as well as the performance indices for algorithm evaluation. Section IV gives the experimental results to exhibit the advantage of our proposed method. A discussion is included in Section V.

## II. METHODS

In this section, we first present the complex-valued, shift-invariant CPD algorithm (named csCPD). We then impose a phase sparsity constraint on the shared SMs of the csCPD (named pcsCPD) to form our proposed method. Finally, we present a detailed implementation of the proposed pcsCPD algorithm.

*Notations:* The scalars, vectors, matrices and tensors are denoted by italic lower-case letters (e.g.,  $a$ ), bold lower-case letters (e.g.,  $\mathbf{a}$ ), bold upper-case letters (e.g.,  $\mathbf{A}$ ), and underlined bold capital letters (e.g.,  $\underline{\mathbf{A}}$ ), respectively. The  $n$ -mode matricizing of the tensor is  $\mathbf{A}_{(n)}$ . Superscripts “ $T$ ”, “ $*$ ”, “ $H$ ” and “ $\dagger$ ” denote the transpose, conjugate, conjugate transpose, and pseudo-inverse, respectively. “ $\|\cdot\|$ ” is a  $\ell_2$  norm function. “ $\text{Re}(\cdot)$ ”, “ $\text{Im}(\cdot)$ ”, “ $|\cdot|$ ” and “ $\theta(\cdot)$ ” denote the real, imaginary, magnitude, and phase parts of a complex-valued variable, respectively.

### A. Complex-Valued Shift-Invariant CPD Algorithm

1) *The Model:* Let the three-way (voxel  $\times$  time  $\times$  subject) multi-subject fMRI data be  $\underline{\mathbf{X}} = \{x_{v,j,k}\} \in \mathbb{C}^{V \times J \times K}$ , where  $V$  is the number of in-brain voxels,  $J$  is the number of time points for each subject,  $K$  is the number of subjects,  $v = 1, \dots, V$ ,  $j = 1, \dots, J$ , and  $k = 1, \dots, K$ . Assume  $N$  is the number of components, and  $\mathbf{S} \triangleq \{\mathbf{s}_1, \dots, \mathbf{s}_N\} = \{s_{v,n}\} \in \mathbb{C}^{V \times N}$ ,  $\mathbf{B} \triangleq \{\mathbf{b}_1, \dots, \mathbf{b}_N\} = \{b_{j,n}\} \in \mathbb{C}^{J \times N}$  and  $\mathbf{C} \triangleq \{\mathbf{c}_1, \dots, \mathbf{c}_N\} = \{c_{k,n}\} \in \mathbb{C}^{K \times N}$  ( $n = 1, \dots, N$ ) correspond to shared SMs, shared TCs, and subject-specific intensities, respectively. The model for csCPD is as follows [8], [11]:

$$x_{v,j,k} = \sum_{n=1}^N s_{v,n} b_n(j - \tau_{k,n}) c_{k,n} + e_{v,j,k}, \quad (1)$$

where  $\mathcal{T} = \{\tau_{k,n}\} \in \mathbb{R}^{K \times N}$  is the time delays for  $N$  components of  $K$  subjects;  $b_n(j - \tau_{k,n})$  denotes  $b_{j,n}$  with the

time delay  $\tau_{k,n}$ , and  $\tau_{k,n}$  is the time delay for component  $n$  of subject  $k$ . We, thus, have the shared TCs with time delays for subject  $k$   $\mathbf{B}^{(k)} \triangleq [\mathbf{b}_1^{(k)}, \dots, \mathbf{b}_N^{(k)}] = \{b_n(j - \tau_{k,n})\} \in \mathbb{C}^{J \times N}$ . If  $\tau_{k,n} > 0$ ,  $\mathbf{b}_n^{(k)} = [b_n(1 - \tau_{k,n}), \dots, b_n(J - \tau_{k,n})]^T$  is obtained by cyclic left shifting  $\mathbf{b}_n$  with  $\tau_{k,n}$  points, otherwise it is obtained by cyclic right shifting  $\mathbf{b}_n$  with  $\tau_{k,n}$  points.  $\underline{\mathbf{E}} = \{e_{v,j,k}\} \in \mathbb{C}^{V \times J \times K}$  is the residual reflecting the inter-subject variability effect.

Performing a discrete Fourier transform on (1), we have the model for the csCPD algorithm in the frequency domain, as follows:

$$\tilde{x}_{v,f,k} = \sum_{n=1}^N s_{v,n} \tilde{b}_{f,n} c_{k,n} \exp\left\{-i2\pi \frac{f-1}{J} \tau_{k,n}\right\} + \tilde{e}_{v,f,k}, \quad (2)$$

where  $f = 1, \dots, F$  is the frequency point, and here, we let  $F = J$ ;  $\tilde{\mathbf{X}} = \{\tilde{x}_{v,f,k}\} \in \mathbb{C}^{V \times F \times K}$ ,  $\tilde{\mathbf{B}} \triangleq [\tilde{\mathbf{b}}_1, \dots, \tilde{\mathbf{b}}_N] = \{\tilde{b}_{f,n}\} \in \mathbb{C}^{F \times N}$  and  $\tilde{\mathbf{E}} = \{\tilde{e}_{v,f,k}\} \in \mathbb{C}^{V \times F \times K}$  are the frequency-domain forms of  $\mathbf{X}$ ,  $\mathbf{B}$  and  $\underline{\mathbf{E}}$ , respectively;  $i = \sqrt{-1}$ ;  $\exp\{\cdot\}$  denotes the exponential function, and  $\tilde{b}_{f,n} \exp\{-i2\pi(f-1)\tau_{k,n}/J\}$  is the frequency form of  $b_n(j - \tau_{k,n})$ .

In the csCPD algorithm, the loading matrices  $\mathbf{S}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  and  $\mathcal{T}$  are unknown. Thus, we randomly initialize  $\mathbf{S}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$ , and  $\mathcal{T}$ , and then update these four loading matrices in each iteration until convergence or reaching the number of maximum iterations. We first present update rules for  $\mathbf{S}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$ , and then derive the update rule for  $\mathcal{T}$  in the framework of the real-valued shift-invariant CPD method [8].

**2) Updates of  $\mathbf{S}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$ :** Assume vectors  $\mathbf{s}_v$ ,  $\mathbf{x}_{(1)v}$  and  $\mathbf{e}_{(1)v}$  are constructed by the  $v$ th row of  $\mathbf{S}$ ,  $\mathbf{X}_{(1)} \in \mathbb{C}^{V \times J \times K}$ , and  $\mathbf{E}_{(1)} \in \mathbb{C}^{V \times J \times K}$ ;  $\tilde{\mathbf{b}}_f$ ,  $\tilde{\mathbf{x}}_{(2)f}$  and  $\tilde{\mathbf{e}}_{(2)f}$  by the  $f$ th row of  $\tilde{\mathbf{B}}$ ,  $\tilde{\mathbf{X}}_{(2)} \in \mathbb{C}^{F \times V \times K}$ , and  $\tilde{\mathbf{E}}_{(2)} \in \mathbb{C}^{F \times V \times K}$ ; and  $\mathbf{c}_k$ ,  $\mathbf{x}_{(3)k}$  and  $\mathbf{e}_{(3)k}$  by the  $k$ th row of  $\mathbf{C}$ ,  $\mathbf{X}_{(3)} \in \mathbb{C}^{K \times J \times V}$ , and  $\mathbf{E}_{(3)} \in \mathbb{C}^{K \times J \times V}$ , respectively. Based on (1) and (2), the csCPD model can be further rewritten as follows:

$$\mathbf{x}_{(1)v} = \mathbf{Z}\mathbf{s}_v + \mathbf{e}_{(1)v}, \quad (3)$$

$$\tilde{\mathbf{x}}_{(2)f} = (\tilde{\mathbf{C}}^{(f)} \odot \mathbf{S})\tilde{\mathbf{b}}_f + \tilde{\mathbf{e}}_{(2)f}, \quad (4)$$

$$\mathbf{x}_{(3)k} = (\mathbf{B}^{(k)} \odot \mathbf{S})\mathbf{c}_k + \mathbf{e}_{(3)k}, \quad (5)$$

where  $\mathbf{Z} \in \mathbb{C}^{J \times K \times N}$  is the aggregating mixing matrix and its element is  $z_{j+k(J-1)n} = c_{k,n} b_n(j - \tau_{k,n})$ , i.e., the Khatri-Rao product between  $\mathbf{C}$  and  $\mathbf{B}^{(k)}$  ( $k = 1, \dots, K$ );  $\tilde{\mathbf{C}}^{(f)} = \mathbf{C} \cdot \exp\{-i2\pi(f-1)\mathcal{T}/J\}$ , and  $\cdot$  is the dot product. Thus, the alternative least squares (ALS) updates of  $\mathbf{S}$ ,  $\mathbf{B}$  and  $\mathbf{C}$  can be derived from (3) to (5) [8] as follows:

$$\mathbf{s}_v \leftarrow \mathbf{Z}^\dagger \mathbf{x}_{(1)v}, \quad (6)$$

$$\tilde{\mathbf{b}}_f \leftarrow (\tilde{\mathbf{C}}^{(f)} \odot \mathbf{S})^\dagger \tilde{\mathbf{x}}_{(2)f}, \quad (7)$$

$$\mathbf{c}_k \leftarrow (\mathbf{B}^{(k)} \odot \mathbf{S})^\dagger \mathbf{x}_{(3)k}, \quad (8)$$

where  $\mathbf{S}$  and  $\mathbf{C}$  are updated in the time domain, while  $\mathbf{B}$  is updated in the frequency domain.

**3) Update of Subject-Specific Time Delays  $\mathcal{T}$ :** As stated earlier,  $\mathcal{T} = \{\tau_{k,n}\} \in \mathbb{R}^{K \times N}$  totally includes  $K \times N$  time delays. Each of the time delays is optimized using the rule

of ALS, i.e., the remaining time delays are fixed when a time delay is updated. Next, we take an exemplifying time delay  $\tau_{k,n'}$  for component  $n'$  ( $n' = 1, \dots, N$ ) of subject  $k$  ( $k = 1, \dots, K$ ) to derive the update rule.

Referring to (5), the update of the time delay  $\tau_{k,n'}$  minimizes the following least squared error:

$$\arg \min_{\tau_{k,n'}} \|\mathbf{x}_{(3)k} - (\mathbf{B}^{(k)} \odot \mathbf{S})\mathbf{c}_k\|^2. \quad (9)$$

Because the shared SMs  $\mathbf{S} \in \mathbb{C}^{V \times N}$  is not relevant to  $\mathcal{T}$ , and the number of in-brain voxels  $V$  is very large (usually larger than  $5 \times 10^4$ ), we project (3) into the  $\mathbf{S}$  subspace to reduce the dimensionality of the tensor as follows:

$$\mathbf{S}^T \mathbf{X}_{(1)} = \mathbf{S}^T (\mathbf{S}\mathbf{Z}^T + \mathbf{E}_{(1)}) = \hat{\mathbf{S}}\mathbf{Z}^T + \hat{\mathbf{E}}_{(1)} = \hat{\mathbf{X}}_{(1)}. \quad (10)$$

Transform  $\hat{\mathbf{X}}_{(1)} \in \mathbb{C}^{N \times J \times K}$  into a new tensor  $\hat{\mathbf{X}} \in \mathbb{C}^{N \times J \times K}$ . The size of  $\hat{\mathbf{X}}$  is significantly smaller than that of  $\mathbf{X} \in \mathbb{C}^{V \times J \times K}$ , as  $N \ll V$ . As a result, the estimate of  $\mathcal{T}$  is accelerated and the computation memory is saved.

With the new tensor  $\hat{\mathbf{X}}$ , we rewrite (5) as follows:

$$\hat{\mathbf{x}}_{(3)k} = (\mathbf{B}^{(k)} \odot \hat{\mathbf{S}})\mathbf{c}_k + \hat{\mathbf{e}}_{(3)k}, \quad (11)$$

where  $\hat{\mathbf{x}}_{(3)k} \in \mathbb{C}^{J \times N}$  and  $\hat{\mathbf{S}} = \mathbf{S}^T \mathbf{S} \in \mathbb{C}^{N \times N}$ . We then have the following vector presentation for the Khatri-Rao product between  $\mathbf{B}^{(k)}$  and  $\hat{\mathbf{S}}$ :

$$\hat{\mathbf{x}}_{(3)k} = \sum_{n=1}^N c_{k,n} (\mathbf{b}_n^{(k)} \odot \hat{\mathbf{s}}_n) + \hat{\mathbf{e}}_{(3)k}. \quad (12)$$

To explicitly show the time delay  $\tau_{k,n'}$ , we rewrite (9) based on (12) as follows:

$$\arg \min_{\tau_{k,n'}} \|\hat{\mathbf{x}}_{(3)k} - \sum_{n \neq n'} c_{k,n} (\mathbf{b}_n^{(k)} \odot \hat{\mathbf{s}}_n) - c_{k,n'} (\mathbf{b}_{n'}^{(k)} \odot \hat{\mathbf{s}}_{n'})\|^2. \quad (13)$$

Let  $\mathbf{q}_{n'}^{(k)} = \hat{\mathbf{x}}_{(3)k} - \sum_{n \neq n'} c_{k,n} (\mathbf{b}_n^{(k)} \odot \hat{\mathbf{s}}_n)$  denote the remaining signal when projecting all but component  $n'$  ( $n' = 1, \dots, N$ ) out of  $\hat{\mathbf{x}}_{(3)k}$ , thus (13) becomes the following:

$$\begin{aligned} & \arg \min_{\tau_{k,n'}} \|\mathbf{q}_{n'}^{(k)} - c_{k,n'} (\mathbf{b}_{n'}^{(k)} \odot \hat{\mathbf{s}}_{n'})\|^2 \\ & = \arg \min_{\tau_{k,n'}} \left[ \begin{aligned} & \|\mathbf{q}_{n'}^{(k)}\|^2 + \|c_{k,n'} (\mathbf{b}_{n'}^{(k)} \odot \hat{\mathbf{s}}_{n'})\|^2 \\ & - 2\operatorname{Re}\{\mathbf{q}_{n'}^{(k)T}\} \operatorname{Re}\{c_{k,n'} (\mathbf{b}_{n'}^{(k)} \odot \hat{\mathbf{s}}_{n'})\} \\ & - 2\operatorname{Im}\{\mathbf{q}_{n'}^{(k)T}\} \operatorname{Im}\{c_{k,n'} (\mathbf{b}_{n'}^{(k)} \odot \hat{\mathbf{s}}_{n'})\} \end{aligned} \right] \quad (14) \end{aligned}$$

In the square brackets of (14), the first term does not contain the information of  $\tau_{k,n'}$ , and the second term does not vary with  $\tau_{k,n'}$ , as  $\mathbf{b}_{n'}^{(k)}$  is cyclic shifted based on the time delay  $\tau_{k,n'}$ . Thus, the update of the time delay  $\tau_{k,n'}$  maximizes the sum of the third and fourth terms as follows:

$$\arg \max_{\tau_{k,n'}} \left[ \begin{aligned} & \operatorname{Re}\{\mathbf{q}_{n'}^{(k)T}\} \operatorname{Re}\{c_{k,n'} (\mathbf{b}_{n'}^{(k)} \odot \hat{\mathbf{s}}_{n'})\} \\ & + \operatorname{Im}\{\mathbf{q}_{n'}^{(k)T}\} \operatorname{Im}\{c_{k,n'} (\mathbf{b}_{n'}^{(k)} \odot \hat{\mathbf{s}}_{n'})\} \end{aligned} \right] \quad (15)$$

By expanding each term in (15) to extract the time delay,  $\tau_{k,n'}$  is finally obtained by the following:

$$\hat{\tau}_{k,n'} = \arg \max_{1 \leq j \leq J} |\varphi_{k,n'}(j)|, \quad \tau_{k,n'} = \hat{\tau}_{k,n'} - J + 1. \quad (16)$$

The detailed derivation of (16) from (15) can be found in the supplementary materials.<sup>1</sup>

The time delay we estimate here is an integer. A non-integer time delay can also be estimated using a gradient based iteration update, but this can result in a local minima [30] and, therefore, be time-consuming.

### B. Phase Sparsity Constraint on Shared Spatial Maps S

In the above-described csCPD algorithm, the shared SMs  $\mathbf{S}$  is estimated using (6). However, each component of  $\mathbf{S}$  includes a large number of high-magnitude unwanted voxels due to the high noise nature of complex-valued fMRI data [15]. As such, we propose the addition of a sparsity constraint on  $\mathbf{S}$  to remove these unwanted voxels.

The sparsity constraint is usually imposed on the magnitude of a complex-valued signal. However, the high-magnitude, unwanted voxels involved in the shared SM components lead to the failure of placing a sparsity constraint directly on the magnitude. For this reason, we propose to impose sparsity on the phase of shared SMs according to the small phase characteristic of BOLD-related voxels [15]. To achieve more direct and efficient effects on the newest estimate of shared SMs, the phase sparsity constraint is performed as a sequential update for  $\mathbf{S}$ , instead of being incorporated into the csCPD algorithm given in (1). Next, we obtain the corrected phase of shared SMs via phase de-ambiguity; we then present the model of phase sparsity constraint based on the smoothed  $\ell_0$  norm and the second update rule of  $\mathbf{S}$ .

**1) Phase de-Ambiguity:** Given a shared SM component  $\mathbf{s}_n = [s_{1,n}, \dots, s_{V,n}]^T$  of  $\mathbf{S}$  ( $n = 1, \dots, N$ ), its associated TC estimate is  $\mathbf{z}_n$ , the  $n$ th column of the aggregating mixing matrix  $\mathbf{Z}$ . The rotation angle  $\hat{\theta}_n$  for  $\mathbf{s}_n$  is determined by maximizing the power of the real part of the rotated  $\mathbf{z}_n$  as follows [15]:

$$\hat{\theta}_n = \arg \max_{0 \leq \theta \leq \pi} \|\operatorname{Re}\{\exp\{-i\theta\}\mathbf{z}_n\}\|_2. \quad (17)$$

The phase without ambiguity (i.e., spatial source phase) for  $\mathbf{s}_n$  is then obtained as follows [15]:

$$\theta(\mathbf{s}_n) = \theta \left( \exp\{i\hat{\theta}_n\}\mathbf{s}_n \right). \quad (18)$$

Note that the phase de-ambiguity in (18) does not change the magnitude  $|s_{v,n}|$  but it changes the phase  $\theta(s_{v,n})$  of each voxel in  $\mathbf{S}$ . The phase  $\theta(s_{v,n}) \in [-\pi, \pi]$  is capable of identifying the BOLD-related voxels (with smaller phase  $\theta(s_{v,n}) \in [-4/\pi, 4/\pi]$ ) from the unwanted voxels (with larger phase  $\theta(s_{v,n}) \in [-\pi, -4/\pi]$ , or  $\theta(s_{v,n}) \in (4/\pi, \pi]$ ). As such, the phase sparsity constraint can be imposed on the voxels with larger  $\theta(s_{v,n})$  in the shared SMs.

**2) The Phase Sparsity Constraint Model:** Given  $\mathbf{s}_v = [s_{v,1}, \dots, s_{v,N}]$  estimated by (6), we use a smoothed  $\ell_0$  function to minimize large-phase voxels in  $\mathbf{s}_v$  with an additional regularization term satisfying (3), as follows:

$$\begin{cases} \arg \min_{\mathbf{s}_v} \sum_{v=1}^V F_\sigma \{|\mathbf{s}_v|, \theta(s_{v,n})\} \\ \text{s.t. } \mathbf{x}_{(1)v} = \mathbf{Z}\mathbf{s}_v + \mathbf{e}_{(1)v}, \end{cases} \quad (19)$$

<sup>1</sup>Supplementary materials are available in the supporting documents/multimedia tab.

where

$$F_\sigma \{|\mathbf{s}_v|, \theta(s_{v,n})\} = \sum_{n=1}^N f_\sigma \{|s_{v,n}|, \theta(s_{v,n})\}, \quad (20)$$

and

$$f_\sigma \{|s_{v,n}|, \theta(s_{v,n})\} = \begin{cases} 1 - \exp\left\{-|s_{v,n}|^2/2\sigma^2\right\}, & |\theta(s_{v,n})| \geq \theta_n^{th} \\ 1, & |\theta(s_{v,n})| < \theta_n^{th} \end{cases} \quad (21)$$

where  $\theta_n^{th}$  denotes a threshold of  $|\theta(s_{v,n})| \in [0, \pi]$  for adding a sparsity constraint on the unwanted voxels in  $\mathbf{s}_n$ . Although larger phase changes of the unwanted voxels account for theoretically 75% ( $|\theta(s_{v,n})| \in (\pi/4, \pi]$ ) and actually 60–70% of the full phase changes [15], a smaller percentage of the voxels should be removed in the separation process to retain more BOLD-related voxels. Here, we select  $\theta_n^{th}$  as the phase value segmenting the largest  $V/3$  values of  $|\theta(s_{v,n})|$ . As a result, only 33% of the total voxels are gradually removed in the separation process via phase sparsity.

The function  $F_\sigma \{|\mathbf{s}_v|, \theta(s_{v,n})\}$  in (19) is equivalent to the minimization of the  $\ell_0$  norm for a sufficiently small  $\sigma$ . When  $|s_{v,n}| \ll \sigma$ ,  $|f_\sigma \{|s_{v,n}|, \theta(s_{v,n})\}| \rightarrow 0$ , while when  $|s_{v,n}| \gg \sigma$ ,  $|f_\sigma \{|s_{v,n}|, \theta(s_{v,n})\}| \rightarrow 1$ . The larger the value of  $\sigma$  is, the smoother the  $f_\sigma \{|s_{v,n}|, \theta(s_{v,n})\}$  is, i.e., the smaller the value of  $\sigma$  is, the closer the behavior of  $f_\sigma \{|s_{v,n}|, \theta(s_{v,n})\}$  is to the  $\ell_0$  norm [31], [32],  $\lim_{\sigma \rightarrow 0} f_\sigma \{|s_{v,n}|, \theta(s_{v,n})\} = \|s_{v,n}\|_0$  [31].

**3) Update of S Based on Phase Sparsity:** For the model of the phase sparsity constraint given in (19), we use the steepest descent method to minimize the smoothed  $\ell_0$  function on the feasible set satisfying  $\mathbf{x}_{(1)v} = \mathbf{Z}\mathbf{s}_v + \mathbf{e}_{(1)v}$ . The update rule for  $\mathbf{s}_v$  is obtained as follows:

$$\begin{cases} \mathbf{s}_v \leftarrow \mathbf{s}_v - \lambda \Delta \mathbf{s}_v \\ \mathbf{s}_v \leftarrow \mathbf{s}_v - \mathbf{Z}^\dagger (\mathbf{Z}\mathbf{s}_v - \mathbf{x}_{(1)v}), \end{cases} \quad (22)$$

where

$$\Delta \mathbf{s}_v = \begin{bmatrix} \exp\{\theta(s_{v,1})\} f'_\sigma \{|s_{v,1}|, \theta(s_{v,1})\} \\ \vdots \\ \exp\{\theta(s_{v,N})\} f'_\sigma \{|s_{v,N}|, \theta(s_{v,N})\} \end{bmatrix}, \quad (23)$$

$f'_\sigma \{|s_{v,n}|, \theta(s_{v,n})\} = \partial f_\sigma \{|s_{v,n}|, \theta(s_{v,n})\} / \partial |s_{v,n}|$  is the first deviation of  $f_\sigma \{|s_{v,n}|, \theta(s_{v,n})\}$ :

$$f'_\sigma \{|s_{v,n}|, \theta(s_{v,n})\} = \begin{cases} \frac{|s_{v,n}|}{\sigma^2} \exp\left\{-\frac{|s_{v,n}|^2}{2\sigma^2}\right\}, & |\theta(\hat{s}_{v,n})| \geq \theta_n^{th} \\ 0, & |\theta(\hat{s}_{v,n})| < \theta_n^{th}, \end{cases} \quad (24)$$

and  $\lambda$  is a positive step size.

In summary, the proposed algorithm updates  $\mathbf{S}$  twice by using (6) and (22), correspondingly. To escape from local minima and singular values,  $\sigma$  in (24) should be slowly decreased [31], [32]. We let  $\sigma_{\text{iter}} = \gamma \sigma_{\text{iter}-1}$ , where ‘iter’ denotes the iteration index,  $\gamma$  is the decrease rate, and  $0.9 < \gamma < 1$  (we set  $\gamma = 0.99$ ). The update order for each iteration is  $\mathbf{B}$ ,  $\mathbf{T}$ ,  $\mathbf{S}$ , and  $\mathbf{C}$ . A detailed procedure for the proposed algorithm is described in Table I.

TABLE I  
IMPLEMENTATION OF THE PROPOSED ALGORITHM

---



---

**Input:** complex-valued multi-subject fMRI data  $\mathbf{X} \in \mathbb{C}^{V \times J \times K}$ , the number of components  $N$ , the initial  $\sigma$  named  $\sigma_0$ , the decrease rate  $\gamma$ , the step size  $\lambda$ , the termination threshold of relative error changes  $\Delta\epsilon_{\text{iter\_min}}$ , the termination threshold of error  $\epsilon_{\text{iter\_min}}$ , and the maximum number of iterations  $\text{iter}_{\text{max}}$ .

---

Randomly initialize  $\mathbf{B}$ ,  $\mathcal{T}$ ,  $\mathbf{S}$ , and  $\mathbf{C}$ ,  $\text{iter} = 0$ ,  $\Delta\epsilon_{\text{iter}} = 1$ ;

calculate the initial error  $\epsilon_0 = \sum_{v,j,k} e_{v,j,k} / VJK$  based on (1);

**while**  $\Delta\epsilon_{\text{iter}} > \Delta\epsilon_{\text{iter\_min}}$ ,  $\epsilon_{\text{iter}} > \epsilon_{\text{iter\_min}}$  or  $\text{iter} < \text{iter}_{\text{max}}$  **do**  
   $\text{iter} = \text{iter} + 1$ ;  
  update  $\mathbf{B}$  based on (7);  
  update  $\mathcal{T}$  based on (16);  
  first update  $\mathbf{S}$  based on (6);  
  **for**  $n=1:N$  **do**  
    perform phase de-ambiguity on  $\mathbf{s}_n$  using (17) and (18) to obtain  $\theta(\mathbf{s}_n)$ ;  
    sort  $|\theta(\mathbf{s}_n)|$  in descending order, determine  $\theta_n^{v/3}$ , and let  $\theta_n^{\text{th}} = \theta_n^{v/3}$ ;  
  **end for**  
  second update  $\mathbf{S}$  using (22);  
   $\sigma_{\text{iter}+1} = \gamma\sigma_{\text{iter}}$ ;  
  update  $\mathbf{C}$  using (8);  
  calculate  $\epsilon_{\text{iter}}$  and  $\Delta\epsilon_{\text{iter}} = |(\epsilon_{\text{iter}-1} - \epsilon_{\text{iter}}) / \epsilon_{\text{iter}-1}|$ ;  
**end while**

---

**Output:**  $\mathbf{B}$ ,  $\mathcal{T}$ ,  $\mathbf{S}$ , and  $\mathbf{C}$ .

---



---

### III. EXPERIMENTAL METHODS

To evaluate the efficacy of the proposed pcsCPD method, we carry out extensive experiments using both simulated and experimental complex-valued fMRI data, with comparison to the following three algorithms: (1) csCPD (without a spatial constraint, see subsection II.A); (2) CPD (without spatiotemporal constraints) using the COMFAC algorithm, which is a fast implementation of trilinear ALS [33]; and (3) T-sICA using the complex-valued entropy bound minimization (EBM) algorithm [34] to perform ICA. Each algorithm is repeated 20 times for each case. The estimated integer time delays range from  $-10$  points to  $10$  points. We test  $\sigma_0$  values from  $0.5$  to  $20$  and  $\lambda$  values from  $0.1$  to  $5$  according to the suggestions in [31], [32] and finally select  $\sigma_0 = 1$  and  $\lambda = 0.5$  for the simulated fMRI data, as well as  $\sigma_0 = 2$  and  $\lambda = 4$  for the experimental fMRI data. In addition, we choose  $\Delta\epsilon_{\text{iter\_min}} = 10^{-6}$ ,  $\epsilon_{\text{iter\_min}} = 10^{-4}$  and  $\text{iter}_{\text{max}} = 500$  for csCPD, CPD, and pcsCPD.

For each of shared SM estimate, i.e.,  $\mathbf{s}_n$ , phase de-ambiguity is first performed using (17) and (18) to obtain  $\theta(s_{v,n})$ , phase de-noising is then performed based on  $\theta(s_{v,n})$ . Specifically, a binary mask  $\mathbf{m}_n = \{m_n(v)\}$ ,  $v = 1, \dots, V$ , is generated as follows:

$$m_n(v) = \begin{cases} 1, & \text{if } |\theta(s_{v,n})| \leq \pi/4 \text{ and } |s_{v,n}| \geq 0.5 \\ 0, & \text{otherwise} \end{cases} \quad (25)$$

The phase-denoised  $\mathbf{s}_n$  is obtained by masking the following:

$$\mathbf{s}_n = \mathbf{s}_n \cdot \mathbf{m}_n. \quad (26)$$

Each shared TC estimate  $\mathbf{b}_n$  is also phase corrected. Replacing  $\mathbf{z}_n$  in (17) with  $\mathbf{b}_n$  to estimate its rotation angle  $\hat{\theta}_n$ , we obtain the phase corrected TC as follows:

$$\mathbf{b}_n = \exp\{-i\hat{\theta}_n\}\mathbf{b}_n. \quad (27)$$

The denoised SMs  $\mathbf{s}_n$  by (26) and the phase-corrected TCs  $\mathbf{b}_n$  ( $n = 1, \dots, N$ ) are used for evaluating the performance of the proposed algorithm.

When comparing with magnitude-only fMRI data analysis, we select the real-valued ICA-sCPD algorithm, which simultaneously utilizes spatial independence and temporal delay constraints [11]. Comparisons with real-valued sCPD [8] and CPD are provided in the supplementary materials.<sup>1</sup> The ICA in ICA-sCPD is selected as the fastICA algorithm in deflation mode with a ‘‘tanh’’ nonlinearity function. The stopping criterion is set to less than  $10^{-6}$  or a maximum iteration of  $1000$ .

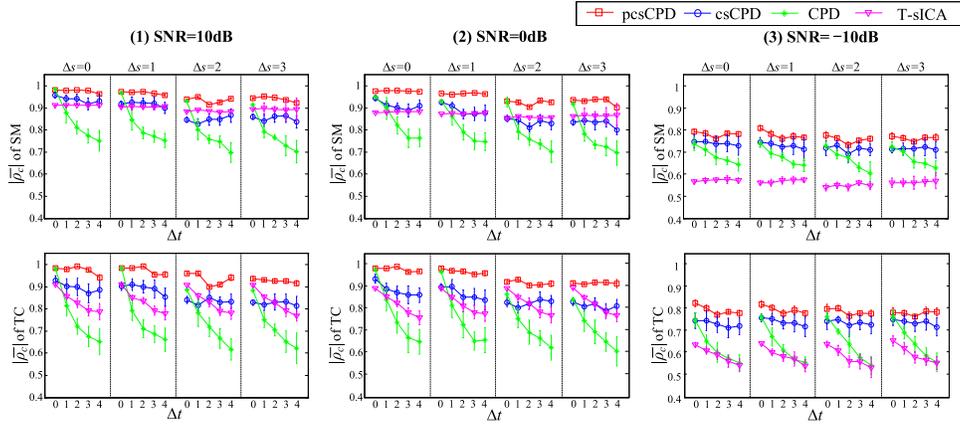
#### A. Simulated fMRI Data

We generate simulated complex-valued fMRI data sets with  $K = 10$  subjects, and each subject includes eight fMRI-like complex-valued sources (<http://mlsp.umbc.edu/resources.html>). One is task-related (S1), two are transiently task-related (S2, S6), and five are artifact related (S3, S4, S5, S7, and S8). Each simulated source is a spatial map of  $V = 60 \times 60$  voxels with a time course of  $J = 100$  time points. Reshaping each spatial map into a one-dimensional vector ( $8 \times 3600$ ) and mixing with eight time courses ( $100 \times 8$ ), we obtain simulated complex-valued fMRI data sets ( $3600 \times 100 \times 10$ ). To simulate the experimental fMRI data with spatial and temporal variability, we further change the above-generated ideal SMs and TCs by adding various spatial changes and different time delays for each subject. Specifically, we randomly decrease the activated SM voxels in the range of  $0\%$  (no change),  $10\%$ ,  $20\%$ , and  $30\%$  (denoted as  $\Delta s = 0, 1, 2, 3$ ), with respect to the original dataset. As the periods of the task-related and transiently task-related TCs are  $25$  points and  $20$  points, respectively, we randomly shift the time courses in the range of  $0\%$ ,  $2\%$ ,  $4\%$ ,  $8\%$  and  $10\%$  (denoted as  $\Delta t = 0, 1, 2, 3, 4$ ), with respect to the total points of  $100$ . Thus, the time delay change is within  $[-10, 10]$ . Moreover, we add complex-valued Gaussian noise to the simulated fMRI data with different signal-to-noise ratio (SNR) levels from  $-20$  dB to  $20$  dB with a  $2.5$  dB interval (a total of  $17$  different levels of SNR). The SNR is defined as  $20 \lg(\hat{\sigma}_s / \hat{\sigma}_n)$ , where  $\hat{\sigma}_s$  and  $\hat{\sigma}_n$  are the temporal standard deviations of the source signal and Gaussian noise, respectively. We set the number of components for each algorithm to be the true number of components  $N = 8$ .

We use the simulated fMRI data to evaluate the effects of spatial and temporal changes and the noise effect on pcsCPD, csCPD, CPD, and T-sICA.

#### B. Experimental fMRI Data

In this study, we use complex-valued fMRI data sets from  $10$  subjects performing a finger-tapping motor task while receiving auditory instructions [15], [35]. All participants signed IRB-approved informed consent at the University of



**Fig. 1.** Comparison of pcsCPD, csCPD, CPD and T-sICA in terms of the mean and standard deviation of  $|\bar{\rho}_c|$  for the SM and TC estimates of all eight components, under 20 cases of spatial and temporal changes ( $\Delta s = 0, 1, 2, 3$  and  $\Delta t = 0, 1, 2, 3, 4$ ) and three levels of noise. (1) SNR = 10 dB, (2) SNR = 0 dB, (3) SNR = -10 dB.

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. 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 =  $64 \times 64$ , TE = 29 ms, TR = 2 s, and flip angle = 70 degrees. The Statistical Parametric Mapping (SPM) software package was used to perform the preprocessing of the data. The magnitude datasets were coregistered to compensate for movements in the fMRI time series images. The images were then spatially normalized into standard Montreal Neurological Institute space. Following spatial normalization, the data (real and imaginary) were slightly resampled using trilinear interpolation, resulting in  $53 \times 63 \times 46$  voxels. Motion correction and spatial normalization parameters were calculated from the magnitude data and applied to the phase data. Both the real and imaginary images were then spatially smoothed with a  $10 \times 10 \times 10$  mm<sup>3</sup> full-width-at-half-maximum Gaussian kernel.

Prior to the analysis, we filtered the experimental fMRI data by an 8–150 mHz band-pass filter [36], [37]. The model order of each algorithm was obtained by first estimating the model order for each single subject of fMRI data using the minimum description length (MDL) criterion with a sub-sampling scheme [40] and then averaging over 10 subjects. The final model order was set to 50, which is consistent with our previous study [35].

### C. Performance Indices

We calculate (1) the absolute correlation coefficient  $|\rho_c|$  between the estimates of the shared SM and TC, the subject-specific time delay and intensity and their corresponding references; and (2) the mean and standard deviation of  $|\rho_c|$  over all 20 runs for each case, to evaluate the algorithm performance for both the simulated and experimental fMRI data.

The simulated fMRI data includes references as the ground truth. For the experimental fMRI data, we use the model TC as the reference for the task-related TC, which is acquired

by convolving the stimuli with the canonical SPM hemodynamic response function; and we use a group general linear model (GLM) map as the reference for the task-related SM. The GLM reference is specifically obtained by performing one sample  $t$ -test ( $p < 0.05$ ) on the single-subject GLM results [11], [15]. The time delay reference is the estimates corresponding to the largest  $|\rho_c|$  values between the lag-shifted TC estimates of complex-valued EBM and the model TC. The subject intensity is measured via the similarity between the proposed method and the other algorithms in terms of  $|\rho_c|$ .

In addition, the quality of the phase of the SM and TC estimates is evaluated by the performance indices described above for the experimental fMRI data. Since the SM phase has frequently varying polarity, we use the source phase mask  $\mathbf{m}_n$  given in (25) to calculate the correlation with the GLM reference.

## IV. RESULTS

### A. Simulated fMRI Data

We first compare the proposed pcsCPD with csCPD, CPD and T-sICA under 20 cases of spatial and temporal changes ( $\Delta s = 0, 1, 2, 3$ ;  $\Delta t = 0, 1, 2, 3, 4$ ), and take SNR = -10 dB, 0 dB and 10 dB as examples to test the noise effects. Fig. 1 shows the results. The average  $|\rho_c|$  over all eight components is analyzed and denoted as  $|\bar{\rho}_c|$ . We observe from Fig. 1 that our proposed algorithm yields the highest means and the lowest standard deviations of  $|\bar{\rho}_c|$  for all cases when compared with the other three algorithms. CPD can obtain better performance for the case of  $\Delta s, \Delta t = 0$  where the simulated fMRI data conforms to the CPD model, but it sharply degrades with the increase in  $\Delta s$  and  $\Delta t$ .

We examine the effects of spatial and temporal changes on the algorithms by considering the case of SNR = 10 dB, as shown in Fig. 1(1). The proposed approach exhibits the slightest effects, whereas CPD shows the most severe effects in terms of both the means and standard deviations of  $|\bar{\rho}_c|$ . As expected, csCPD without a spatial constraint demonstrates larger degradation with the increase in  $\Delta s$  than in  $\Delta t$ . In contrast, T-sICA without a temporal constraint illustrates a larger degradation with the increase in  $\Delta t$  than in  $\Delta s$ .

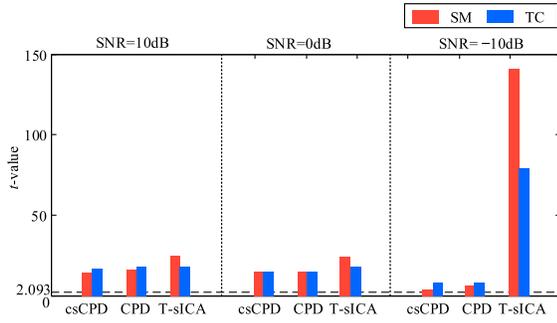


Fig. 2. The  $t$ -values of a two-sample  $t$ -test for evaluating the difference in the mean  $|\bar{\rho}_c|$  values for 20 cases of spatial and temporal changes, as displayed in Fig. 1, between pcsCPD and csCPD/CPD/T-sICA, under three noise levels. The  $t$ -value = 2.093 ( $p$ -value = 0.05,  $df = 19$ ) is shown as a threshold.

Generally speaking, T-sICA is better than csCPD in detecting SM, while csCPD is better than T-sICA in detecting TC.

We investigate the effects of the noise levels on the algorithms by decreasing the SNR from 10 dB to 0 dB and -10 dB, see Figs. 1(2) and (3). T-sICA shows considerable degradation for SNR = -10 dB, while the other three algorithms exhibit reasonable decreases in the means and standard deviations of  $|\bar{\rho}_c|$  for the shared SM and TC estimates.

Fig. 2 includes  $t$ -values calculated by a two-sample  $t$ -test for evaluating the difference in the mean  $|\bar{\rho}_c|$  for all 20 cases of spatial and temporal changes under SNR = -10 dB, 0 dB and 10 dB, as displayed in Fig. 1, between the proposed pcsCPD and each of the csCPD, CPD and T-sICA. The  $t$ -values range from 3.913 to 140.812, all of which are larger than 2.093 ( $p$ -value = 0.05, degree of freedom = 19). This result verifies that the proposed approach achieves significant improvements over csCPD, CPD and T-sICA.

We further present the results of the noise influence from -20 dB to 20 dB (17 levels of SNR) in two example cases, a smaller spatial and temporal change case ( $\Delta s = 1$ ,  $\Delta t = 1$ ) and a larger spatial and temporal change case ( $\Delta s = 2$ ,  $\Delta t = 3$ ). Fig. 3 includes the means and standard deviations of the  $|\bar{\rho}_c|$  values for all four algorithms. The proposed algorithm exhibits the highest mean  $|\bar{\rho}_c|$  values with smaller standard deviations for the two cases. CPD and T-sICA alternatively obtain the lowest mean  $|\bar{\rho}_c|$  values from high to low noise levels. The csCPD algorithm obtains moderate performance among these four algorithms.

For the larger spatiotemporal change ( $\Delta s = 2$ ,  $\Delta t = 3$ ), the distinction between the algorithms becomes wider than the smaller spatiotemporal change ( $\Delta s = 1$ ,  $\Delta t = 1$ ). T-sICA benefits from incorporating spatial independence in estimating the shared SMs (see Fig. 3B(1)), whereas csCPD benefits from incorporating the time delay in estimating the shared TC (see Fig. 3B(2)). The  $t$ -value of the two-sample  $t$ -test ranges from 5.927 to 16.335, all of which are larger than 2.120 ( $p$ -value = 0.05,  $df = 16$ ), as displayed in Fig. 4, confirming the significant improvement of the proposed pcsCPD over csCPD, CPD and T-sICA.

Finally, we compare the means and standard deviations of the  $|\rho_c|$  values for the shared SM and TC, and the subject-specific time delay and intensity estimates obtained in two example cases, i.e., an easy case  $\Delta s = 1$ ,  $\Delta t = 1$ ,

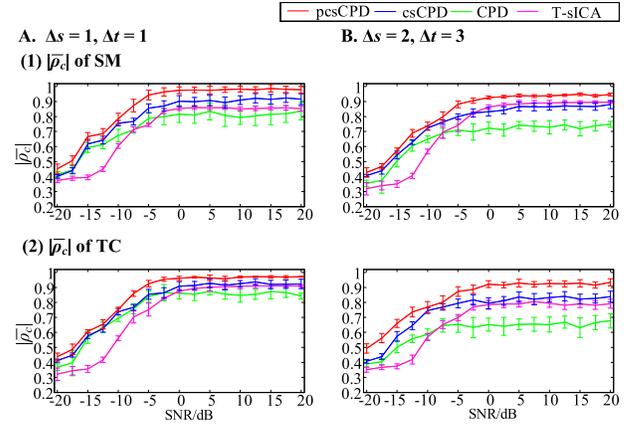


Fig. 3. Comparison of noise effects on pcsCPD, csCPD, CPD and T-sICA in terms of the mean and standard deviation of  $|\bar{\rho}_c|$  computed for the (1) SM and (2) TC estimates from all eight components. (A) Smaller spatiotemporal change ( $\Delta s = 1$ ,  $\Delta t = 1$ ). (B) Larger spatiotemporal change ( $\Delta s = 2$ ,  $\Delta t = 3$ ).

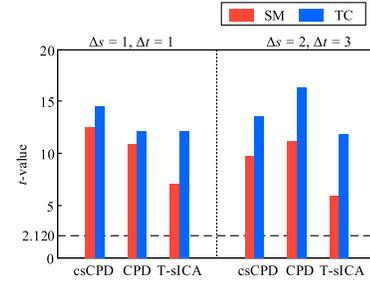


Fig. 4. The  $t$ -values of the two-sample  $t$ -test for evaluating the difference in the mean  $|\bar{\rho}_c|$  values for the two cases, as displayed in Fig. 3, between pcsCPD and csCPD/CPD/T-sICA. The  $t$ -value = 2.120 ( $p$ -value = 0.05,  $df = 16$ ) is shown as a threshold.

SNR = 10 dB, and a relatively hard case  $\Delta s = 2$ ,  $\Delta t = 3$ , SNR = -10 dB. Components 1, 6 and 8 are selected to represent the task-related, transiently task-related, and noise components, respectively. Fig. 5 shows the results. The proposed method yields the best performance for the three types of components under the two cases. In contrast, CPD, csCPD and T-sICA show sensitivity to the type of components, and the sensitivity increases in the hard case. Generally speaking, these three algorithms obtain the best estimations for the task-related component, which is a component of interest with a higher SNR for task fMRI data.

## B. Experimental fMRI Data

We examine the task-related component for motor task fMRI data and analyze the magnitude and phase of the task-related SM and TC. Table II illustrates the means and standard deviations of  $|\rho_c|$  for our proposed pcsCPD, as well as the csCPD, CPD and T-sICA. Our approach yields the highest average  $|\rho_c|$  values, followed by T-sICA, csCPD and CPD. In addition, the proposed method achieves relatively lower standard deviations compared to the other three algorithms. CPD obtains the worst performance in terms of both the means and standard deviations of  $|\rho_c|$ , indicating that the experimental fMRI data does not conform to the CPD model.

Fig. 6 shows the magnitude and phase images of the shared SM estimates ( $|Z| \geq 0.5$ ). For the SM magnitudes

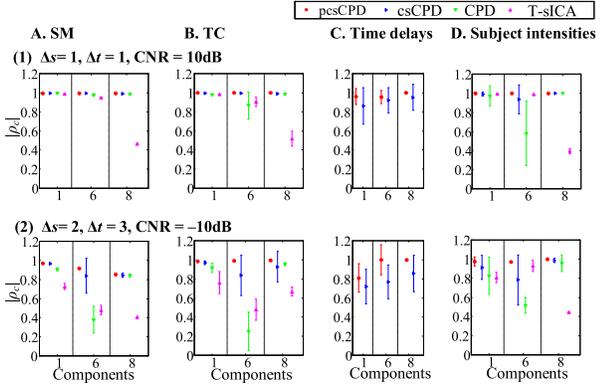


Fig. 5. Comparison of pcsCPD, csCPD, CPD, and T-sICA in terms of the mean and standard deviation of  $|\rho_c|$  computed for (A) SM, (B) TC, (C) Time delays, and (D) Subject intensities from Components 1, 6 and 8 under (1)  $\Delta s = 1$ ,  $\Delta t = 1$ , SNR = 10 dB and (2)  $\Delta s = 2$ ,  $\Delta t = 3$ , SNR = -10 dB.

TABLE II

COMPARISON OF PCSCPD, CS CPD, CPD, AND T-SICA FOR ANALYZING THE EXPERIMENTAL MULTI-SUBJECT FMRI DATA OVER 20 RUNS IN TERMS OF THE MEAN AND STANDARD DEVIATION OF  $|\rho_c|$  FOR THE MAGNITUDE AND PHASE OF THE SHARED TASK-RELATED SM AND TC ESTIMATES. THE BOLD VALUES ARE THE LARGEST MEANS AND THE LOWEST STANDARD DEVIATIONS IN EACH ROW

	pcsCPD	csCPD	CPD	T-sICA
SM magnitude	<b>0.56±0.05</b>	0.41±0.13	0.16±0.05	0.46±0.08
SM phase	<b>0.45±0.04</b>	0.37±0.12	0.08±0.08	0.44±0.08
TC magnitude	<b>0.85±0.02</b>	0.73±0.21	0.66±0.18	0.80±0.12
TC phase	<b>0.65±0.18</b>	0.36±0.26	0.39±0.19	0.52±0.16

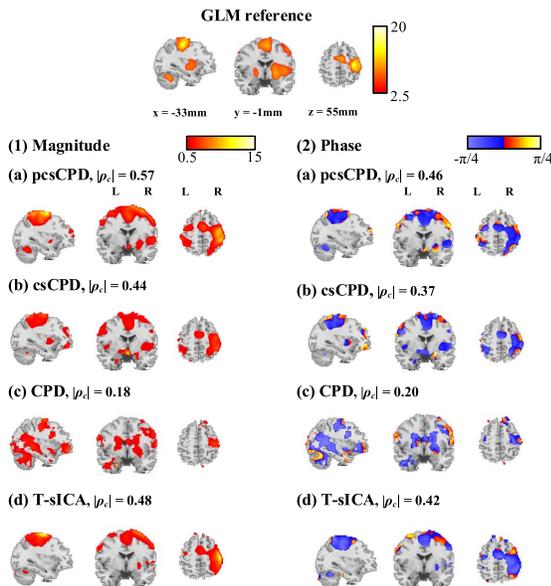


Fig. 6. Magnitude (1) and phase (2) maps of the shared task-related SMs estimated from the experimental fMRI data by (a) pcsCPD, (b) csCPD, (c) CPD, and (d) T-sICA. The absolute Pearson correlation coefficients ( $|\rho_c|$ ) with the GLM reference are also shown.

given in Fig. 6 (1), the proposed method detects the largest contiguous and reasonable activations in the left primary motor areas (LPMA), right primary motor areas (RPMA), and supplementary motor areas (SMA). T-sICA is better than

TABLE III

COMPARISON OF PCSCPD, CS CPD, CPD, AND T-SICA IN TERMS OF THE TOTAL NUMBER OF ACTIVATED VOXELS AND THE VOXELS INSIDE AND OUTSIDE THE GLM MASK FOR THE SM MAGNITUDES DISPLAYED IN Fig. 6(1)

	pcsCPD	csCPD	CPD	T-sICA
Total voxels	8348	7141	14103	5725
Inside GLM mask	3234	2401	2317	2665
Outside GLM mask	5114	4740	11786	3060

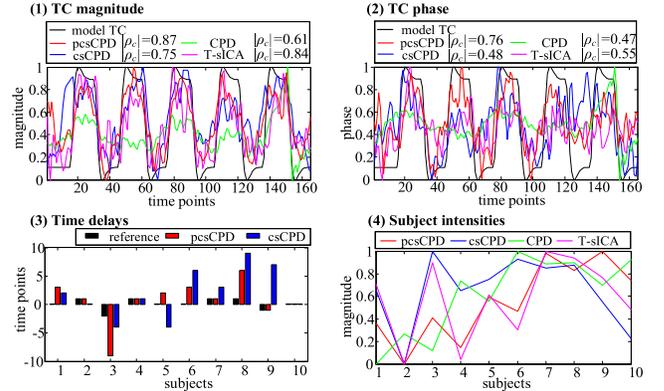
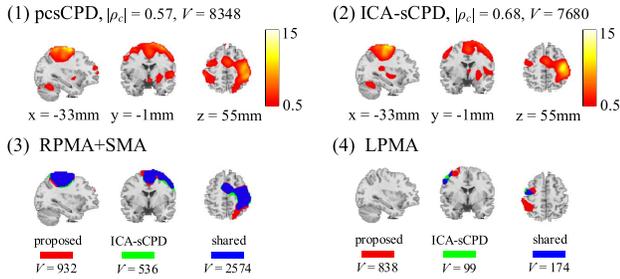


Fig. 7. Comparison of pcsCPD, csCPD, CPD, and T-sICA for the experimental fMRI data. (1) Magnitude and (2) phase parts of the TC estimates as well as their absolute Pearson correlation coefficients ( $|\rho_c|$ ) with the model TC. (3) Subject-specific time delays estimated by pcsCPD and csCPD. (4) Magnitude of the original magnitude and phase parts of the TC estimates and subject-specific intensities are 0-1 normalized before making comparisons among the four methods.

csCPD in terms of having less noisy voxels involved. CPD ranks last due to fewer motor-related activations and more noisy voxels. Table III includes the number of total activated voxels and the voxels inside and outside the GLM mask. Our proposed approach detects the highest number of voxels both inside and outside the GLM mask compared to T-sICA, and the voxels outside the GLM mask are mostly located in motor-related areas (see the supplementary materials for details<sup>1</sup>). CPD provides the largest number of total activated voxels, but with the minimal number of voxels inside the GLM mask and the maximal number of voxels unrelated to motor activations. When observing the SM phase maps shown in Fig. 6(2), the proposed pcsCPD obtains the highest  $|\rho_c|$  value, followed by T-sICA, csCPD, and CPD.

The shared TC, subject-specific time delay and intensity estimates are shown in Fig. 7. The proposed method yields the highest  $|\rho_c|$  values for the TC magnitude and phase compared to the other three algorithms. The TC phase is highly correlated with the model TC ( $|\rho_c| = 0.76$ ), which is consistent with previous studies [38]. Fig. 7(3) displays the time delays estimated by the proposed pcsCPD and csCPD. The average  $|\rho_c|$  values between the time delay estimates and the references from all subjects are 0.78 for the proposed method and 0.31 for csCPD. The time delays are accurately estimated for 5 subjects by pcsCPD and for 2 subjects by csCPD. Finally, the subject intensities are shown in Fig. 7(4). The  $|\rho_c|$  values between our method and the other three algorithms are 0.76 for T-sICA, 0.54 for CPD, and 0.32 for csCPD. The higher  $|\rho_c|$  between



**Fig. 8.** Comparison of the task-related magnitude maps estimated by (1) pcsCPD from the complex-valued fMRI data and (2) ICA-sCPD from the magnitude-only fMRI data. The  $|\rho_c|$  values and the number of activated voxels (shorted as “ $V$ ”) are shown. The maps and number of voxels detected by pcsCPD (in red), by ICA-sCPD (in green), or by both (in blue) are also compared for the task-related regions (3) RPMA+SMA and (4) LPMA, respectively.

pcsCPD and T-sICA indicates the high quality of the subject intensity estimates, as these two methods provide better results than csCPD and CPD for the SM and TC estimates.

### C. Comparison With the Magnitude-Only Method

Fig. 8 shows the comparison of the proposed method with ICA-sCPD in terms of the magnitude maps,  $|\rho_c|$  values and the number of activated voxels for the shared task-related SM estimates ( $|Z| \geq 0.5$ ). The proposed method obtains a lower  $|\rho_c|$  value but a larger number of activated voxels than ICA-sCPD, as shown in Figs. 8(1) and 8(2). Figs. 8(3) and 8(4) show the activated voxels detected by the proposed pcsCPD (in red), by ICA-sCPD (in green), or by both (in blue) in task-related regions, including RPMA and SMA (denoted as “RPMA+SMA”) and LPMA. The number of voxels detected individually by the proposed method is much larger than ICA-sCPD. More precisely, the proposed algorithm detects 73.9% more contiguous voxels in RPMA+SMA (932 vs. 536) and 746.5% more contiguous voxels in LPMA (838 vs. 99) than ICA-sCPD. In summary, our proposed method extracts 178.7% more contiguous task-related activations than ICA-sCPD, consistent with the previous finding that phase data provides additional brain information [15], [18], [20], [35], [39].

## V. DISCUSSION

This study proposes the use of the spatial source phase in the complex-valued CPD algorithm for decomposing complex-valued multi-subject fMRI data into shared SMs and TCs and subject-specific time delays and intensities. The spatial source phase sparsity and inter-subject temporal delays are simultaneously incorporated to address the inter-subject spatiotemporal variability among multi-subject fMRI data. The results from both the simulated and experimental fMRI data demonstrate the improvements of the proposed pcsCPD method over the complex-valued T-sICA, csCPD, and CPD algorithms, as well as the real-valued ICA-sCPD algorithm, in detecting more contiguous task-related activations.

The incorporation of spatial source phase sparsity is one contribution of this study to the analysis of complex-valued fMRI data. Although the spatial sparsity yields promising results in the magnitude-only analysis [24]–[29], it is not straightforward to incorporate sparsity for the high noise

complex-valued fMRI data analysis. A huge number of unwanted voxels have magnitude values as high as or even much higher than the BOLD-related voxels in the complex-valued SMs [15]. Therefore, the magnitude sparsity does not hold for complex-valued components as magnitude-only components do. Fortunately, the spatial source phase is qualified due to the small-phase characteristic of the BOLD-related voxels after phase de-ambiguity [15], which facilitates the sparse representation of complex-valued fMRI data to examine the intrinsic characteristic of brain activation.

In the proposed method, the spatial activation is mainly affected by three phase-sparsity-related parameters, the threshold  $\theta_n^{th}$  for adding the sparsity constraint on the unwanted voxels in  $s_n$ , the initial  $\sigma$  ( $\sigma_0$ ) in the smoothed  $\ell_0$  function given  $\sigma$  is slowly decreased, and the step size  $\lambda$  in the second update of  $\mathbf{S}$  based on phase sparsity. The selection of these parameters has a more dramatic influence on the experimental fMRI data than the simulated one because the simulated fMRI data have much sparser spatial activations than the experimental data. A reasonable criterion for choosing  $\theta_n^{th}$  is removing a smaller percentage (here 33%) of the total voxels in the separation process to retain more BOLD-related voxels. A larger  $\theta_n^{th}$  leads to the loss of BOLD-related voxels, while a smaller one causes the interference of unwanted voxels. In addition, a different sparsity of the shared SMs supports a different selection of  $\sigma_0$  and  $\lambda$  for the simulated and experimental fMRI data. Larger  $\sigma_0$  and  $\lambda$  values should be used for the experimental fMRI data to encourage greater sparsity.

Imposing both spatial and temporal constraints is essential for analyzing multi-subject complex-valued fMRI data. The CPD without spatiotemporal constraints obtains the worst separation performance. T-sICA shows better SM estimates than csCPD when the spatial changes are larger due to imposing the spatial independence constraint, and csCPD exhibits better TC estimates than T-sICA when the TC changes are larger due to incorporating the inter-subject TC delays. In view of the performance of the above three compared algorithms, our proposed method demonstrates the least sensitivity to spatiotemporal changes and robustness to noise in both the simulated and experimental fMRI data experiments, indicating the match between the multiple-subject fMRI data and the CPD model with spatiotemporal constraints.

Our previous studies have verified that the complex-valued fMRI data analysis can detect more contiguous and reasonable activations than magnitude-only fMRI data analysis [15], [18], [20], [35]. By using complex-valued ICA and IVA algorithms, the complex-valued analyses extracted 139% more voxels for ICA and 393% more voxels for IVA than the magnitude-only analyses for the task fMRI data [15], [35]. The proposed method also extracts more contiguous task-related activations (73.9% more for RPMA+SMA and 746.5% more for LPMA) than the magnitude-only ICA-sCPD method, supporting the uniqueness and complementarity of phase fMRI data to the magnitude fMRI data.

The use of the number of activated voxels and spatial correlation as performance metrics may be a possible limitation of the study. Future work includes the following: exploiting new parameter metrics; applying our proposed method to

resting-state fMRI data, which are more easily obtained and analyzed from patients with neurological disorders, to exact shared spatial activations and shared time courses with subject-specific time delays; examining how well a complex-valued approach can predict, e.g., patients from controls compared to a magnitude-only approach; and incorporating the identification of the real rank before calculating the loading matrices [41], [42]. Adaptive selection of the threshold of the spatial source phase in imposing the phase sparsity constraint on spatial maps also deserves future investigation.

## 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] C. F. Beckmann and S. M. Smith, "Tensorial extensions of independent component analysis for multisubject fMRI analysis," *NeuroImage*, vol. 25, no. 1, pp. 294–311, 2005.
- [3] 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.
- [4] J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an  $n$ -way generalization of 'Eckart-Young' decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, Sep. 1970.
- [5] R. A. Harshman, "Foundations of the PARAFAC procedure: Model and conditions for an explanatory multi-mode factor analysis," in *Proc. UCLA Working Papers Phonetics*, vol. 16, 1970, pp. 1–84.
- [6] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Appl.*, vol. 18, no. 2, pp. 95–138, 1977.
- [7] A. Stegeman, "Comparing independent component analysis and the Parafac model for artificial multi-subject fMRI data," Univ. Groningen, Groningen, The Netherlands, Tech. Rep., 2007. [Online]. Available: <http://www.alwinstegeman.nl/docs/Stegeman%20-%20parafac%20&%20ica.pdf>
- [8] 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.
- [9] N. E. Helwig and S. Hong, "A critique of tensor probabilistic independent component analysis: Implications and recommendations for multi-subject fMRI data analysis," *J. Neurosci. Methods*, vol. 213, no. 2, pp. 263–273, Mar. 2012.
- [10] G. Zhou and A. Cichocki, "Canonical polyadic decomposition based on a single mode blind source separation," *IEEE Signal Process. Lett.*, vol. 19, no. 8, pp. 523–526, Aug. 2012.
- [11] 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.
- [12] J.-H. Lee, T.-W. Lee, F. A. Jolesz, and S.-S. Yoo, "Independent vector analysis (IVA): Multivariate approach for fMRI group study," *NeuroImage*, vol. 40, no. 1, pp. 86–109, 2008.
- [13] M. De Vos, D. Nion, S. Van Huffel, and L. De Lathauwer, "A combination of parallel factor and independent component analysis," *Signal Process.*, vol. 92, no. 12, pp. 2990–2999, Dec. 2012.
- [14] S. Hong and R. A. Harshman, "Shifted factor analysis—Part III:  $N$ -way generalization and application," *J. Chemometrics*, vol. 17, no. 7, pp. 389–399, Jul. 2003.
- [15] M.-C. Yu, Q.-H. Lin, L.-D. Kuang, X.-F. Gong, F. Cong, and V. D. Calhoun, "ICA of full complex-valued fMRI data using phase information of spatial maps," *J. Neurosci. Methods*, vol. 249, pp. 75–91, Jul. 2015.
- [16] L.-D. Kuang, Q.-H. Lin, X.-F. Gong, F. Cong, and V. D. Calhoun, "Post-ICA phase de-noising for resting-state complex-valued fMRI data," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 856–860.
- [17] L. D. Kuang, Q.-H. Lin, X.-F. Gong, Y.-G. Chen, F. Cong, and V. D. Calhoun, "Model order effects on independent vector analysis applied to complex-valued fMRI data," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 81–84.
- [18] L. D. Kuang, Q.-H. Lin, X.-F. Gong, F. Cong, J. Sui, and V. D. Calhoun, "Model order effects on ICA of resting-state complex-valued fMRI data: Application to schizophrenia," *J. Neurosci. Methods*, vol. 304, pp. 24–38, Jul. 2018.
- [19] D. W. Adrian, R. Maitra, and D. B. Rowe, "Complex-valued time series modeling for improved activation detection in fMRI studies," *Ann. Appl. Statist.*, vol. 12, no. 3, pp. 1451–1478, 2018.
- [20] Y. Qiu *et al.*, "Spatial source phase: A new feature for identifying spatial differences based on complex-valued resting-state fMRI data," *Hum. Brain Mapping*, vol. 40, no. 9, pp. 2662–2676, Jun. 2019.
- [21] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1," *Vis. Res.*, vol. 37, no. 23, pp. 3311–3325, 1997.
- [22] P. Lennie, "The cost of cortical computation," *Current Biol.*, vol. 13, no. 6, pp. 493–497, Mar. 2003.
- [23] 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, Jun. 2009.
- [24] J. Lv *et al.*, "Sparse representation of group-wise fMRI signals," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, Nagoya, Japan, Sep. 2013, pp. 608–616.
- [25] J. Lv *et al.*, "Sparse representation of whole-brain fMRI signals for identification of functional networks," *Med. Image Anal.*, vol. 20, no. 1, pp. 112–134, Feb. 2015.
- [26] J. Lv *et al.*, "Task fMRI data analysis based on supervised stochastic coordinate coding," *Med. Image Anal.*, vol. 38, pp. 1–16, May 2017.
- [27] H. Shen *et al.*, "Making group inferences using sparse representation of resting-state functional MRI data with application to sleep deprivation," *Hum. Brain Mapping*, vol. 38, no. 9, pp. 4671–4689, Sep. 2017.
- [28] A.-K. Seghouane and A. Iqbal, "Sequential dictionary learning from correlated data: Application to fMRI data analysis," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 3002–3015, Jun. 2017.
- [29] J. Xie, P. K. Douglas, Y. N. Wu, A. L. Brody, and A. E. Anderson, "Decoding the encoding of functional brain networks: An fMRI classification comparison of non-negative matrix factorization (NMF), independent component analysis (ICA), and sparse coding algorithms," *J. Neurosci. Methods*, vol. 282, no. 17, pp. 81–94, 2017.
- [30] M. Mørup, K. H. Madsen, and L. K. Hansen, "Shifted independent component analysis," in *Proc. Int. Conf. Independ. Compon. Anal. Signal Separat.*, Sep. 2007, pp. 89–96.
- [31] G. H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "Complex-valued sparse representation based on smoothed  $\ell^0$  norm," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2008, pp. 3881–3884.
- [32] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed  $\ell^0$  norm," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 289–301, Jan. 2009.
- [33] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 810–823, Mar. 2000.
- [34] X.-L. Li and Tulay Adali, "Complex independent component analysis by entropy bound minimization," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 7, pp. 1417–1430, Jul. 2010.
- [35] L.-D. Kuang, Q.-H. Lin, X.-F. Gong, F. Cong, and V. D. Calhoun, "Adaptive independent vector analysis for multi-subject complex-valued fMRI data," *J. Neurosci. Methods*, vol. 281, pp. 49–63, Apr. 2017.
- [36] F. Cong *et al.*, "Key issues in decomposing fMRI during naturalistic and continuous music experience with independent component analysis," *J. Neurosci. Methods*, vol. 223, pp. 74–84, Feb. 2014.
- [37] C. Du, N. D. Volkow, A. P. Koretsky, and Y. Pan, "Low-frequency calcium oscillations accompany deoxyhemoglobin oscillations in rat somatosensory cortex," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 43, pp. 4677–4686, Oct. 2014.
- [38] H. Li, N. M. Correa, P. A. Rodriguez, V. D. Calhoun, and T. Adali, "Application of independent component analysis with adaptive density model to complex-valued fMRI data," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 10, pp. 2794–2803, Oct. 2011.
- [39] Z. Chen, A. Caprihan, E. Damaraju, S. Rachakonda, and V. Calhoun, "Functional brain connectivity in resting-state fMRI using phase and magnitude data," *J. Neurosci. Methods*, vol. 293, pp. 299–309, Jan. 2018.
- [40] W. Xiong, Y.-O. Li, N. Correa, X.-L. Li, V. D. Calhoun, and T. Adali, "Order selection of the linear mixing model for complex-valued fMRI data," *J. Signal Process. Syst.*, vol. 67, no. 2, pp. 117–128, May 2012.
- [41] X. Luciani and L. Albera, "Joint eigenvalue decomposition of non-defective matrices based on the LU factorization with application to ICA," *IEEE Trans. Signal Process.*, vol. 63, no. 17, pp. 4594–4608, Sep. 2015.
- [42] B. Yang, G. Wang, and N. D. Sidiropoulos, "Tensor completion via group-sparse regularization," in *Proc. IEEE 50th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2016, pp. 1750–1754.