

Received September 27, 2021, accepted October 12, 2021, date of publication October 19, 2021, date of current version November 1, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3121417

Taking the 4D Nature of fMRI Data Into Account Promises Significant Gains in Data Completion

IRINA BELYAEVA¹, (Graduate Student Member, IEEE), SUCHITA BHINGE¹,
QUNFANG LONG, (Member, IEEE), AND TÜLAY ADALI¹, (Fellow, IEEE)

Department of CSEE, University of Maryland, Baltimore County, Baltimore, MD 21250, USA

Corresponding author: Irina Belyaeva (irinbell@umbc.edu)

This work was supported in part by the NSF Grant NCS 1631838, and in part by the NIH Grant R01MH123610 and Grant R01MH118695.

ABSTRACT Functional magnetic resonance imaging (fMRI) is a powerful, noninvasive tool that has significantly contributed to the understanding of the human brain. fMRI data provide a sequence of whole-brain volumes over time and hence are inherently four dimensional (4D). Missing data in fMRI experiments arise from image acquisition limits, susceptibility and motion artifacts or during confounding noise removal. Hence, significant brain regions may be excluded from the data, which can seriously undermine the quality of subsequent analyses due to the significant number of missing voxels. We take advantage of the four dimensional (4D) nature of fMRI data through a tensor representation and introduce an effective algorithm to estimate missing samples in fMRI data. The proposed Riemannian nonlinear spectral conjugate gradient (RSCG) optimization method uses tensor train (TT) decomposition, which enables compact representations and provides efficient linear algebra operations. Exploiting the Riemannian structure boosts algorithm performance significantly, as evidenced by the comparison of RSCG-TT with state-of-the-art stochastic gradient methods, which are developed in the Euclidean space. We thus provide an effective method for estimating missing brain voxels and, more importantly, clearly show that taking the full 4D structure of fMRI data into account provides important gains when compared with three-dimensional (3D) and the most commonly used two-dimensional (2D) representations of fMRI data.

INDEX TERMS 4D fMRI processing, tensor completion, fMRI missing data completion, tensor train decomposition.

I. INTRODUCTION

Functional magnetic resonance imaging (fMRI) provides a noninvasive and indirect measure of brain activity and is a vital methodology in basic and applied neuroscience research. One of the common problems in fMRI analysis is the problem of missing data. The data collected during fMRI experiments often suffer from scanner instabilities, magnetic susceptibility artifacts [1], subject motion, and task-related noise [2]. An additional challenge arises when the fMRI data are processed in their native, complex domain, as the phase images exhibit a challenging noise structure [3]. The denoising technique is often used in complex-valued fMRI as a preprocessing step to increase the sensitivity of analyses [4], [5]. However, the denoising process results in the exclusion of a significant number of voxels from neuroimaging studies [3], [4]. The excluded brain regions at the cortical

boundaries increase the risk of false negatives. The removal of voxels with missing data may increase the probability of false positives when spatial extent thresholds are applied to establish significance because fewer voxels generate smaller cluster sizes to trigger significance [2]. Hence, missing data present an important challenge for fMRI analysis. This challenge can result in the loss of statistical power and affect the consequent inferences motivating the use of data completion techniques to recover/estimate the excluded or missing voxels.

The missing data in fMRI studies can be either missing at random (MAR) or not missing at random (MNAR), which implies random groups or contiguous groups of voxels are missing [6]. Magnetic field inhomogeneity is the main source of missing MAR values due to the variability in the response of various tissues to magnetization [7]. The variability in tissue magnetization may result in geometric distortions, which are especially pronounced at tissue/border edges, the frontal lobe proximal to the sinuses, and the temporal lobe proximal

The associate editor coordinating the review of this manuscript and approving it for publication was Kin Fong Lei¹.

to the mastoid air cells and air canals [8]. The MNAR pattern is a more challenging missing value pattern in fMRI data. For example, task-related or resting-state fMRI (rs-fMRI) experiments may lead to the nonrandom exclusion of voxels that are associated with brain activity or the effect of interest due to standard fMRI preprocessing and motion censoring techniques [9], [10]. Another example of MNAR pattern can arise as a result of complex-valued fMRI data quality control when noisy regions are eliminated from the phase images [4].

One approach for solving the incomplete fMRI data problem is to use robust estimation techniques to infer missing voxels. Data completion methods are suitable for this purpose [11]–[14]. Data completion methods infer the missing values from partially observed entries and structural properties of the data. Real-world fMRI datasets exhibit correlations among their samples and hence highly correlated latent factors, which make data completion a feasible task. The completion methods used to estimate missing values fall into three broad categories: methods based on expectation maximization (EM) [13], nonlinear optimization [15], and nuclear norm minimization [11], [16], [17]. These approaches are further categorized into low-rank matrix and tensor completion methods. The algorithms that have been used in [18]–[21] utilize low-rank matrix completion methods, in which each brain volume is transformed into a two-dimensional (2D) matrix $\mathbf{Y} \in \mathbb{R}^{T \times V}$, where T denotes the number of timepoints and V denotes the number of voxels for each brain volume.

The majority of tensor completion methods developed for medical imaging mainly focus on three-dimensional (3D) fMRI data [22]–[24]. In [22], authors made use of tensor completion for fMRI data to recover continuously missing 3D scans at select timepoints with the multiway delay embedding transform (MDT) and alternating least square (ALS) approach based on Tucker decomposition. However, the fMRI dataset used in [22] was transformed into a 3D representation for data completion purposes. In [23] and [24], authors formulated the task of sampling and reconstruction of fMRI data as a 3D tensor completion problem. Hence, the advantages of the four-dimensional (4D) structure of fMRI data have not been widely exploited for the data completion purposes. The state-of-the-art tensor-based methods that focus on a higher-order fMRI data representation for brain imaging analyses include [25]–[27].

In [25], block term decomposition (BTD) is used to present a 4D analysis approach based on blind source separation (BSS) demonstrating the potential of such a representation. There has been considerable interest on tensor-based methods for analysis [25], [28] and functional connectivity estimation [29] that use higher-order models in the medical imaging context. However, less attention has been paid to the full utilization of all available spatial and temporal modes of fMRI data for the data completion task, which could directly highlight the importance of taking the full 4D information into account.

Recently, a nonconvex TT-rank minimization method for 4D fMRI tensor completion has been developed in [30] with promising results reported for the MAR voxels pattern. Similar to this work, we view 4D fMRI data as a sequence of 3D whole-brain volumes observed at a number of consecutive time points for data completion. Mathematically speaking, an fMRI dataset is a 4D tensor with 3D space \times time. Tensors capture correlation information across multiple dimensions via a high-order decomposition. Since fMRI data are inherently 4D, the 4-way representation allows us to naturally preserve the global structure of the data and maximize the use of both the spatial and temporal structures, i.e., the correlations, in the data. In addition, we make use of tensor train (TT) decomposition, which provides scalable numerical algebra operations and hierarchical compressed representations of structurally rich data sets; hence, it results in both reduced computational complexity and the efficient utilization of the covariance information.

In addition to not considering the time dimension, the majority of the algorithms for estimating missing data for medical imaging modalities were derived from schemes originally developed for Euclidean spaces. However, representing the geometry of medical images in Euclidean spaces has been shown to undermine the ability to account for joint variability either at the subject or group level [31], [32]. It was shown in [32] that brain images belong to a low-dimensional manifold that can provide a better approximation for the magnetic resonance imaging (MRI) data.

To overcome this problem, we propose to design a Riemannian framework for statistical learning in the context of fMRI tensor completion directly on a manifold, which is equipped with a Riemannian affine-invariant metric [33]. In [33] authors show that when tensors are endowed with the affine-invariant Riemannian distance, they provide a powerful framework for a robust interpolation and restoration of missing data by using generalized geometrical statistics. Therefore, the estimation error can be substantially reduced by learning tensor-valued fMRI data using Riemannian methods. As demonstrated in [32], [34]–[36], the use of the Riemannian methods yields a high discriminative power in fMRI/MRI data analyses. The Riemannian geometry offers a principled mathematical framework for applying standard optimization algorithms that work with Euclidean geometry. The proposed approach involves a projection of the quantities expressed in the Euclidean space onto the tangent space, which is equipped with the invariant information-preserving distance. Our study seeks to adapt the classical optimization algorithms into their Riemannian counterparts by substituting the Euclidean distance with the Riemannian distance and optimizing the cost function over the Riemannian space.

In recent years, much progress has been made in the research of various conjugate gradient methods that use Riemannian geometry [37], [38]. The extension of the conjugate gradient algorithm to Riemannian manifolds is accomplished by replacing the classical gradient with the natural

gradient [39]. This extension is relatively simple, as it is sufficient that all the vector operations take into account the Riemannian nature of the problem space. Yao *et al.* [40] proposed the Riemann Fletcher-Reeves conjugate gradient method to solve the stochastic eigenvalue problem, and Steinlechner [41] proposed the Riemannian nonlinear conjugate gradient scheme for the approximation of high-dimensional multivariate functions. To account for the non-linearity of the search space of the original data, in this paper, we propose a novel Riemannian approach for 4D tensor completion of fMRI data. We combine the strength of two ideas — natural 4D tensor representation of fMRI data and the model parametrization that best fits to the fMRI data. TT decomposition guarantees numerical stability, and the affine-invariant Riemannian distance allows volume-preserving transformations of the shape of the data. We show the validity of this approach and its desirable properties with fMRI data from the Center for Biomedical Research Excellence [42].

In addition to the successful demonstration of the great potential of considering the full 4D information in fMRI data, our contributions in this paper are summarized as follows:

- Viewing fMRI data as a 4D hypervolume allows us to take both temporal and spatial correlations into consideration, and therefore, allows us to establish a joint 4D learning scheme. The joint 4D learning scheme fully captures the intrinsic relationship between the spatial and temporal modes, thus guaranteeing superior performance compared to 2D and 3D fMRI completion methods in recovering missing voxels.
- We cast the problem of estimating missing fMRI data as a smooth optimization problem on Riemannian manifolds due to the nonlinear structure of the parameter space. To the best of our knowledge, this is the first time Riemannian geometry has been used for the estimation of missing fMRI data in the form of a 4D tensor in a practical setting.
- We develop a new geometric version of the spectral-scaling conjugate gradient method. Our method exploits second-order information and offers enhanced numerical stability.
- We extend our Riemannian optimization method in TT format, which allows efficient numerical algebra operations without exponential scaling of the memory consumption and the computational complexity with the number of dimensions [43]. It has been shown in [17] that TT decomposition captures the global correlation of the tensor entries, and thus, it is an efficient tool for tensor completion methods.
- We demonstrate that the proposed Riemannian nonlinear spectral conjugate gradient with TT (RSCG-TT) method provides a robust estimate of the original data for both the MAR and the MNAR missing value patterns encountered in real fMRI datasets and has performance superior to that of the state-of-the-art tensor completion methods across different rates of missing values. We show that the proposed method outperforms the state-of-the-art

gradient algorithms based on canonical polyadic decomposition (CPD) and Tucker decomposition formulated in Euclidean space.

The remainder of the paper is organized as follows. Section II-A introduces some notations and preliminaries for TT decomposition. Section B-A describes the Riemannian manifold learning framework with TT decomposition, followed by the formulation of the tensor completion problem on TT manifolds in Section II-B. In Section II-C, we propose a new spectral conjugate gradient scheme in conventional Euclidean space. In Section III, we present the algorithmic development of the proposed RSCG-TT method in full. We provide a description of the experimental setup in Section IV. The efficiency and robustness of the proposed algorithm on real fMRI data with missing data patterns that reflect simulated and real-world scenarios are demonstrated in Section V. We discuss the experimental results in Section VI. Conclusions and future work are given in Section VII.

II. THEORY

In this work, we consider TT tensors as our structured tensor format of interest [44], [45]. As noted in Section B-A2, the set of all such tensors of a fixed multilinear rank \mathbf{r} , which we equip with an affine invariant Riemannian metric [46], forms a smooth embedded submanifold of $\mathbb{R}^{I_1 \times \dots \times I_N}$ and were originally studied in [47]. With this differentiable Riemannian structure, we can construct optimization algorithms to solve the tensor completion problem for N -dimensional tensors. In Section II-B, we formulate high-dimensional data completion as a Riemannian tensor completion problem on TT manifolds. We provide the preliminaries on Riemannian manifold learning for tensor decomposition in TT format in Appendix B, and notations and definitions of the necessary ingredients used in Riemannian TT learning are summarized in Table 2.

To provide the reader with some context on TT decomposition, we provide the basic concepts and results in Section II-A. For an overview of the most common choices of tensor formats, we refer the reader to [48]. Although our presentation is self-contained, we recommend the reader to refer to the original content [44], [47], [49] for further clarification. The mathematical notations and definitions are adopted from [50] and [43], [51]. In this paper, we denote vectors with boldface lowercase letters ($\mathbf{x}, \mathbf{y}, \mathbf{z}, \dots$), matrices with boldface capital letters ($\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \dots$), and tensors with bold calligraphic letters ($\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \dots$); see Table 1.

A. TT DECOMPOSITION

A TT decomposition [44], shown in Fig. 1, represents an N th-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ defined as

$$\mathcal{X} = \mathcal{G}^{(1)} \times^1 \mathcal{G}^{(2)} \times^1 \dots \times^1 \mathcal{G}^{(N)}, \quad (1)$$

where $\mathcal{G}^{(n)} \in \mathbb{R}^{R_{n-1} \times I_n \times R_n}$ is the n th 3D TT core and $\mathcal{G}^{(1)} \in \mathbb{R}^{I_1 \times R_1}$, $\mathcal{G}^{(N)} \in \mathbb{R}^{R_{N-1} \times I_N}$ are matrices. The rank of TT

TABLE 1. Notation used throughout the paper.

Notation	Definition
\mathcal{X}	tensor ($\in \mathbb{R}^{I_1 I_2 \times \dots \times I_N}$)
N	dimension of \mathcal{X}
I_n	length of the n th mode of \mathcal{X}
$\mathbf{X}, \mathbf{x}, x$	matrix, vector, scalar
$\mathbf{X}_{\langle n \rangle} \in \mathbb{R}^{I_1 I_2 \dots I_n \times I_{n+1} \dots I_N}$	mode $(1, \dots, n)$ matricization of \mathcal{X}
$\mathbf{x} = \text{vec}(\mathcal{X}) \in \mathbb{R}^{I_1 I_2 \dots I_N}$	vectorization of \mathcal{X}
*	Hadamard (element-wise) product
\times^1	mode-(N,1) contracted tensor product
\odot, \otimes	Khatri-Rao and Kronecker products
$\ \cdot\ _F$	Frobenius norm of the argument
Ω	the sampling set ^a
$\bar{\Omega}$	the missing values set ^b
$\mathbf{r} = (R_0, R_1, \dots, R_N)$	TT rank
$\text{tr}(\cdot)$	trace of the argument
$\langle \cdot, \cdot \rangle$	inner product of the arguments

^b the set of indices of the observed entries of \mathcal{X}

^a the set of indices of unknown entries of \mathcal{X}

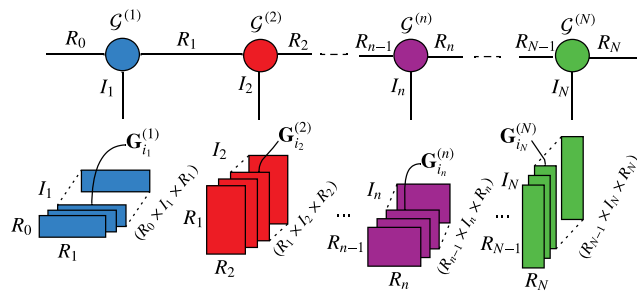


FIGURE 1. TT decomposition of an N th-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ with TT rank $\mathbf{r} = (R_0, R_1, R_2, \dots, R_n, \dots, R_{N-1}, R_N)$, where $R_0 = R_N = 1$. Conceptually, TT decomposition is a collection of TT cores connected in a cascade, where $\mathcal{G}^{(n)} \in \mathbb{R}^{R_{n-1} \times I_n \times R_n}$ are 3D tensors and $\mathcal{G}^{(1)} \in \mathbb{R}^{I_1 \times R_1}$ and $\mathcal{G}^{(N)} \in \mathbb{R}^{R_{N-1} \times I_N}$ are matrices [43].

decomposition (TT rank) [44] is an $N + 1$ tuple of ranks

$$\text{rank}_{\text{TT}}(\mathcal{X}) = \mathbf{r} = [R_0, R_1, R_2, \dots, R_{N-1}, R_N],$$

$$R_n = \text{rank}(\mathbf{X}_{\langle n \rangle}), \quad (2)$$

where R_n is the n th component of the TT rank, which controls the complexity of TT decomposition, and $R_0 = R_N = 1$. The rank R_n is the rank of matrix $\mathbf{X}_{\langle n \rangle} \in \mathbb{R}^{I_1 I_2 \dots I_n \times I_{n+1} \dots I_N}$ obtained by the n th splitting operator [44] of tensor \mathcal{X} as follows:

$$\mathbf{X}_{\langle n \rangle} = \left(\mathcal{X}, \prod_{i=1}^n I_i, \prod_{i=n+1}^N I_i \right). \quad (3)$$

Fig. 1 shows that the TT cores $\mathcal{G}^{(n)}$ are connected in a cascade, and the leaf nodes are identities. The entries of TT decomposition for an N th-order tensor can be computed as a multilayer multiplication of slices of TT cores, where the internal edges represent the TT rank [43]. The inner product of TT tensors can be defined in terms of the standard Euclidean product for vectors:

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \text{vec}(\mathcal{X})^T \text{vec}(\mathcal{Y}) = \text{tr}(\left(\mathbf{X}_{\langle n \rangle} \right)^T \mathbf{Y}_{\langle n \rangle}), \quad (4)$$

where $\text{vec}(\mathcal{X})$ is the vectorization operator and $n = 1, \dots, N$.

B. TENSOR COMPLETION ON TT MANIFOLDS

We consider the completion of an N -dimensional ($N \geq 3$) tensor $\mathcal{T} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ with partially observed entries defined in TT format. We assume the data are low-rank, and we present tensor completion as an optimization problem on a Riemannian manifold with a fixed TT rank as follows:

$$\min_{\mathcal{X}} f(\mathcal{X}) := \frac{1}{2} \|\text{P}_{\Omega} \mathcal{X} - \text{P}_{\Omega} \mathcal{T}\|_F^2$$

$$\text{subject to } \mathcal{X} \in \mathcal{M}_r := \{ \mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_N} \mid \text{rank}_{\text{TT}}(\mathcal{X}) = \mathbf{r} \}, \quad (5)$$

where P_{Ω} is the projection onto the sampling set $\Omega \subset \{1, 2, \dots, I_1\} \times \dots \times \{1, 2, \dots, I_N\}$, which defines the indices of the known entries of \mathcal{T} ,

$$\text{P}_{\Omega} \mathcal{X} := \begin{cases} \mathcal{X}(i_1, i_2, \dots, i_N) & \text{if } (i_1, i_2, \dots, i_N) \in \Omega \\ 0 & \text{if } (i_1, i_2, \dots, i_N) \notin \Omega. \end{cases} \quad (6)$$

We choose TT decomposition as our structured tensor representation to solve the problem in (5). The choice of the TT format is driven by the requirement to address tensor completion for very high-dimensional ($N \geq 4$) problems and the property of TT decomposition, which captures the global correlation of the tensor entries [17] due to its well-balanced matricization scheme. The TT format allows us to develop an algorithm that scales linearly in the number of dimensions N when the tensor size is I and polynomially in the TT rank \mathbf{r} , yielding a total asymptotic complexity of $N|\Omega|\mathbf{r}^2$. We exploit the Riemannian structure of TT manifold [47] by viewing (5) as an unconstrained optimization problem on \mathcal{M}_r , which allows for the use of Riemannian optimization techniques [37], [49].

The Euclidean gradient for problem (5) with respect to (w.r.t.) \mathcal{X} is given by

$$\nabla f_{\mathcal{X}}(\mathcal{X}) = \text{P}_{\Omega} \mathcal{X} - \text{P}_{\Omega} \mathcal{T}, \quad (7)$$

where $f_{\mathcal{X}}(\mathcal{X}) \in \mathbb{R}^{I_1 \times \dots \times I_N}$, and the Riemannian gradient is obtained by projecting (7) onto the tangent space $T_{\mathcal{X}} \mathcal{M}_r$:

$$\text{grad} f(\mathcal{X}) = \text{P}_{T_{\mathcal{X}} \mathcal{M}_r} \nabla f_{\mathcal{X}}(\mathcal{X})$$

$$= \text{P}_{T_{\mathcal{X}} \mathcal{M}_r} (\text{P}_{\Omega} \mathcal{X} - \text{P}_{\Omega} \mathcal{T}). \quad (8)$$

C. A NONLINEAR SPECTRAL CONJUGATE GRADIENT (SCG) SCHEME

To solve the tensor completion problem in (5), we use a nonlinear SCG scheme as our method of interest due its global superlinear convergence property, low memory requirements, and suitability for solving large-scale unconstrained optimization problems.

We develop a nonlinear SCG procedure that seeks the conjugate gradient direction \mathcal{N}_k closest to the direction of the double parameter self-scaled preconditioned Broyden-Fletcher-Goldfarb-Shanno (BFGS) update. To improve the

TABLE 2. Notations and symbols used in Riemannian TT learning.

Symbol	Definition	Computation
γ	geodesic curve	
\mathfrak{g}	Riemannian metric	
\mathcal{M}_r	Riemannian manifold of fixed TT rank r equipped with metric \mathfrak{g}	
ξ	tangent vector at point \mathcal{X}	
$T_{\mathcal{X}}\mathcal{M}_r$	tangent space of \mathcal{M}_r at point $\mathcal{X} \in \mathcal{M}_r$	
$P_{T_{\mathcal{X}}\mathcal{M}_r}(\cdot)$	the orthogonal projection of the argument onto $T_{\mathcal{X}}\mathcal{M}_r$	
$\text{grad}f(\mathcal{X})$	Riemannian gradient at $\mathcal{X} \in \mathcal{M}_r$, $\text{grad}f(\mathcal{X}) \equiv \xi \in T_{\mathcal{X}}\mathcal{M}_r$	$\text{grad}f(\mathcal{X}) = P_{T_{\mathcal{X}}\mathcal{M}_r}(\nabla f_{\mathcal{X}}(\mathcal{X}))$
$\tau_{\mathcal{X} \rightarrow \mathcal{Y}}$	vector transport of $\xi \in T_{\mathcal{X}}\mathcal{M}_r$, $\tau_{\mathcal{X} \rightarrow \mathcal{Y}} : T_{\mathcal{X}}\mathcal{M}_r \rightarrow T_{\mathcal{Y}}\mathcal{M}_r$	$\tau_{\mathcal{X} \rightarrow \mathcal{Y}}(\xi) = P_{T_{\mathcal{Y}}\mathcal{M}_r}(\xi)$
$P_r^{\text{TT-SVD}}(\cdot)$	projection operator by TT rank truncation	$P_r^{\text{TT-SVD}}(\cdot) = \text{TT-SVD}_r(\cdot)$
$R_{\mathcal{X}}(\xi)$	retraction of $\xi \in T_{\mathcal{X}}\mathcal{M}_r$, $R_{\mathcal{X}}(\xi) : T_{\mathcal{X}}\mathcal{M}_r \rightarrow \mathcal{M}_r$	$R_{\mathcal{X}}(\xi) = P_r^{\text{TT-SVD}}(\mathcal{X} + \xi)$

numerical stability of our method, we use the spectral-scaling secant condition [52] while deriving the search direction. The choice of the values for the spectral parameters in the search direction are determined to minimize the spectral condition number of the BFGS approximation to the inverse Hessian of the objective function. To achieve global convergence for general functions, we use a modified BFGS update (MBFGS) [53], which possesses the global convergence property without the convexity assumption on the objective function. Hence, by incorporating the spectral-scaling preconditioner in the MBFGS update, we simultaneously use second-order information and improve the numerical stability of our method.

Next, we introduce the key ingredients of our nonlinear SCG method in Euclidean space and provide a definition of our nonlinear SCG method in Riemannian space in Section III.

Starting from the initial guess $\mathcal{X}_0 \in \mathbb{R}^{I_1 \times \dots \times I_N}$, the nonlinear conjugate gradient method generates a sequence \mathcal{X}_k as follows:

$$\mathcal{X}_{k+1} = \mathcal{X}_k + \alpha_k \mathcal{N}_k, \quad k = 0, 1, \dots, \quad (9)$$

where $\alpha_k > 0$ is the step length determined by a suitable line search strategy. The spectral search direction \mathcal{N}_k [54] is given by

$$\mathcal{N}_k = -\theta_k \mathcal{G}_k + \beta_k \mathcal{N}_{k-1}, \quad \mathcal{N}_0 = -\theta_0 \mathcal{G}_0, \quad (10)$$

where θ_k and β_k are spectral and conjugate gradient parameters, respectively, and \mathcal{G}_k denotes the Euclidean gradient $\nabla f(\mathcal{X}_k)$. We denote $\mathcal{S}_{k-1} = \mathcal{X}_k - \mathcal{X}_{k-1} = \alpha_{k-1} \mathcal{N}_{k-1}$ and $\mathcal{Y}_{k-1} = \mathcal{G}_k - \mathcal{G}_{k-1}$.

A well-known drawback of the spectral conjugate gradient methods with the search direction (10) is that they cannot always satisfy the sufficient descent conditions [55], but they perform very well in practice for solving large-scale unconstrained optimization problems. Here, the sufficient descent condition means that there exists a positive constant c such that

$$\mathcal{G}_k^T \mathcal{N}_k \leq -c \|\mathcal{G}_k\|^2 \quad \forall k \geq 0, \quad (11)$$

where c is a positive constant. The condition in (11) is crucially important in establishing the global convergence of

our method. To take advantage of the attractive properties of the spectral gradient methods and enforce the generation of the search directions that satisfy (11), we require that the step length α_k satisfies the improved Wolfe conditions [56], which ensure that the condition (11) always holds. To support this requirement, we employ a nonmonotone line search strategy, which produces the step length α_k to satisfy the improved Wolfe conditions [56]. In Section II-C1, we derive the search direction (10) in Euclidean space and propose a SCG method based on the revised MBFGS update [53]. In Section III, we generalize the proposed SCG method to the Riemannian space and present our RSCG-TT method using the Riemannian optimization scheme on the TT manifold.

1) DERIVATION OF THE SPECTRAL SEARCH DIRECTION

It was shown in [57], [58] that the BFGS variable metric method can be regarded as a conjugate gradient algorithm by incorporating the BFGS updating scheme of the inverse Hessian approximation within the framework of a memoryless quasi-Newton method. We propose a new efficient spectral conjugate gradient algorithm based on the MBFGS update [53]. We require that the search direction (10) is close to a quasi-Newton direction; that is,

$$\mathcal{N}_k^{SGC} \approx -\mathcal{H}_k \mathcal{G}_k, \quad (12)$$

where \mathcal{H}_k is the symmetric approximation to the inverse Hessian $\nabla^2 f^{-1}(\mathcal{X}_k)$, expressed as a memoryless BFGS update [59]. During each iteration, the quantity $\mathcal{H}_k \approx \nabla^2 f^{-1}(\mathcal{X}_k)$ is replaced with a diagonal preconditioner $\theta_k \mathcal{I}$ such that the spectral-scaling secant condition [52] is satisfied:

$$\mathcal{H}_k \mathcal{Z}_{k-1} = \mathcal{S}_{k-1}, \quad \theta_k \mathcal{I} \mathcal{Z}_{k-1} = \mathcal{S}_{k-1}, \quad (13)$$

$$\mathcal{Z}_{k-1} = \mathcal{Y}_{k-1} + h_{k-1} \|\mathcal{G}_{k-1}\|^q \mathcal{S}_{k-1}, \quad (14)$$

$$h_{k-1} = p + \max \left\{ \frac{-\mathcal{S}_{k-1}^T \mathcal{Y}_{k-1}}{\|\mathcal{S}_{k-1}\|^2}, 0 \right\} \|\mathcal{G}_{k-1}\|^{-q}, \quad (15)$$

where $q > 0$, $p > 0$ and the related variables are defined after (10). The spectral-scaling secant condition (13) takes into account the second-order information of the objective function, and by extension, it incorporates the proven properties of the MBFGS update (14)–(15) [53]. Furthermore, since we require our line search strategy to fulfill the improved

Wolfe conditions [56], the curvature condition holds; that is, $\mathbf{S}_{k-1}^T \mathbf{y}_{k-1} > 0$. As a result, the expression for \mathbf{z}_{k-1} in (14) can be simplified as follows:

$$\mathbf{z}_{k-1} = \mathbf{y}_{k-1} + p \|\mathbf{g}_{k-1}\|^q \mathbf{s}_{k-1}. \quad (16)$$

To further enhance the numerical stability of our method, we propose to employ a double parameter variable BFGS metric update [60], which is given by:

$$\begin{aligned} \mathcal{H}_k^{\theta, \tau} = & \theta_k \mathcal{I} - \theta_k \frac{\mathbf{z}_{k-1} \mathbf{S}_{k-1}^T + \mathbf{s}_{k-1} \mathbf{z}_{k-1}^T}{\mathbf{z}_{k-1}^T \mathbf{s}_{k-1}} \\ & + \left(1 + \tau_k \frac{\mathbf{z}_{k-1}^T \mathbf{z}_{k-1}}{\mathbf{z}_{k-1}^T \mathbf{s}_{k-1}}\right) \frac{\mathbf{s}_{k-1} \mathbf{S}_{k-1}^T}{\mathbf{z}_{k-1}^T \mathbf{s}_{k-1}}, \end{aligned} \quad (17)$$

where $\tau_k > 0$ is set to satisfy the bound $\tau_k = \gamma \theta_k$, $1 \leq \gamma < 2$. Alternatively, (17) can be written as

$$\mathcal{H}_k^{\theta, \tau} = \mathcal{H}_k^\theta + (\tau_k - \theta_k) \frac{\|\mathbf{z}_{k-1}\|^2}{(\mathbf{z}_{k-1}^T \mathbf{s}_{k-1})^2} \mathbf{s}_{k-1} \mathbf{S}_{k-1}^T. \quad (18)$$

It is easy to see that if \mathcal{H}_k^θ is symmetric positive definite, the choice of $\theta_k > 0$, $\tau_k \geq \theta_k$ results in the sufficient descent condition for the search direction (12). The choice of the spectral parameter $\theta_k > 0$ is driven by the requirement to minimize the condition number of $\nabla^2 f^{-1}(\mathcal{X}_k)$, which can enhance the numerical stability of the tensor-based computational process. Therefore, the quantity $\theta_k \mathcal{I}$ should be an approximation of $\nabla^2 f^{-1}(\mathcal{X}_k)$. We obtain parameter θ_k as suggested in [61], [62] by minimizing $\|\theta_k \mathbf{z}_{k-1} - \mathbf{s}_{k-1}\|_F$, which yields

$$\theta_k^* = \frac{\mathbf{s}_{k-1}^T \mathbf{s}_{k-1}}{\mathbf{z}_{k-1}^T \mathbf{s}_{k-1}}. \quad (19)$$

By its derivation, (19) minimizes the spectral condition number of (17) and thereby improves the numerical stability of our method. The optimal upper bound of the parameters θ_k and τ_k was derived in [60], [63], and the parameters θ_k and τ_k are computed by

$$\theta_k = \frac{1}{2 - \gamma} \frac{\mathbf{s}_{k-1}^T \mathbf{s}_{k-1}}{\mathbf{z}_{k-1}^T \mathbf{s}_{k-1}}, \quad \tau_k = \gamma \theta_k, \quad 1 \leq \gamma < 2. \quad (20)$$

To ensure the sufficient descent condition for the search direction (12), we restrict θ_k as follows:

$$\tilde{\theta}_k = \max\{\min\{\theta_k, \frac{1}{m_1}\}, m_1\}, \quad \tilde{\tau}_k = \gamma \tilde{\theta}_k, \quad \leq \gamma < 2, \quad (21)$$

where m_1 is a small positive constant.

By multiplying (17) by the gradient \mathbf{g}_k and after some algebraic transformations, our spectral search direction (12) is given by

$$\mathcal{N}_k^{SCG} = -\tilde{\theta}_k \mathbf{g}_k + \beta_k \mathbf{s}_{k-1} + \zeta_k \mathbf{z}_{k-1}, \quad (22)$$

where the spectral parameter $\tilde{\theta}_k$ is given in (21) and the conjugate gradient parameter β_k is computed as follows:

$$\beta_k(\tilde{\theta}_k, \tilde{\tau}_k) = \tilde{\theta}_k \frac{\mathbf{g}_k^T \mathbf{z}_{k-1}}{\mathbf{z}_{k-1}^T \mathbf{s}_{k-1}} \quad (23)$$

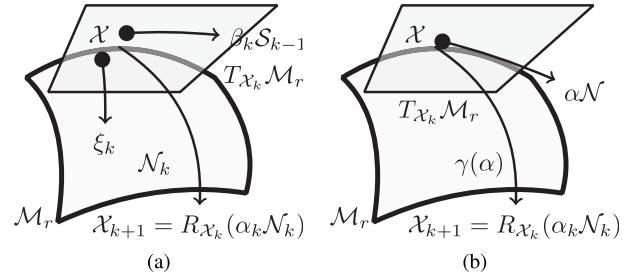


FIGURE 2. SCG on a Riemannian TT manifold. (a) Riemannian SCG iteration. (b) Minimizing along the search direction \mathcal{N} on \mathcal{M} . The step length α is obtained by minimizing the tangent line passing through $\mathcal{X} \in \mathcal{M}_r$ with direction $\mathcal{N}_k \in T_{\mathcal{X}_k} \mathcal{M}_r$. We minimize along the search curve to obtain an optimal estimate of the step length α .

$$\begin{aligned} & - \left(1 + \tilde{\tau}_k \frac{\mathbf{z}_{k-1}^T \mathbf{z}_{k-1}}{\mathbf{z}_{k-1}^T \mathbf{s}_{k-1}}\right) \frac{\mathbf{g}_k^T \mathbf{s}_{k-1}}{\mathbf{z}_{k-1}^T \mathbf{s}_{k-1}}, \\ \zeta = & \tilde{\theta}_k \left(\frac{\mathbf{g}_k^T \mathbf{s}_{k-1}}{\mathbf{z}_{k-1}^T \mathbf{s}_{k-1}} \right). \end{aligned} \quad (24)$$

Here, we discuss the global convergence properties of the SCG method inherited from the MBFGS update. Notably, the MBFGS update formula in (14) has a favorable property that for each k , it always holds that $\mathbf{z}_{k-1}^T \mathbf{s}_{k-1} > 0$, which ensures that \mathcal{H}_k inherits the positive definiteness of \mathcal{H}_{k-1} [64] as long as we start with the initial positive definite quantity $\mathcal{H}_0 = \theta_0 \mathcal{I}$. It was shown in [53] [Theorem 2.1], [64] [Theorem 3.3] that if the update for \mathcal{H}_k is given by (13)–(15), then the global convergence property holds. Therefore, as previously argued [53], [64] the SCG method is seen to possess the global convergence property. For more details regarding the MBFGS method, we refer the reader to the original articles [53], [64], and [65].

III. A RIEMANNIAN SCG METHOD FOR TENSOR COMPLETION

This section is devoted to the algorithmic development for solving the tensor completion problem (5).

A. RSCG-TT ITERATIVE ALGORITHM

With the concepts introduced in Sections II-B and II-C we have all the requirements for performing Riemannian optimization on the TT manifold \mathcal{M}_r . The SCG algorithmic framework developed in Section II-C yields Algorithm 1. It can be seen as an extension of SCG scheme introduced in Section II-C, with the Euclidean gradient replaced by the Riemannian gradient. Here, we generalize the quantities we derived in Section II-C to Riemannian settings. When the Euclidean space $\mathbb{R}^{I_1 \times \dots \times I_N}$ is replaced by the Riemannian TT manifold $(\mathcal{M}_r, \mathfrak{g})$, the components in spectral search direction (22) belong to different tangent spaces; hence, the Euclidean vector addition on $(\mathcal{M}_r, \mathfrak{g})$ makes no sense. To modify the vector addition into a suitable operation, we use the vector transport $\tau_{\mathcal{X}_{k-1} \rightarrow \mathcal{X}_k} := P_{T_{\mathcal{X}_k} \mathcal{M}_r}(\mathcal{X}_{k-1})$ (see Table 2). Thus, the Riemannian SCG update rule is as follows:

$$\xi_k = \text{grad}f(\mathcal{X}_k), \quad \text{defined in (8)} \quad (25)$$

Algorithm 1 RSCG-TT Algorithm

Require: The observed tensor $\mathcal{T} \in \mathcal{M}_r$, sampling set Ω , multilinear TT rank \mathbf{r} , prescribed tolerance ϵ , and positive parameters p, m_1, γ, q .

Ensure: The completed tensor \mathcal{X} is an approximation of \mathcal{T} .

- 1: **Initialize:** $\mathcal{X}_0 \in \mathcal{M}_r$ randomly
- 2: Compute the initial gradient
 $\xi_0 \leftarrow \text{grad}f(\mathcal{X})$
- 3: Compute the initial search direction as the steepest descent
 $\theta_0 \leftarrow 1, \mathcal{N}_0 \leftarrow -\theta_0 \xi_0$
- 4: Compute the initial step size by exact minimizer (37)
 $\alpha_0 \leftarrow \min_{\alpha} f(\mathcal{X}_0) + \alpha \mathcal{N}_0$
- 5: Obtain the next iterate by retraction (33)
 $\mathcal{X}_1 \leftarrow R_{\mathcal{X}_0}(\alpha_0 \mathcal{N}_0)$
- 6: **while** $\frac{\|\mathcal{X}_{\Omega} - \mathcal{T}_{\Omega}\|_F^2}{\|\mathcal{T}_{\Omega}\|_F^2} \leq \epsilon$ **do**
- 7: Compute gradient (25)
 $\xi_k \leftarrow \text{grad}f(\mathcal{X}_k)$
- 8: Compute the spectral search direction (32)
 $\mathcal{N}_k^{SCG} \leftarrow -\tilde{\theta}_k \xi_k + \beta_k \mathcal{S}_{k-1} + \zeta_k \mathcal{Z}_{k-1}$
- 9: Compute the step length by updating rules (30) and (31)
 $\alpha_k \leftarrow \alpha_k^*$
- 10: Compute the next iterate by retraction (33)
 $\mathcal{X}_{k+1} \leftarrow R_{\mathcal{X}_k}(\alpha_k \mathcal{N}_k)$
- 11: **end while**

$$\mathcal{S}_{k-1} = \tau_{\mathcal{X}_{k-1} \rightarrow \mathcal{X}_k}(\alpha_{k-1} \mathcal{N}_{k-1}), \tag{26}$$

$$\xi_{k-1}^{\text{transp}} = \tau_{\mathcal{X}_{k-1} \rightarrow \mathcal{X}_k}(\xi_{k-1}), \tag{27}$$

$$\mathcal{Y}_{k-1} = \xi_k - \xi_{k-1}^{\text{transp}}, \tag{28}$$

$$\mathcal{Z}_{k-1} = \mathcal{Y}_{k-1} + p \|\xi_{k-1}^{\text{transp}}\|^q \mathcal{S}_{k-1}, \tag{29}$$

$$\beta_k(\tilde{\theta}_k, \tilde{\tau}_k) = \tilde{\theta}_k \frac{\xi_k^T \mathcal{Z}_{k-1}}{\mathcal{Z}_{k-1}^T \mathcal{S}_{k-1}} - \left(1 + \tilde{\tau}_k \frac{\mathcal{Z}_{k-1}^T \mathcal{Z}_{k-1}}{\mathcal{Z}_{k-1}^T \mathcal{S}_{k-1}}\right) \frac{\xi_k^T \mathcal{S}_{k-1}}{\mathcal{Z}_{k-1}^T \mathcal{S}_{k-1}},$$

$$\zeta = \tilde{\theta}_k \left(\frac{\xi_k^T \mathcal{S}_{k-1}}{\mathcal{Z}_{k-1}^T \mathcal{S}_{k-1}} \right),$$

$$\alpha_k^0 = \min_{\alpha} f(\mathcal{X}_k) + \alpha \mathcal{N}_k \quad \text{defined in (37)} \tag{30}$$

$$\alpha_k = \alpha_k^* \quad \text{defined in (38) and (39)}, \tag{31}$$

$$\mathcal{N}_k^{SCG} = -\tilde{\theta}_k \xi_k + \beta_k \mathcal{S}_{k-1} + \zeta_k \mathcal{Z}_{k-1}, \tag{32}$$

$$\mathcal{X}_{k+1} = R_{\mathcal{X}_k}(\alpha_k \mathcal{N}_k). \tag{33}$$

The resulting optimization scheme is given in Algorithm 1. We call it the Riemannian SCG algorithm for tensor completion via TT. Algorithm 1 represents a geometrical version of the nonlinear SCG algorithm proposed in Section II-C. The major difference is that we have to ensure that every point \mathcal{X}_k in the optimization sequence stays on the Riemannian manifold \mathcal{M}_r . To move in the direction of the steepest ascent

of $f(\mathcal{X}_k)$, we use the Riemannian gradient $\xi_k \in T_{\mathcal{X}_k} \mathcal{M}_r$, which points in the direction of the greatest increase within tangent space $T_{\mathcal{X}_k} \mathcal{M}_r$. The geometrical version of the algorithm is visualized in Fig. 2a and consists of the following core concepts.

In the first step, we initialize tensor \mathcal{X}_0 with random entries and move in the direction of the negative Riemannian gradient $\mathcal{N}_0 = -\theta_0 \xi_0$. This produces a new iterate that is not an element of \mathcal{M}_r . We bring the new iterate $\mathcal{X}_0 + \alpha_0 \mathcal{N}_0$ back to manifold \mathcal{M}_r using retraction (33). The spectral search direction $\mathcal{N}_k \in T_{\mathcal{X}_k} \mathcal{M}_r$ is computed by using the updating rule in (32). It requires taking a linear combination of the Riemannian gradient ξ_k and quantities \mathcal{S}_{k-1} and \mathcal{Z}_{k-1} , which belong to a different tangent space $T_{\mathcal{X}_{k-1}} \mathcal{M}_r$. We map \mathcal{S}_{k-1} and \mathcal{Z}_{k-1} to the current tangent space with the vector transport $\tau_{\mathcal{X}_{k-1} \rightarrow \mathcal{X}_k}$. To generate the next iteration in the direction of sufficient descent, we compute the step length α_k using a nonmonotone line search strategy based on the improved Wolfe condition proposed in [56]. Next, with the spectral conjugate direction \mathcal{N}_k , we perform a step along the search curve γ and obtain the next iteration by retraction (33) so that $\mathcal{X}_k \in \mathcal{M}_r$. Evaluating cost function (5) reduces the computation of the entries of the sparse tensor $P_{\Omega} \mathcal{X}$ since the observed tensor $P_{\Omega} \mathcal{T}$ is provided as input. The convergence condition is reached when the relative error between observed tensor $P_{\Omega} \mathcal{T}$ and its approximation $P_{\Omega} \mathcal{X}$ is smaller than a prescribed tolerance ϵ . We find the step length α , which is a solution to the cost function, defined on the geodesic search curve $\gamma(\alpha) := R_{\mathcal{X}_k}(\alpha \mathcal{N}_k)$, where $R_{\mathcal{X}_k}(\alpha \mathcal{N}_k)$ is the retraction of $\alpha \mathcal{N}_k$ at point \mathcal{X}_k . The cost function for the step length α is given as

$$\phi(\alpha) := f(\gamma(\alpha)) := f(R_{\mathcal{X}_k}(\alpha \mathcal{N}_k)). \tag{34}$$

We determine the step length α_* such that the value $\phi(\alpha_*)$ provides the sufficient reduction in cost function (5) and generates a descent direction such that $\mathcal{N}_k^T \xi_k < 0$. The one-dimensional minimizer for (34) given in [37]

$$\alpha_k^* := \text{argmin}_{\alpha > 0} \phi(\alpha) := \text{argmin}_{\alpha > 0} f(R_{\mathcal{X}_k}(\alpha \mathcal{N}_k)) \tag{35}$$

We minimize along the search curve to obtain the optimal value of the step length. The concept is illustrated in Fig. 2b. Our tensor completion problem (5) is a smooth quadratic function. Therefore, a good initial guess for the step length can be obtained with exact minimization of the tangent space without retraction.

$$\min_{\alpha} f(\mathcal{X} + \alpha \mathcal{N}) = \frac{1}{2} \min_{\alpha} \|P_{\Omega} \mathcal{X} + \alpha P_{\Omega} \mathcal{N} - P_{\Omega} \mathcal{T}\|_F^2 \tag{36}$$

The closed-form solution of (36) is derived by setting the α -derivative to 0 and is given by

$$\alpha_* = \frac{\langle P_{\Omega} \mathcal{N}, P_{\Omega}(\mathcal{X} - \mathcal{T}) \rangle}{\langle P_{\Omega} \mathcal{N}, P_{\Omega} \mathcal{N} \rangle}. \tag{37}$$

To guarantee the global convergence of our RSCG-TT method, we use the nonmonotone line search strategy based

on the improved Wolfe condition [56] with a combination of curvature conditions as follows:

$$\phi(\alpha) \leq \phi_k(0) + \min\{\epsilon_c |\phi_k(0)|, \delta \alpha \phi_k(0) + l_k\} \quad (38)$$

$$\phi'_k(\alpha) \geq \sigma \phi'_k(0), \quad (39)$$

$$l_k = \frac{1}{k^2} \quad (40)$$

where $\epsilon_c > 0$, $0 < \delta < \sigma < 1$, l_k is a positive term so that a positive summable sequence $\{l_k\}$ satisfies condition $\sum_{k \geq 1} l_k < +\infty$. As noted by [56], the positive term in (40) allows slight increase in the function value, and it helps to avoid the numerical drawback of the standard Wolfe condition [66] due to the existence of the numerical round-off errors or when the tolerance error is very low [56].

B. COMPLEXITY ANALYSIS OF THE RSCG-TT METHOD

We analyze the time complexity of RSCG-TT for one iteration in Algorithm 1. The computational complexity of RSCG-TT is influenced by the following computations: the Riemannian gradient $\text{grad}f(\mathcal{X}_k)$ in (8), the search direction \mathcal{N}_k in (32), the step-length α_k in (37), and the next iteration \mathcal{X}_{k+1} obtained by the retraction $R_{\mathcal{X}_k}(\alpha_k \mathcal{N}_k)$ in (33). The computational complexities of TT decomposition and the Riemannian operators are given in [44] and [67], respectively. For simplicity, we assume $I = \max\{I_1, I_2, \dots, I_N\}$ and the upper bound on TT rank is $R_{\max} = \max\{R_0, R_1, \dots, R_N\}$. The Riemannian gradient $\text{grad}f(\mathcal{X}_k)$ is computed as a projection of the sparse Euclidean gradient in (7) onto the tangent space of \mathcal{X} , which results in a cost of $\mathcal{O}(NIR_{\max}^3 + N|\Omega|R_{\max}^2)$. The estimates for the search direction \mathcal{N}_k and the next iteration \mathcal{X}_{k+1} can be calculated in $\mathcal{O}(NIR_{\max}^3)$, and the cost for the step length α_k is $|\Omega|(N-1)R_{\max}^2$. Hence, for one iteration of RSCG-TT, the total computational cost is dominated by the computation of the Riemannian gradient and is given by

$$\mathcal{O}(NIR_{\max}^3 + N|\Omega|R_{\max}^2). \quad (41)$$

IV. NUMERICAL EXPERIMENTS

We assess the numerical performance of the RSCG-TT method with real fMRI data from the Center of Biomedical Research Excellence (COBRE), which is available from the COllaborative Informatics and Neuroimaging Suite (COINS) data exchange repository [68].¹ We compare RSCG-TT with two other tensor completion stochastic gradient descent (SGD) methods, namely, Tucker-SGD and CPD-SGD.

A. IMAGE ACQUISITION AND PROCESSING

We employed data from the COBRE study of the Mind Research Network (MRN). The rs-fMRI data consists of 149 volumes of T2*-weighted functional images each, acquired through a gradient-echo EPI sequence: TR = 2 s, TE = 29 ms, flip angle = 75°, number of axial slices = 33 in

¹COINS data exchange repository <https://coins.trendscenter.org/#exchange>

sequential ascending order, slice thickness = 3.5 mm, slice gap = 1.05 mm, field of view = 240 mm, matrix size = 64 × 64 and voxel size = 3.75 mm × 3.75 mm × 4.55 mm. To remove the T1 equilibration effects, the first four volumes were discarded. Slice-timing correction was used to realign images. The data were spatially normalized to the standard Montreal Neurological Institute space, resampled to 3 mm × 3 mm × 3 mm voxels, and smoothed using a Gaussian kernel with a full width at a half maximum of 5 mm.

B. EXECUTION DETAILS

All experiments were performed with Python in TensorFlow on a Linux workstation with 4 Quad-Core Intel Xeon 3.1 GHz processors and 16 GB RAM. For all experiments, the input is a tensor $P_{\Omega}\mathcal{T}$ projected onto the sampling set Ω that reflects the positions of the observed entries, where \mathcal{T} is the fully observed true tensor, and $*$ is the Hadamard (element-wise) product. The completed tensor $\hat{\mathcal{X}}$ is computed as

$$\hat{\mathcal{X}} = (\mathcal{I} - P_{\Omega}\mathcal{I}) * \mathcal{X} + P_{\Omega}\mathcal{T}, \quad (42)$$

where \mathcal{I} is the tensor of all ones the same size as \mathcal{T} , and \mathcal{X} is the output tensor estimated by each algorithm. RSCG-TT (Algorithm 1) was implemented using TT decomposition in the TensorFlow (T3F) [67] toolbox with its inbuilt operations to manipulate the tensor in TT format as well as its Riemannian operations. To find a step length α , we used the nonmonotone Wolfe line-search algorithm in [56].

The following parameters were used in the implementation of the proposed method:

$\epsilon = 10^{-8}$, $p = 0.001$, $m_1 = 10^{-8}$, $\gamma = 1.2$, $q = 3$, $\delta = 0.0001$, and $\sigma = 0.9$. Here, the values for p , m_1 , γ , q are the values recommended in [60], and the values for δ , σ were selected based on the typical values used in the nonmonotone Wolfe line-search algorithms suggested in [55], [56], [69]. We further fine-tuned the selection of the optimization hyperparameters based on the recommended ranges in the literature using a grid search with cross validation on the validation set Ω_{Γ} . We created a training set based on the sampling set Ω by randomly selecting the indices i_k for each tensor dimension I_k from a uniform distribution on $\{1, \dots, I_k\}$. We generated a test set Ω_{Γ} and a validation set Ω_{Γ} with random sampling from the missing value set $\bar{\Omega}$, where 80% of the set was allocated to the test set and 20% was allocated to the validation set. The stopping criteria for RSCG-TT were set as follows: reaching a maximum 500 iterations or achieving the convergence tolerance $\epsilon \leq 10^{-8}$.

C. EXPERIMENTAL DESIGN

Our goal is to study the performance of the proposed method in terms of its ability to recover missing voxels with different missing patterns and proportions in real fMRI data. We conduct hybrid experiments in different tensor dimensions and study the effectiveness of exploiting the multidimensional structure through 3D and 4D representations of fMRI data in terms of reconstruction errors. In addition, we compare

TABLE 3. fMRI data completion design factors.

Factors	Levels
Pattern of missing values	RMV ¹ , SMV ²
Percentage of missing values ³	10, 20, 30, 40, 50, 60, 70, 80, 90
Tensor dimension	2D, 3D, 4D

¹ Random missing values.

² Structural missing values.

³ SMV pattern experiments conducted with 5, 10, 15, and 20 percent missing values.

the numerical performance and convergence rate of the proposed algorithm with the state-of-the-art SGD optimization [70], [71] for tensor completion in the form of Tucker and CPD decompositions. The experiments were carried out considering the percentage and pattern of missing values and tensor dimensionality. The design conditions are summarized in Table 3. We executed 50 runs for each setup and reported the averages.

D. APPLICATION TO REAL fMRI DATA

As we stated in Section II-B, tensor completion performs best on low-rank datasets. We exploit the low-rank property of fMRI data [19] and cast fMRI tensor completion as a specific case of the tensor completion problem on the TT manifold (5). TT decomposition applied on fMRI data captures the global correlation between spatial and temporal modes and thereby allows the optimal estimation of missing voxels. We evaluate our method on two types of missing value patterns. First, random missing values are simply used, and the other type is a more realistic structural missing values pattern based on 3D spatial ellipsoids. We briefly describe the background and the experimental setup of these test cases in Sections IV-F and IV-E–IV-G.

E. MAGNETIC SUSCEPTIBILITIES IN fMRI DATA

fMRI is inherently susceptible to geometric distortions caused by magnetic field inhomogeneity [8] given its relatively long total readout times. Furthermore, this data distortion is often significant in the phase encoding direction. The primary source of field inhomogeneity is differences in the susceptibilities of neighboring tissues [72]. Specifically, there are significant spatial variations at tissue borders/edges. The most vulnerable areas in brain MRI to geometric distortions are the frontal lobe proximal to the paranasal sinuses, as well as the temporal lobe proximal to the mastoid air cells and ear canals [72], [73]. Geometric distortion is often most severe in regions where air-filled sinuses border bone or tissue, including the frontal and temporal lobes. Consequently, these spatial variations produce rapid changes in magnetic field gradients. Different techniques for correcting fMRI image distortion have been reported [3], [72], [74]–[77]. Most of these approaches exclude brain regions with significant variations in the magnetic field gradients in the phase encoding direction or slice selection direction based on a field quality map. The field quality map is

computed based on the magnitude or phase derivative variance [3], [72], [74], [77]. The field quality maps are robust in identifying noisy areas in fMRI images and are used to address geometric distortions in fMRI studies. The acquired quality maps are used to develop binary quality masks to threshold the brain images. One such method is the quality map phase denoising (QPMD) method for fMRI data [3]; this method identifies the noisy regions based on the partial derivatives $\nabla_{x,y,z}$ of the phase image, where the high gradient values correspond to low signal quality, and therefore, the corresponding voxels are excluded from the analysis. This results in MNAR missing voxels in areas, including the sinuses, which are susceptible to gradient changes, and the ventral/medial prefrontal areas. This study proposed addressing this issue through 4D fMRI tensor completion, where the missing voxels are represented by a 3D brain mask.

F. RANDOM MISSING VALUE (RMV) PATTERN

This problem is identical to the one described in [7], [8] and corresponds to the MAR pattern [6] observed in fMRI studies. To simulate an RMV pattern, we generated a sampling set Ω w.r.t different missing value rates (MR) defined as

$$MR_{RMV} = \frac{m}{\prod_{k=1}^N I_k}, \quad (43)$$

where m is the number of missing entries. The indices of observed entries in the sampling set Ω are chosen by randomly sampling each dimension index I_n from a uniform distribution on $\{1, \dots, I_n\}$.

G. STRUCTURAL MISSING VALUES (SMV) PATTERN

This test case is equivalent to the MNAR fMRI pattern described in Section IV-E. To evaluate the proposed algorithm on a realistic MNAR test case, we generate analytical ellipsoids with missing voxels that satisfy the 3D ellipsoid equation

$$\frac{(x - x_0)^2}{r_x^2} + \frac{(y - y_0)^2}{r_y^2} + \frac{(z - z_0)^2}{r_z^2} = 1, \quad (44)$$

where (x_0, y_0, z_0) is the origin of the ellipsoid and (r_x, r_y, r_z) is the radius in the direction of each spatial dimension. A spatial location (x_0, y_0, z_0) was manually selected in a brain volume, and ellipsoid (44) with missing voxels was injected into this region. Furthermore, to simulate the case where several 3D fMRI brain volumes are impacted by spatial ellipsoid (44), we place the spatial ellipsoid at multiple randomly selected timepoints in the 4D fMRI brain scan. The number of affected timepoints is selected to simulate the percentage of missing values in the temporal sense. In addition, we study the effect of the size of the spatial ellipsoid to evaluate the estimation capabilities of the proposed method that are close to a real-world scenarios.

Next, we define the quantities that we use to measure the performance of algorithms completing the data affected by SMV patterns as follows:

$$V_{e3D} = (4/3)\pi r_x r_y r_z,_{(\text{voxels})^3}, \quad (45)$$

TABLE 4. Ellipsoid volumes.

Ellipsoid label	r_x, r_y, r_z	$V_{e^{3D}}$	$MR_{V_{e^{3D}}}, \%$
Size 1	(7,10,8)	2345	1.53
Size 2	(8,10,8)	2680	1.75
Size 3	(9,10,8)	3015	1.96
Size 4	(10,10,8)	3351	2.18
Size 5	(11,10,8)	3686	2.40
Size 6	(12,10,8)	4021	2.62

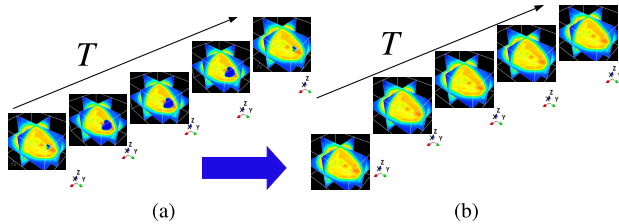


FIGURE 3. 4D fMRI tensor completion of missing voxels corrupted by the SMV pattern. (a) 4D fMRI hypervolume (3D voxels \times T) is corrupted by the spatial ellipsoids with missing voxels embedded in 3D brain volumes. (b) Completed 4D fMRI hypervolume (3D voxels \times T).

$$V_{e^{4D}} = (4/3)\pi r_x r_y r_z \times T,_{(\text{voxels})^3 \times \text{time}}, \quad (46)$$

$$V_{3D} = X \times Y \times Z,_{(\text{voxels})^3}. \quad (47)$$

where $V_{e^{3D}}$ and $V_{e^{4D}}$ are 3D spatial and 4D hypervolumes of ellipsoid (44), V_{3D} is the total volume of the fMRI brain scan, X, Y, Z are the numbers of voxels in each spatial dimension, and T is the number of timepoints comprising the fMRI brain scan. The metadata of the generated ellipsoids are given in Table 4. The missing value rate metrics for the SMV pattern are given as

$$MR_{V_{e^{3D}}} = \frac{V_{e^{3D}}}{V_{3D}}, \quad (48a)$$

$$MR_T = \frac{M_T}{T}, \quad (48b)$$

where $MR_{V_{e^{3D}}}$ is the spatial missing value rate measured as the ratio of the spatial volume of the ellipsoid with missing voxels in (45) to the total spatial volume in (47). MR_T is the temporal missing value rate across the entire 4D fMRI brain scan, where M_T is the number of timepoints affected by the artifact in (44). In the SMV experiment, we use the discrete values missing ratio set $MR_T(\%) \in \{5, 10, 15, 20\}\%$, and for each value of MR_T , we vary the volume of the ellipsoid artifact according to values in Table 4 and the tensor dimension values with levels as per Table 3. Conceptually, the 4D fMRI tensor completion process for the SMV pattern is visualized in Fig. 3. The 4D fMRI hypervolume is corrupted adding spatial ellipsoids with missing voxels into a 3D whole-brain volume at randomly selected timeframes. We feed the entire 4D fMRI tensor with corrupted 3D whole-brain volumes at timepoints T_i into the proposed algorithm, which completes the 4D tensor simultaneously for all corrupted 3D spatial brain volumes.

H. RANK ESTIMATION

The tensor completion problem in (5) requires that the rank of a high-dimensional dataset be known in advance. Specifically, for TT decomposition, we have to determine the multilinear TT rank in (2). The TT ranks R_n have an upper bound determined by a CPD approximation with tensor rank R and accuracy ϵ_{TT} , such that $R_n \leq R$ and TT decomposition can be computed with a relative accuracy of $\sqrt{N-1}\epsilon_{TT}$ [44], where N is the tensor dimension. We estimate the CPD-rank R and multilinear Tucker-rank with the aid of automatic relevance determination (ARD) [78] and the L-curve algorithm [79], [80]. The stopping criterion for the ARD algorithm is set to when the relative tolerance of the negative log-likelihood is below 10^{-8} , which results in the explained variance being in the range of 0.9998 – 0.9999. The ranks computed by the ARD algorithm are $R_{CPD_{4D}} = 81$, $R_{CPD_{3D}} = 46$ for CPD decomposition and $\mathbf{r}_{Tucker_{4D}} = (53, 63, 46, 91)$, $\mathbf{r}_{Tucker_{3D}} = (1630, 43, 144)$ for Tucker decomposition.

We use the computed CPD-rank R as the maximal rank constraint R_{max} and as the input to the TT-SVD algorithm [44], and the values of the TT rank are determined by a quasi-optimal TT approximation. The multilinear \mathbf{r}_{4D} and \mathbf{r}_{3D} TT ranks are estimated as a result of the TT-rounding algorithm with the imposed maximal rank constraints $R_{4D_{max}} = R_{CPD_{4D}} = 81$, $R_{3D_{max}} = R_{CPD_{3D}} = 46$, and $R_{2D_{max}} = 144$ for 4D, 3D, and 2D tensor representations accordingly. The maximal rank for the 2D representation is selected as a column rank of the reshaped 2D matrix of the input data tensor. The following TT rank values were estimated for each tensor dimensionality: $\mathbf{r}_{4D_{TT}} = (R_0, R_1, R_2, R_3, R_4) = (1, 81, 81, 81, 1)$, $\mathbf{r}_{3D_{TT}} = (R_0, R_1, R_2, R_3) = (1, 46, 46, 1)$, and $\mathbf{r}_{2D_{TT}} = (R_0, R_1, R_2) = (1, 144, 1)$. The ranks R_0 and R_N are restricted to one by the definition of TT decomposition, whereas ranks R_n are selected to be the minimum values of the column ranks of the unfolding matrices $\mathbf{X}_{\langle n \rangle} \in \mathbb{R}^{I_1 \times \dots \times I_n \times I_{n+1} \times \dots \times I_N}$ [44] and the maximal rank R_{max} .

I. QUANTITATIVE EVALUATION

We compare RSCG-TT with two other methods: Tucker-SGD and CPD-SGD. The compared algorithms employ the SGD optimization method called Adaptive Moment Estimation (Adam) [70], which is known for its prominent performance. The Adam's update is given as follows:

$$x_{k+1} = x_k - \frac{d}{\sqrt{v_k(f_k(x_k), \beta_2, v_{k-1}) + \epsilon}} m_k(f_k(x_k), \beta_1, m_{k-1}), \quad (49)$$

where d, ϵ, β_1 , and β_2 are hyperparameters and m_k and v_k are the first- and second-moment estimates of gradient $\nabla f_k(x_k)$. We selected the default values for the hyperparameters and learning rate $\alpha: \epsilon = 10^{-8}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\alpha = 0.001$, as suggested in [70]. All three comparison algorithms are implemented in the TensorLy [81] framework.

To measure the numerical performance of the algorithms, we report the relative square error (RSE) between the completed tensor $\hat{\mathcal{X}}$ and the true tensor \mathcal{T} , which is defined as

$$RSE = \frac{\|\hat{\mathcal{X}} - \mathcal{T}\|_F}{\|\mathcal{T}\|_F}. \quad (50)$$

However, to assess the reconstruction quality of the missing values, we use the tensor completion score (TCS) [14], which is given as

$$TCS = \frac{\|(\mathcal{I} - P_{\Omega}\mathcal{I}) * (\hat{\mathcal{X}} - \mathcal{T})\|_F}{\|(\mathcal{I} - P_{\Omega}\mathcal{I}) * \mathcal{T}\|_F}. \quad (51)$$

The RSE metric in (50) is used to evaluate the overall global numerical performance, whereas the TCS in (51) is used to assess the RSE in the missing values [14]. To accommodate the specific needs of fMRI data, we define a new threshold metric based on (51), which measures the recovery of missing voxels in the data regions after thresholding Z-scores by two times the standard deviation

$$TCS_Z = \frac{\|(\mathcal{I} - P_{\Omega}\mathcal{I}) * (\hat{\mathcal{X}}_Z - \mathcal{T}_Z)\|_F}{\|(\mathcal{I} - P_{\Omega}\mathcal{I}) * \mathcal{T}_Z\|_F}. \quad (52)$$

We detect the stagnation and stop the algorithms when a relative residual of the objective function in ϵ_{Ω} between the successive iterations drops below the prescribed tolerance ϵ :

$$\epsilon_{\Omega}(f(\mathcal{X}_k)) = \frac{|f(\mathcal{X}_k) - f(\mathcal{X}_{k-1})|}{|f(\mathcal{X}_{k-1})|} \leq \epsilon \quad (53)$$

where $\epsilon = 10^{-8}$ and the maximum number of iterations $\max_{iter} = 500$.

Furthermore, we evaluate whether there is a significant effect between study groups using one-way analysis of variance (ANOVA) and reported F -statistics along with F -values and p -values. Post hoc analyses with two-tailed parametric t-tests and corrections for multiple comparisons that use the false discovery rate (FDR) are performed to determine whether the difference between the performance of the contrasts of interest is significant and are described by the p -value, T -value, and H_0 -value. The latter shows the result from evaluating the null hypothesis and that there is a significant difference between two contrasts (i.e., $H_0 = 1$ indicates that the null hypothesis is rejected, and if $H_0 = 0$, the null hypothesis is not rejected). We also provided the mean (M) of the measure of interest and the standard deviation (SD). In this paper, we use the 95% ($p < 0.05$) confidence interval (CI) to reject the null hypothesis. Of note, the relevant statistical results of group fMRI analyses are corrected for multiple comparisons with the familywise error (FWE), where $p < 0.05$.

V. RESULTS

This section is divided into three subsections. The first subsection describes the data completion results with the RMV pattern (see Section IV-F) as a performance baseline for RSCG-TT. The second and third subsections are used to evaluate the robustness of RSCG-TT on more challenging

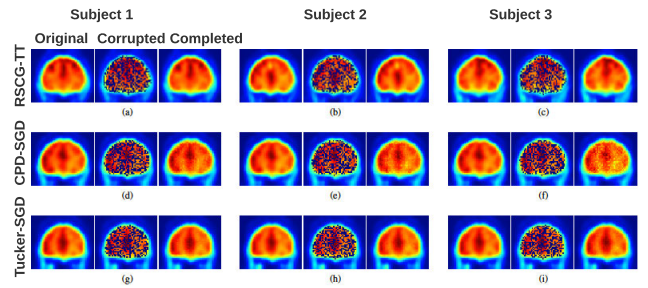


FIGURE 4. Visual comparison of 4D fMRI data completion with RSCG-TT, CPD-SGD, and Tucker-SGD on multiple subjects when subject scans are corrupted by the RMV pattern with $MR_{RMV}(\%) = 50\%$. The subject-specific scans are shown from left to right using axial projection: original scan, corrupted scan, and completed scan. (a)-(c) RSCG-TT. (d)-(f) CPD-SGD. (g)-(i) Tucker-SGD.

TABLE 5. Mean and standard deviation of the TCS for the RMV pattern with different percentages of missing values by applying the RSCG-TT algorithm. The plots corresponding to this table can be seen in Fig. 7.

$MR_{RMV}(\%)$	D=2		D=3		D=4	
	TCS, (10^{-3})		TCS, (10^{-3})		TCS, (10^{-3})	
	M	SD	M	SD	M	SD
10	9.121	1.728	5.406	0.829	0.639	0.084
20	9.169	1.646	4.187	0.647	0.654	0.087
30	11.352	2.037	4.307	0.666	0.680	0.093
40	11.587	2.085	4.403	0.671	0.693	0.097
50	13.921	2.915	4.567	0.694	0.732	0.102
60	12.827	2.919	4.758	0.717	0.773	0.109
70	17.466	3.826	5.758	0.827	0.831	0.113
80	20.086	4.433	6.014	0.964	0.975	0.152
90	48.884	13.965	11.626	2.817	1.541	0.336

tasks such as SMV pattern completion (see Section IV-G) and resting network estimation. Throughout the paper, we provide the results averaged across $N = 50$ subjects from the COBRE dataset described in Section IV-A. In this paper, the individual subject datasets are normalized to have unit variance and a zero mean before they are input into each tensor completion algorithm. Fig. 4 presents examples of three randomly selected subjects with fMRI scans corrupted by the RMV pattern and subsequently recovered by the RSCG-TT algorithm as well as other algorithms for comparison.

A. RMV PATTERN

In this section, RSCG-TT is evaluated for RMV pattern completion that uses 4D, 3D, and 2D representations. We compare the accuracy and convergence of RSCG-TT with the state-of-the-art CPD-SCD and Tucker-SCG methods for different problem sizes and tensor dimensionalities.

1) COST COMPARISON BY MISSING VALUE RATES

We investigate the performance of RSCG-TT using the 4D spatiotemporal structure of fMRI scans from the COBRE dataset with 10 – 90% voxels missing at random. See Section IV-F for a detailed description of the setup. The size of a 4D fMRI tensor is $\mathcal{X} \in \mathbb{R}^{53 \times 63 \times 46 \times 144}$.

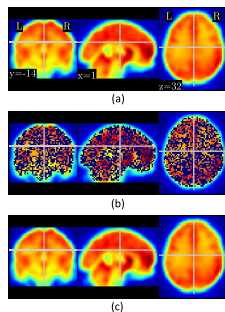


FIGURE 5. 4D fMRI tensor completion for the RMV pattern with $MR_{RMV}(\%) = 60\%$ at the 69th timepoint by applying the RSCG-TT algorithm. Orthogonal projections of the image are shown from left to right: coronal, sagittal, and axial views. (a) Original fMRI image. (b) fMRI image with randomly missing voxels within the brain mask. (c) Completed fMRI image with the following performance metrics $TCS = 0.773 \times 10^{-3}$ and $TCS_Z = 1.04 \times 10^{-3}$.

We show an example of completed fMRI brain volume with $MR_{RMV}(\%) = 60\%$ at specific timepoints in Fig. 5. We use the TCS in (51) and its threshold counterpart TCS_Z in (52) to evaluate the quality of data completion. We summarize the results of 4D tensor completion for the RMV pattern using the RSCG-TT algorithm with different $MR_{RMV}(\%)$ in Fig. 6. As shown in Fig. 5, RSCG-TT yields very low TCS values, and the recovered image in Fig. 5 is almost the same as the original image. The qualitative assessment of the completed images with $MR_{RMV}(\%) \in \{25\%, 50\%, 75\%, 90\%\}$, shown in Fig. 6a–d, suggests that RSCG-TT can recover the 4D fMRI tensor with high accuracy when there is a significant amount of missing values. We present the convergence curves of the mean RSE and the objective function w.r.t. to the number of iterations by applying RSCG-TT in Fig. 6e–f. The convergence curves shown in Fig. 6e–f illustrate the monotonic conversion of cost function (5). Furthermore, for the RMV pattern, the RSCG-TT algorithm exhibits rapid convergence even for larger $MR_{RMV}(\%)$. The loss curve in Fig. 6e demonstrates that the stationary point is reached in fewer than fifty iterations.

2) PERFORMANCE OF RSCG-TT w.r.t. TENSOR DIMENSIONALITY

In this section, we show the results of studying the performance of RSCG-TT for different tensor dimensions and different $MR_{RMV}(\%)$ values. We conduct fMRI tensor completion experiments for three different dimensions: yielding 153524×144 (2D), $3339 \times 46 \times 144$ (3D), and $53 \times 63 \times 46 \times 144$ (4D). The $MR_{RMV}(\%)$ varies from 10% to 90%. We cast the original 4D fMRI tensor to 2D and 3D tensors with TT ranks of $\mathbf{r}_{2D_{TT}} = (1, 144, 1)$ and $\mathbf{r}_{3D_{TT}} = (1, 46, 46, 1)$, respectively, and the 4D tensor has a TT rank of $\mathbf{r}_{4D_{TT}} = (1, 81, 81, 81, 1)$. The rank estimation procedure is described in Section IV-H. We summarize the effect of tensor dimensionality on the performance of RSCG-TT by applying the RMV patterns in Figs. 7, 8, Tables 5, and 6. The main effect of tensor dimensionality after the post hoc tests

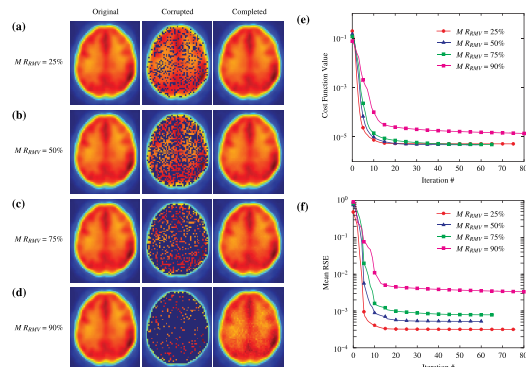


FIGURE 6. RMV pattern 4D fMRI data completion with the RSCG-TT algorithm. The completed 4D fMRI tensor for the RMV pattern with different missing ratios and the RSCG-TT algorithm. Left to right: the original image, its copy with the prescribed missing ratio, and the completed image. (a) $MR_{RMV}(\%) = 25\%$, (b) $MR_{RMV}(\%) = 50\%$, (c) $MR_{RMV}(\%) = 75\%$, (d) $MR_{RMV}(\%) = 90\%$. Convergence curves for the RMV pattern using the 4D tensor structure and RSCG-TT algorithm as a function of the number of iterations with missing values $\in \{25\%, 50\%, 75\%\}$. (e) Cost function value. (f) Mean RSE.

and the distribution of the mean TCS of RSCG-TT on each $MR_{RMV}(\%) \in \{10 - 90\%\}$ are shown in Fig. 7, Tables 6, and 7. Fig. 8 shows that when the fMRI scan is represented by a 4D tensor, the RSCG-TT algorithm has the lowest TCS for each $MR_{RMV}(\%)$ and numerically outperforms the 2D and 3D representations. Furthermore, it is clearly observed that the performance of the RSCG-TT method degrades as the dimension of the data decreases from 4D to 3D and 2D. The superior performance of 4D tensor completion can be explained by the ability of the 4D TT manifold to take full advantage of the spatiotemporal correlations among the modes of the fMRI tensor.

Fig. 8b shows the metric TCS_Z is relatively higher than the TCS. This observation illustrates that RSCG-TT performs better in the areas of the fMRI scan that have high activation since the TCS_Z metric takes only higher values ($|Z| > 2$) into consideration, while the TCS takes all values into consideration. Both the 4D and 3D representations still yield better numerical performance in the range of 0.639×10^{-3} to 1.162×10^{-2} compared with the 2D representation.

We performed ANOVA and showed that the main effect of tensor dimensionality (2D vs 3D vs 4D) was significant; $F(2, 1347) = 553.85, p < 0.0001$. Post hoc analyses with two-tailed t-tests and corrections for multiple comparisons with FDR indicated that the mean TCS for the 4D tensor ($M = 8.354 \times 10^{-4}, SD = 1.184 \times 10^{-4}$) was significantly lower than that for the 2D ($M = 1.715 \times 10^{-2}, SD = 3.632 \times 10^{-3}, t(98) = 31.757, p < 0.0001$) and 3D ($M = 5.406 \times 10^{-3}, SD = 7.285 \times 10^{-4}, t(98) = 38.615, p < 0.0001$) tensors (see Table 6 and Fig. 7). We further studied the ability of RSCG-TT to fully recover original images corrupted with varying $MR_{RMV}(\%)$ levels for different tensor dimensions in the RMV pattern. We illustrate the results of the experiment in Fig. 9 and Table 7. We quantitatively evaluated the statistical significance between the mean voxel intensity values

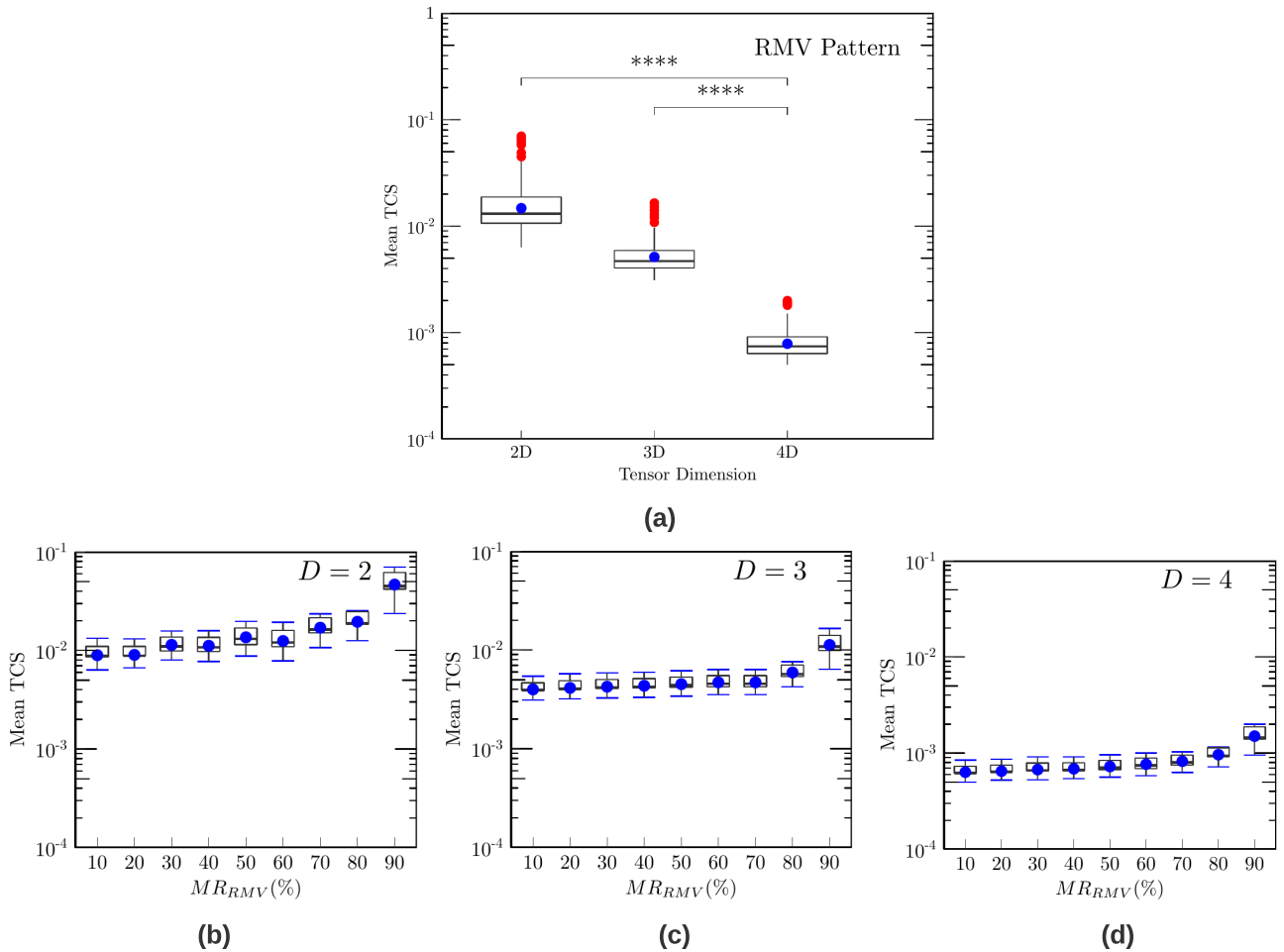


FIGURE 7. Effect of the tensor dimensionality on the distribution of the mean TCS (N=50) for the RMV pattern with the RSCG-TT algorithm. The boxplots summarize the distribution of the mean TCS scores for 2D, 3D, and 4D tensor dimensions. The median values are represented by black lines inside the boxplot, and the tops of the whisker lines indicate the 25th and 75th percentile values. The mean values are plotted in blue, and outliers are represented by red circles. Error bars represent the standard error of the mean. (a) Main effect of the tensor dimensionality on the mean TCS: 2D vs 3D vs 4D. Post hoc analyses with two-tailed t-tests (FDR corrected) indicate that the mean TCS for the 4D tensor ($M = 8.354 \times 10^{-4}$, $SD = 1.184 \times 10^{-4}$) is significantly lower than that for the 2D ($M = 1.715 \times 10^{-2}$, $SD = 3.632 \times 10^{-3}$) and 3D ($M = 5.406 \times 10^{-3}$, $SD = 8.285 \times 10^{-4}$) tensors with $p < 0.0001$. **** indicates $p < 0.0001$ (post hoc and FDR corrected), which means significantly different. Distribution of the mean TCS (N = 50) with $MR_{RMV}(\%) \in \{10 - 90\}$. (b) Distribution of the mean TCS for tensor dimension $D = 2$. (c) Distribution of the mean TCS for tensor dimension $D = 3$. (d) Distribution of the mean TCS for tensor dimension $D = 4$. The mean TCS results are listed in Table 5.

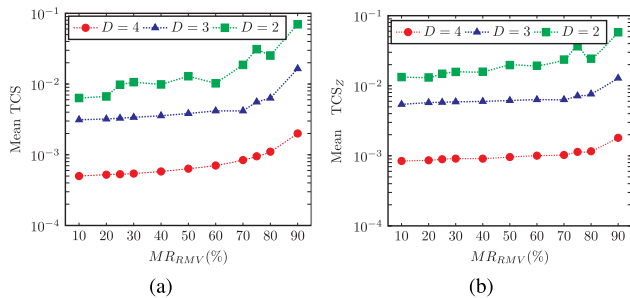


FIGURE 8. Effect of tensor dimensionality on fMRI data completion by missing value rates for tensor dimensions $D=2,3,4$ on the RMV pattern. The mean TCS results are shown. (N = 50). (a) TCS. (b) TCS_z .

of the original fMRI images and the images after data completion for tensor dimensions $D = 2, 3, 4$ and $MR_{RMV}(\%) \in \{10 - 90\}$. Post hoc analyses with two-tailed t-tests (FDR corrected) and the significance level $p < 0.05$ showed that

RSCG-TT can recover the original image with $MRV_{RMV}(\%)$ values in range of $\{10 - 60\}$ for 3D, $\{10 - 80\}$ for 4D tensor, and only 10% for 2D tensor format. As shown in Table 7, and Fig. 9, the mean voxel intensity values between the original and the completed images were statistically different for 2D format with $MR_{RMV} \in \{20 - 90\}$. In contrast, the high-order tensor formats did not fully recover only the original data for the extremely high values of $MR_{RMV}(\%) \in \{70 - 90\}$ for 3D, and for 4D with $MR_{RMV}(\%) = 90\%$. Together, the results presented in this section suggest that tensor dimensionality does have an effect on fMRI data completion. Specifically, our results demonstrate that the original 4D spatiotemporal representation of fMRI data has a higher accuracy than the 2D and 3D representations of the original data. In summary, our experiments show that RSCG-TT can recover corrupted fMRI data in presence of high missing value rates up to 80% when the data are kept in the natural 4D format.

3) COMPARISON OF RSCG-TT WITH THE STATE-OF-THE-ART METHODS

In this section, we compare the performance of RSCG-TT and the state-of-the-art methods on the reconstruction task of 4D fMRI data with $MR_{RMV}(\%) \in \{10\%, 90\%\}$ with an increment of 10% corrupted by the RMV pattern. We evaluated two aspects of the performance: the numerical accuracy in terms of the RSE and TCS and the computational cost w.r.t. the convergence rate and the CPU time. We summarize the computational cost results of all algorithms in Fig. 10 and present the mean metrics of the numerical reconstruction performance results in Fig. 11 and Table 8. The main effect of the tensor completion method (CPD-SGD vs Tucker-SGD vs RSCG-TT) after post hoc analyses is shown in Table 9 and Fig. 12.

We show the convergence rate of the objective function for all methods w.r.t to the iteration in Fig. 10a and present the convergence curves of the objective function w.r.t to CPU time in Fig. 10b. The convergence curves presented in Fig. 10 and the numerical results in Fig. 11 demonstrate that RSCG-TT outperforms both competitor algorithms in terms of numerical accuracy measures (the RSE and TCS) and computational costs. In Fig. 10b, we compare the computational efficiency of RSCG-TT and the state-of-the-art algorithms as a function of the number of iterations and the elapsed time. As shown in Fig. 10a, RSCG-TT converges rapidly compared with both the CPD-SGD and Tucker-SGD algorithms. The state-of-the-art algorithms require a much larger number of iterations to converge. As shown in Fig. 10, while the total elapsed time to converge is similar, the cost function accuracy of RSCG-TT is superior to the state-of-the-art algorithms by two to three fold. We present the mean ($N = 50$) TCS and RSE results of 4D tensor completion with RSCG-TT and compare these results with the results of the state-of-the-art methods: CPD-SGD and Tucker-SGD in Fig. 11f–h. For the fixed rank model in the 4D experiments, we estimate the fixed rank parameters as $\mathbf{r}_{4D_{TT}} = (1, 81, 81, 81, 1)$ for the TT rank, $\mathbf{r}_{Tucker_{4D}} = (53, 63, 46, 91)$ for the Tucker-rank, and $R_{CPD_{4D}} = 81$ for the CPD-rank, as described in Section IV-H. In terms of the RSE, both RSCG-TT and Tucker-SCG demonstrate good numerical performance w.r.t. the RSE for RSCG-TT ($RSE_{RSCG-TT}$) in the range of 0.391×10^{-3} to 0.857×10^{-3} , and the RSE for Tucker-SGD ($RSE_{Tucker-SGD}$) is in the range from 5.529×10^{-3} to 1.709×10^{-2} . The RSE for CPD-SGD ($RSE_{CPD-SGD}$) is in the range of 2.782×10^{-2} to 5.261×10^{-2} , which is higher in order of magnitude than the RSE of both RSCG-TT and Tucker-SCG. The RSE results for the RMV pattern in the case of $D = 4$ are shown in Fig. 11f. In terms of the TCS, we noted a similar range of values as we observed for the RSE of all the compared algorithms: $TCS_{RSCG-TT} \approx 0.639 \times 10^{-3} - 1.154 \times 10^{-3}$, $TCS_{Tucker-SGD} \approx 3.981 \times 10^{-3} - 5.987 \times 10^{-3}$, and $TCS_{CPD-SGD} \approx 2.243 \times 10^{-2} - 2.893 \times 10^{-2}$. The TCS results for the RMV pattern in the case of $D = 4$ are shown in Figs. 11g, 11h, and Table 8. As seen in Fig. 10, CPD-SGD

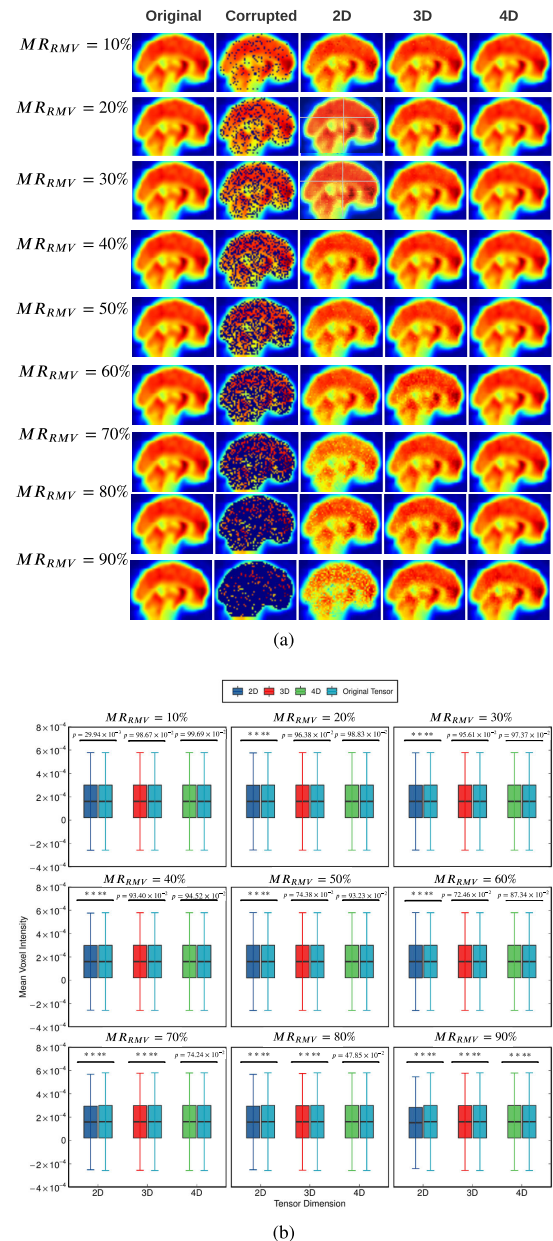


FIGURE 9. Comparison of 2D, 3D, 4D tensor completion on fMRI data for missing value rates $\in \{10 - 90\%$ in the RMV pattern using RSCG-TT algorithm. (a) Left to right: original, corrupted, and completed fMRI images for 2D, 3D, 4D tensor dimensions using sagittal views are shown. (b) Main effect of the tensor dimensionality on the ability to recover the original image for the RMV pattern using different percentages of missing value rates. The boxplots summarize the distribution of the mean voxel intensity values for the $MR_{RMV}(\%) \in \{10 - 90\%$ and tensor dimensions $D = 2, 3, 4$. Post hoc analyses with two-tailed t-tests (FDR corrected) indicate that RSCG-TT can recover the original image with $MR_{RMV}(\%)$ in the range of $\{10 - 60\%$ for 3D, and for 4D tensor with $MR_{RMV}(\%) \in \{10 - 80\%$. **** indicates $p < 0.0001$ (post hoc and FDR corrected), which means significantly different. i

and Tucker-SCG demonstrate a lower computational efficiency compared with RSCG-TT in terms of both the convergence rate and numerical accuracy of the objective function w.r.t. to the CPU time (see Figs. 10a and 10b). Moreover, it is worth noting that RSCG-TT results in consistently smaller

TABLE 6. Comparison of the mean TCS by tensor dimension for the RMV pattern with the RSCG-TT algorithm. The plots corresponding to this table can be seen in Fig. 7.

Evaluation Metric	Tensor Dimension	M	SD	$t(98)$	p	Cohen- d	H_0
TCS	D=2	0.0172	0.0036	31.757	0.0001****a	1.789	1
	D=4	0.0008	0.0001				
	D=3	0.0054	0.0007	38.615	0.0001****a	2.524	1
	D=4	0.0008	0.0001				

^a **** indicates $p < 0.0001$ (post hoc, FDR corrected), which means significantly different.

TABLE 7. Comparison of the mean voxel intensity between the original fMRI image and the completed images using the RSCG-TT algorithm for the RMV pattern and different MR_{RMV} (%). The plots corresponding to this table can be seen in Fig. 9

MR_{RMV} (%)	2D					3D					4D				
	Mean voxel intensity (10^{-4})					Mean voxel intensity(10^{-4})					Mean voxel intensity (10^{-4})				
	M	SD	$t(44235070)$	p	H_0	M	SD	$t(44235070)$	p	H_0	M	SD	$t(44235070)$	p	H_0
10	1.606	1.393	2.170	29.94×10^{-3}	0	1.606	1.393	2.170	98.86×10^{-2}	0	1.606	1.393	-0.017	99.68×10^{-2}	0
20	1.606	1.386	14.737	0.0001****	1	1.606	1.393	0.045	96.34×10^{-2}	0	1.606	1.393	0.017	96.18×10^{-2}	0
30	1.603	1.389	8.945	0.0001****	1	1.606	1.393	0.055	95.61×10^{-2}	0	1.606	1.393	0.033	97.37×10^{-2}	0
40	1.600	1.385	17.406	0.0001****	1	1.606	1.393	0.083	93.40×10^{-2}	0	1.606	1.393	0.069	94.52×10^{-2}	0
50	1.605	1.395	3.846	0.0001****	1	1.606	1.393	0.327	74.37×10^{-2}	0	1.606	1.393	0.085	93.22×10^{-2}	0
60	1.605	1.395	3.846	0.0001****	1	1.606	1.393	0.352	72.46×10^{-2}	0	1.606	1.393	0.159	87.34×10^{-2}	0
70	1.578	1.363	71.327	0.0001****	1	1.606	1.600	16.207	0.0001****	1	1.606	1.393	0.329	74.23×10^{-2}	0
80	1.576	1.362	74.342	0.0001****	1	1.602	1.387	12.487	0.0001****	1	1.607	1.393	0.709	47.85×10^{-2}	0
90	1.519	1.310	217.350	0.0001****	1	1.607	1.380	28.021	0.0001****	1	1.604	1.390	7.086	0.0001****	1

*** indicates $p < 0.0001$ (post hoc, FDR corrected), which means significantly different.

TABLE 8. Mean and standard deviation of the TCS with the RMV pattern with different percentages of missing values for RSCG-TT and the state-of-the-art methods. The plots corresponding to this table can be seen in Fig. 11.

MR_{RMV} (%)	CPD-SGD		Tucker-SGD		RSCG-TT	
	TCS (10^{-3})		TCS (10^{-3})		TCS (10^{-3})	
	M	SD	M	SD	M	SD
10	42.889	13.129	5.745	1.203	0.639	0.084
20	35.439	9.161	5.351	0.985	0.654	0.087
30	38.666	11.304	6.743	1.699	0.680	0.093
40	55.025	19.606	5.558	1.119	0.693	0.097
50	39.684	11.618	6.206	1.323	0.732	0.102
60	34.002	7.929	6.950	1.693	0.773	0.109
70	34.843	8.425	7.769	2.055	0.831	0.113
80	33.456	6.841	8.194	2.170	0.975	0.152
90	39.635	9.937	17.956	5.375	1.541	0.336

TCS_Z values compared with the state-of-the-art methods, as demonstrated in Fig. 11h. This can be interpreted as RSCG-TT having a higher sensitivity in order to complete the data in the areas with very low activation values. Fig. 11a-e presents the results on the completed 4D fMRI data for MR_{RMV} (%) = 75% with the state-of-the-art methods and RSCG-TT. A visual inspection of the completed images clearly shows that the CPD-SGD in Fig. 11c has blurred areas and suffers from poor reconstruction, the Tucker-SGD in Fig. 11d shows better performance, and the images completed by RSCG-TT are almost indistinguishable from the original images shown in Fig. 11a.

Based on the ANOVA results, we identified that the main effect of the tensor completion method (CPD-SGD vs Tucker-SGD vs RSCG-TT) was significant;

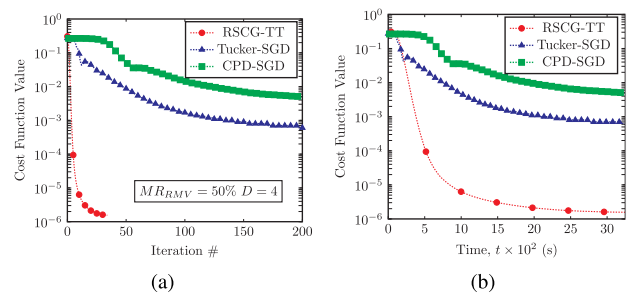


FIGURE 10. Computational cost comparison of RSCG-TT and the state-of-the-art methods with different percentages of missing values in the RMV pattern. (a) The convergence curves with MR_{RMV} (%) = 50% w.r.t. the number of iterations. (b) The convergence curves with MR_{RMV} (%) = 50% w.r.t. to the CPU time.

i.e., $F(2, 1347) = 628.83, p < 0.0001$. Post hoc analyses (FDR corrected) with t-tests indicate that the mean TCS of RSCG-TT ($M = 8.354 \times 10^{-4}, SD = 1.184 \times 10^{-4}$) was significantly lower than those of the CPD-SGD ($M = 3.929 \times 10^{-2}, SD = 1.087 \times 10^{-2}, t(98) = 25.003, p < 0.0001$) and Tucker-SGD ($M = 7.831 \times 10^{-3}, SD = 1.918 \times 10^{-3}, t(98) = 25.735, p < 0.0001$) tensor completion methods (see Table 9 and Fig. 12).

B. SMV PATTERN

In this section, we evaluate RSCG-TT with the SMV pattern described in Section IV-G. We investigate the joint effect of the spatiotemporal missing value rate and the effect of the tensor dimensionality to estimate the missing voxels that originated from the SMV pattern. We show an example of SMV pattern completion in Fig. 13.

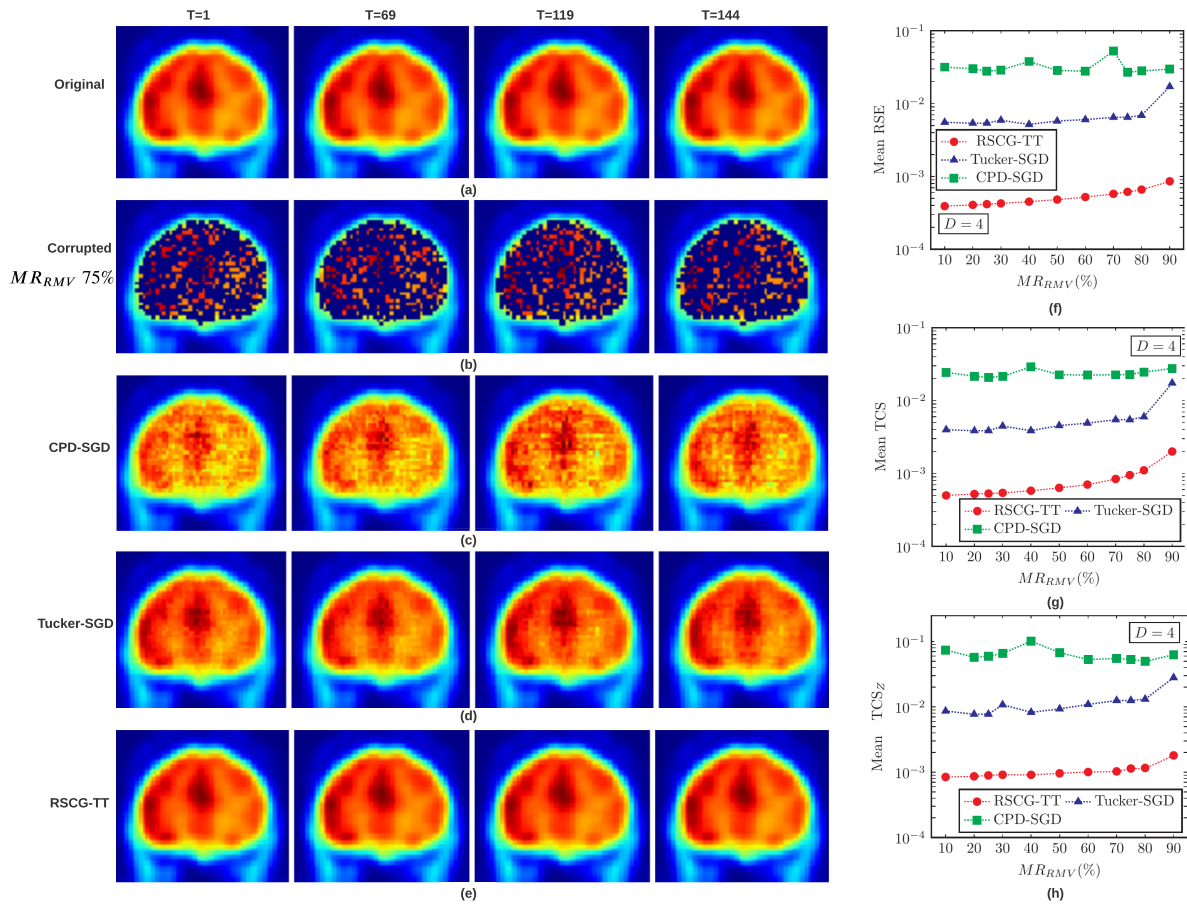


FIGURE 11. Comparison of RSCG-TT, Tucker-SGD, and CPD-SGD 4D tensor completion on fMRI data for missing value rates \in {75%} in the RMV pattern. The 1st, 69th, 119th, and 144th timepoints are shown from left to right with a coronal projection. (a) Original fMRI scan. (b) Missing fMRI data. (c) CPD-SGD. (d) Tucker-SGD. (e) RSCG-TT. Comparison of the numerical accuracies of RSCG-TT and the state-of-the-art methods with different percentages of missing values in the RMV pattern. The mean results are shown ($N = 50$). (f) RSE comparison. (g) TCS comparison. (h) TCS_z comparison.

TABLE 9. Comparison of the mean TCS results of the tensor completion methods for the RMV pattern and tensor dimension $D=4$. The plots corresponding to this table can be seen in Fig. 12.

Evaluation Metric	Tensor Completion Method	M	SD	$t(98)$	p	Cohen- d	H_0
TCS	CPD-SGD	0.0392	0.0109	25.003	0.0001****a	4.183	1
	RSCG-TT	0.0008	0.0001				
	Tucker-SGD	0.0078	0.0192	25.735	0.0001****a	2.261	1
	RSCG-TT	0.0008	0.0001				

^a **** indicates $p < 0.0001$ (post hoc, FDR corrected), which means significantly different.

1) PERFORMANCE OF RSCG-TT FOR TENSORS OF DIFFERENT DIMENSIONS BY VARYING THE SPATIOTEMPORAL MISSING EFFECT SIZE

In this section, we study the joint effect of the missing value rates in the spatial and temporal dimensions affected by the SMV pattern for tensors of different dimensions with RSCG-TT. We summarize the results of fMRI data completion for the SMV pattern with different $MR_T(\%)$ and $MR_{V_{3D},(\%)}$ values and RSCG-TT in Figs. 13 and 14. The results shown in Fig. 13 are part of the experiment when the

SMV pattern is placed at randomly selected timepoints $T_i = \{1, 4, 5, 14, 20, 33, 37, 62, 64, 65, 71, 78, 94, 104, 110, 113, 120, 122, 129, 135, 136\}$ and the temporal missing value rates are $MR_T(\%) = 15\%$. Fig. 13b shows an example in which a spatial ellipsoid with missing voxels of size 1 (see Table 3) is placed in the brain volume at spatial location $(x, y, z) = (2, 32, 22)$ and timepoint $T_i = 20$. The completed fMRI image at timepoint $T_i = 20$ is shown in Fig. 13c. Fig. 14 illustrates the joint effect of both spatiotemporal factors $MR_T(\%)$ and $MR_{V_{3D},(\%)}$ on the quality of data

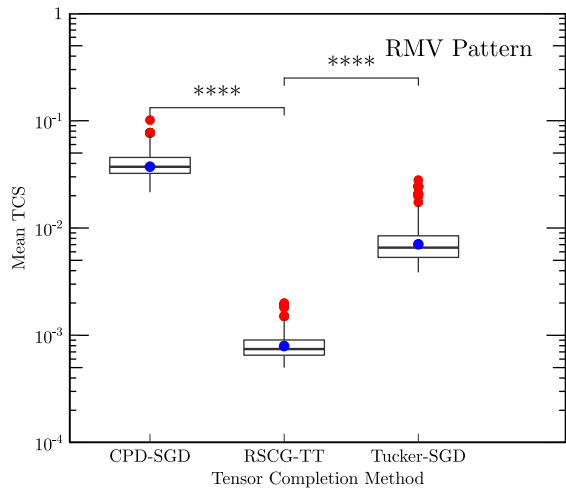


FIGURE 12. Main effect of the tensor completion method on the mean TCS ($N = 50$) for the RMV pattern and $D = 4$. The boxplots summarize the distribution of the mean TCS of the following tensor completion methods: CPD-SGD, Tucker-SGD, and RSCG-TT. The median values are represented by black lines inside the boxplot, and the tops of whisker lines represent the 25th and 75th percentile values. The mean values are plotted in blue, and outliers are represented by red circles. The error bars represent the standard error of the mean. Post hoc analyses with t-tests (FDR corrected) indicate that the mean TCS of RSCG-TT ($M = 8.354 \times 10^{-4}$, $SD = 1.184 \times 10^{-4}$) was significantly lower than those of the CPD-SGD ($M = 3.921 \times 10^{-2}$, $SD = 1.087 \times 10^{-2}$) and Tucker-SGD ($M = 7.831 \times 10^{-3}$, $SD = 1.918 \times 10^{-3}$) tensor completion methods with $p < 0.0001$. **** indicates $p < 0.0001$ (post hoc and FDR corrected), which means significantly different.

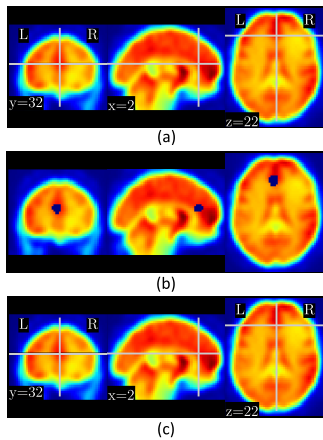


FIGURE 13. 4D fMRI tensor completion on the SMV pattern with $MR_T = 15\%$ and $MR_{V_{e3D}} = 1.75\%$ at the 20th timepoint. The orthogonal projections of the image are shown from left to right: coronal, sagittal, and axial views. (a) Original fMRI image. (b) fMRI image with the spatial ellipsoid in (48a) filled with missing voxels of size 1 within a brain mask. (c) Completed fMRI image with the following performance metrics: $TCS = 6.931 \times 10^{-5}$ and $TCS_Z = 3.92 \times 10^{-3}$.

completion. As we can observe from Fig. 14, the simultaneous increase in both $MR_T(\%)$ and $MR_{V_{e3D},(\%)}$ impacts the numerical accuracy of data completion. However, as shown in Fig. 14, RSCG-TT obtains good quality data completion in terms of the TCS in the range of $1 \times 10^{-4} - 2 \times 10^{-3}$ for the problem sizes of the spatial ellipsoids $\in \{1, 2, 3, 4\}$

and $MR_T(\%)$ up to 20%. The observed results for the SMV pattern are in agreement with those from previous works [13] and [14]. The effect size of the SMV pattern proportionally decreases the number of neighborhood voxels and, as a result, reduces the inferential power of the tensor completion methods.

2) EFFECT OF THE TENSOR DIMENSIONALITY ON TENSOR COMPLETION FOR THE SMV PATTERN

In this section, we study the effect of tensor dimensionality on the estimation of missing voxels with RSCG-TT when fMRI scans are affected by the SMV pattern. We summarize the effect of tensor dimensionality for the SMV pattern in Figs. 15, 16, 17, Table 10 and Table 11.

Fig. 15 presents a visual comparison of the obtained completed images with RSCG-TT for 2D, 3D, and 4D tensor formats corrupted by the SMV pattern with different $MR_T(\%)$. Despite the challenging missing SMV pattern, simultaneous temporal and volumetric corruption, successful reconstruction was obtained for 4D and 3D, but not for the 2D tensor format. In particular, the completed 4D and 3D fMRI images appear visually almost indistinguishable from the original ground truth 4D fMRI tensor. RSCG-TT obtained good quality of data completion for $MR_T(\%)$ in the range of $\{5 - 15\}$. However, RSCG-TT were unable to successfully recover fMRI data in form of 2D representation for all $MR_T(\%)$. The poor performance of RSCG-TT for 2D tensor format can be explained by lack of spatiotemporal structure when the original 4D fMRI is flattened to 2D form, in which, the natural strength of the relationships is lost.

Table 10 and Fig. 16 show quantitative evaluations of RSCG-TT for 2D, 3D, and 4D formats in the SMV pattern and different $MR_T(\%)$. We calculated the mean TCS ($N = 50$) between the data completion results and the ground-truth original 4D fMRI data. In terms of TCS, the reconstruction of 2D representation demonstrated suboptimal performance with the mean TCS in the range from 0.159 to 0.396. Conversely, RSCG-TT was able to obtain good numerical performance for both 4D and 3D formats w.r.t. the mean TCS for 4D in the range from 0.3289×10^{-3} to 0.0146×10^{-3} , and the TCS for 3D format in the range from 0.005 to 0.057. To evaluate the statistical significance between tensor formats using RSCG-TT in the SMV pattern, we performed ANOVA followed by post hoc analyses with two-tailed t-tests corrected for multiple comparisons with $p < 0.05$ (see Table 11 and Fig. 15). The boxplots shown in Fig. 17 summarize the distribution of mean TCS metric for tensor dimensions 2D, 3D, and 4D. The results underline the fact that spatiotemporal data completion leads to statistically significant improvements compared with volumetric 3D, and flattened 2D tensor formats. Furthermore, from Figs. 15, 16 and 17, it can be noticed that the natural spatiotemporal 4D tensor representation outperforms both alternative views of the fMRI tensor casted to the 3D and 2D dimensional forms across all $MR_T(\%)$.

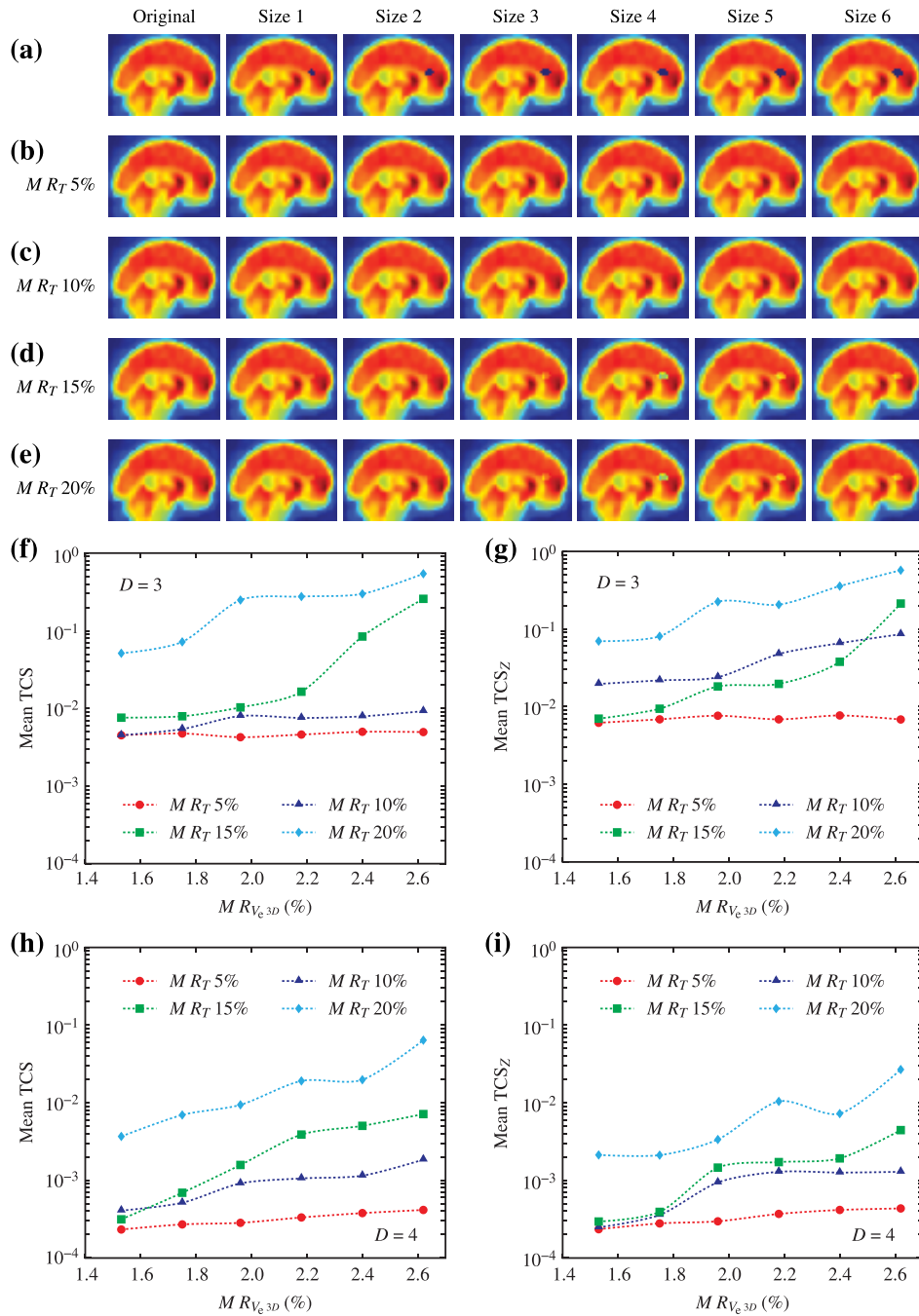


FIGURE 14. Results of 4D fMRI tensor completion on fMRI data with RSCG-TT and the SMV pattern by varying the spatial volume of the missing values for different percentages of missing temporal ratios. Left: The original brain volume at the 20th timepoint. (a) Left to right. Brain volumes at the 20th timepoint with embedded spatial ellipsoids with missing voxel sizes $\in \{1, 2, 3, 4, 5, 6\}$ (see Table 4). (b)-(e) Top to bottom. Completed brain volumes for each spatial ellipsoid size and $MR_T \in \{5\%, 10\%, 15\%, 20\%$. Comparison of fMRI tensor completion for 3D and 4D tensor representations on the SMV pattern for different percentages of spatiotemporal missing value ratios. The mean TCS results are shown ($N = 50$). (f) $MR_T(\%) = 5\%$. (g) $MR_T(\%) = 10\%$. (h) $MR_T(\%) = 15\%$. (i) $MR_T(\%) = 20\%$.

We present the main effect of tensor dimensionality on the results for the SMV pattern in Fig. 17a and Table 11 and the distribution of the mean TCS in Fig. 17b. The main effect of tensor dimensionality (2D vs 3D vs 4D) in the presence of the SMV pattern was significant; i.e., $F(2, 71997) = 40362.579, p < 0.0001$. The subsequent post hoc analyses

with t-tests (FDR corrected) indicated that the mean TCS for the 4D tensor ($M = 4.575 \times 10^{-3}, SD = 1.474 \times 10^{-3}$) was significantly lower than that for the 2D ($M = 2.758 \times 10^{-1}, SD = 4.832 \times 10^{-2}, t(11998) = 431.28, p < 0.0001$) and 3D ($M = 8.475 \times 10^{-2}, SD = 6.405 \times 10^{-2}, t(11998) = 96.735, p < 0.0001$) tensors. These results suggest that

TABLE 10. Mean and standard deviation of the TCS for the SMV pattern with different percentages of temporal missing values $MR_T(\%)$ and tensor dimensions $D \in \{2, 3, 4\}$ and with RSCG-TT. The plots corresponding to this table can be seen in Fig. 16 and Fig. 15.

Tensor Dimension	$MR_T(\%)$							
	5		10		15		20	
	TCS, 10^{-3}		TCS, 10^{-3}		TCS, 10^{-3}		TCS, 10^{-3}	
	M	SD	M	SD	M	SD	M	SD
D=2	159.937 4	56.233 6	208.909 2	47.070 4	337.667 6	65.371 9	396.757 3	13.881 5
D=3	5.841 0	0.887 3	25.437 1	19.998 5	57.506 6	82.221 3	249.907 8	167.304 1
D=4	0.328 9	0.068 7	0.940 7	0.447 7	2.413 5	2.006 9	14.620 4	15.603 6

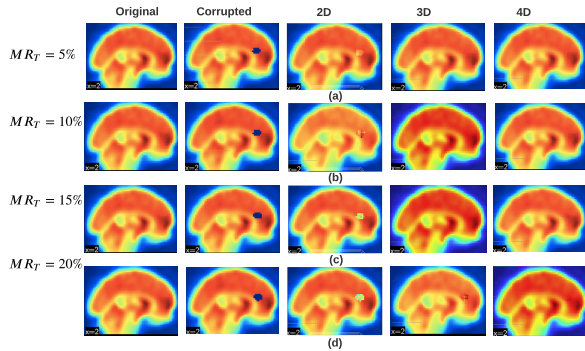


FIGURE 15. Qualitative comparison of 2D, 3D, and 4D tensor completion using RSCG-TT algorithm on fMRI data for spatiotemporal missing value rates $MR_T(\%) \in \{5 - 20\%$ in the SMV pattern. Left to right: original, corrupted, and completed images for 2D, 3D, 4D representations are shown with sagittal projection. (a) $MR_T(\%) = 5\%$. (b) $MR_T(\%) = 10\%$. (c) $MR_T(\%) = 15\%$. (d) $MR_T(\%) = 20\%$.

tensor dimensionality does have an effect on fMRI data completion when the original data are affected by the SMV pattern.

C. EVALUATION OF THE QUALITY OF THE TENSOR COMPLETION METHODS ON DOWNSTREAM fMRI ANALYSES

In this section, we evaluated the effect of each tensor completion method on the sensitivity of fMRI analyses to estimate the resting network (RSN) components. We decomposed the rs-fMRI datasets (see Section IV-A) into functional networks using spatial independent component analysis (ICA) to derive maximally independent RSNs. We conducted group ICA (GICA) [82], [83] to evaluate the sensitivity of RSCG-TT and the state-of-the-art tensor completion methods CPD-SGD and Tucker-SGD to voxel activation. We generate four datasets for each of the $N = 50$ subjects, one original dataset and three datasets from the completion algorithms for $MR_{RMV}(\%) = 50\%$. We used GICA to decompose the datasets and used direct GICA3 [83], [84] back-reconstruction to obtain subject-specific RSN components. We used the Infomax algorithm [85] and estimated $C = 50$ RSN components as suggested by [86]. The whole analysis was implemented with the GIFT toolbox.² The Infomax

²urlhttps://trendscenter.org/software/gift/

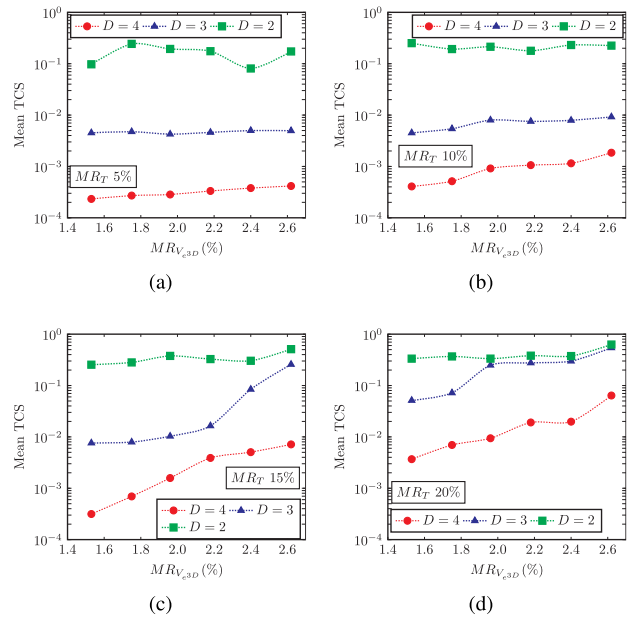


FIGURE 16. Results of fMRI tensor completion affected by the SMV pattern by varying the spatial volumes of missing values for different percentages of missing temporal ratios. The mean TCS results are shown ($N = 50$). (a) $MR_T(\%) = 5\%$. (b) $MR_T(\%) = 10\%$. (c) $MR_T(\%) = 15\%$. (d) $MR_T(\%) = 20\%$.

algorithm was repeated the process ten times, and the final decomposition was chosen based on the most consistent run selection scheme [87]. We estimated 16 independent components (IC) to be RSNs (as opposed to noise artifacts) by inspecting the aggregate ICs and the power spectra [88]. The RSNs were divided into different groups according their functional and anatomical properties [88] and categorized as auditory (AUD), visual (VIS), sensorimotor (SM), cognitive control (CC), frontoparietal (FP), and default-mode networks (DMN).

1) GROUP-LEVEL SENSITIVITY COMPARISON BY TENSOR COMPLETION METHOD

To compare the group-level sensitivity of the three completion methods, we computed voxel-wise t -statistics from a one-sample t -test of the subject-level RSNs and plotted the t -values as a statistical image called the T-map. The threshold of the T-maps is $p < 0.05$, and the T-maps are

TABLE 11. Comparison of the mean TCS results for the RSCG-TT algorithm by tensor dimension for the SMV pattern.

Tensor Dimension	$MR_T(\%)$							
	5		10		15		20	
	M	SD	M	SD	M	SD	M	SD
D=2	159.937 4	56.233 6	208.909 2	47.070 4	337.667 6	65.371 9	396.757 3	13.881 5
D=3	5.841 0	0.887 3	25.437 1	19.998 5	57.506 6	82.221 3	249.907 8	167.304 1
D=4	0.328 9	0.068 7	0.940 7	0.447 7	2.413 5	2.006 9	14.620 4	15.603 6

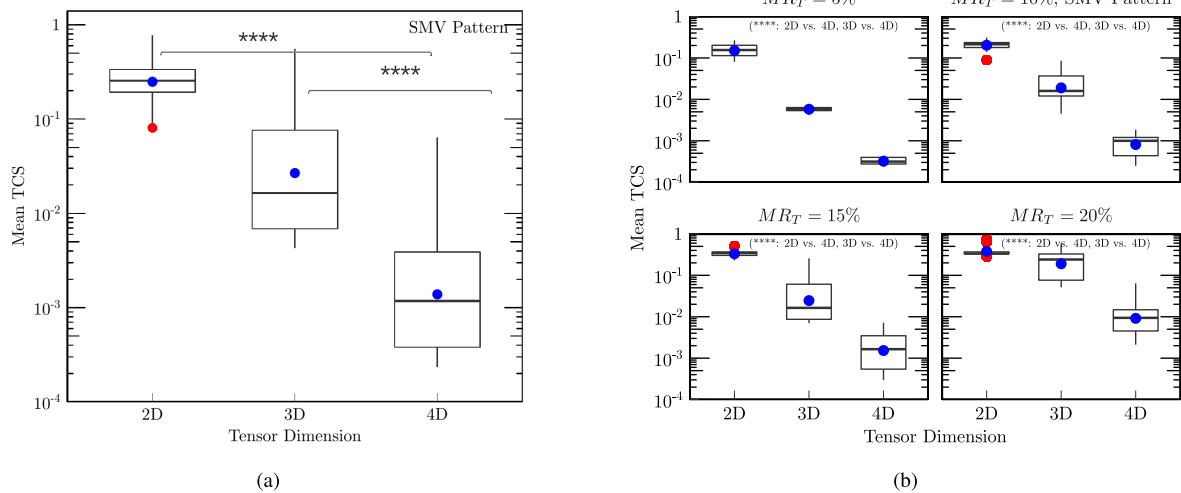


FIGURE 17. Effect of tensor dimensionality on the mean TCS ($N = 50$) for the SMV pattern with the RSCG-TT algorithm. The boxplots summarize the distribution of mean TCS scores for tensor dimensions 2D, 3D, and 4D. The median values are represented by black lines inside the boxplot, and the tops of the whisker lines indicate the 25th and 75th percentile values. The mean values are plotted in blue, and outliers are represented by red circles. The error bars represent the standard error of the mean. The median values are represented by black lines inside the boxplot, and the tops of the whisker lines indicate the 25th and 75th percentile values. The mean values are plotted in blue, and outliers are represented by red circles. The error bars represent the standard error of the mean. (a) The main effect of tensor dimensionality on the mean TCS: 2D vs 3D vs 4D. Post hoc analyses with two-tailed t -tests (FDR corrected) indicate that the mean TCS for the 4D tensor ($M = 4.575 \times 10^{-3}$, $SD = 4.474 \times 10^{-3}$) was significantly lower than those for the 2D ($M = 2.75 \times 10^{-2}$, $SD = 4.849 \times 10^{-3}$) and 3D ($M = 8.467 \times 10^{-3}$, $SD = 6.398 \times 10^{-4}$) tensors with $p < 0.0001$. **** indicates $p < 0.0001$ (post hoc and FDR corrected), which means significantly different. (b) The distribution of the mean TCS with $MR_T(\%) \in \{5\%, 10\%, 15\%, 20\%\}$. The mean TCS results are listed in Table 10.

corrected for multiple comparisons with FWE. We evaluated the group-level sensitivity of each method by studying its ability to recover independent RSNs across six functional RSNs. We present the T-maps of the RSNs comprising six functional networks: DMN, AUD, VIS, SM, CC, and FP in Fig. 18 and Supplementary Fig. S.1. We focused on a reliable estimation of major RSNs, as they are important biomarkers in a number of disorders; see, e.g., [88]–[90].

The T-maps following different tensor completion methods for the canonical DMN component are shown in Fig. 18, and the T-maps for other RSNs are shown in Supplementary Fig. S.1. All six RSNs resulting from RSCG-TT are matched with the reference RSNs (see Fig. 18 and Supplementary Fig. S.1). The results illustrate that the RSN components estimated from the datasets completed by RSCG-TT are the most similar to those estimated from the original datasets. In contrast, it is evident from Fig. 18 and Supplementary Fig. S.1 that the spatial clusters comprising the RSNs are

generally smaller by a certain extent following the CPD-SGD and Tucker-SGD methods. We characterize the performance of each tensor completion method w.r.t. the proportion of significant activations in Table 12 and Supplementary Fig. S.3. In summary, the results presented in Table 12, Fig. 18, Supplementary Fig. S1 and Supplementary Fig. S.3 demonstrate that the t -statistics, the proportion of significant voxels, and the spatial cluster extents are higher in RSCG-TT, indicating better sensitivity.

To quantitatively compare the RSN components across all tensor completion methods, we performed a two-sample t -test on the spatial maps between the reference RSN component estimated from the original datasets and the RSN component generated after each tensor completion method. Fig. 19 and Supplementary Fig. S.2 illustrate the T-map contrasts yielded from the two-sample t -tests: the reference RSN – CPD-SGD RSN, reference RSN – Tucker-SGD RSN, and reference RSN – RSCG-TT. The threshold of the T-map contrasts

TABLE 12. Peak activations of RSN components after the tensor completion algorithms. Cluster ID: the enumerated cluster ID of peak activations within the RSN component; t_{\max} : the maximum t-statistic in each cluster; Coordinate: x, y, z (in mm) of t_{\max} in MNI space; Cluster size: the cluster size (in mm^3); V_l : the number of voxels in each cluster; $V_l(\%)$: the percentage of activated voxels.

Method/Network	Cluster ID	Coordinate	t_{\max}	Cluster Size, mm^3	V_l	$V_l(\%)$
CPD-SGD						
Default Mode Network	1	(0, -57, 18)	6.86	22,207	1204	2.43
Auditory Network	1	(-45, -12, 15)	6.93	11,961	1289	2.11
Visual Network	1	(0, 84, 15)	6.67	26,244	3488	2.85
Sensorimotor Network	1	(-18, -18, 81)	12.29	1134	719	1.19
Cognitive Control Network	1	(-36, -72, 54)	9.66	5346	708	1.72
Fronto-Parietal Network	1	(46, -75, 36)	7.38	7128	660	1.58
Tucker-SGD						
Default Mode Network	1	(-6, -66, 36)	12.37	35,586	1402	2.34
Auditory Network	1	(-51, -24, 6)	10.29	44,901	3670	7.94
Visual Network	1	(-6, -66, 3)	9.84	33,318	1281	2.32
Sensorimotor Network	1	(-54, -27, 57)	9.48	43,713	5689	4.31
Cognitive Control Network	1	(-57, -57, 45)	12.67	14,985	795	1.32
Fronto-Parietal Network	1	(48, -75, 39)	11.35	16,875	810	1.39
RSCG-TT						
Default Mode Network	1	(-3, -66, 33)	83.67	128,034	5161	12.26
Auditory Network	1	(-48, -21, 9)	62.16	68,931	6560	16.84
Visual Network	1	(6, -81, 9)	62.67	92,610	9015	15.02
Sensorimotor Network	1	(-51, -27, 54)	61.38	86,805	3568	9.48
Cognitive Control Network	1	(-39, -69, 51)	70.75	76,005	10,230	17.04
Fronto-Parietal Network	1	(51, -51, 45)	56.97	72,172	5585	15.41

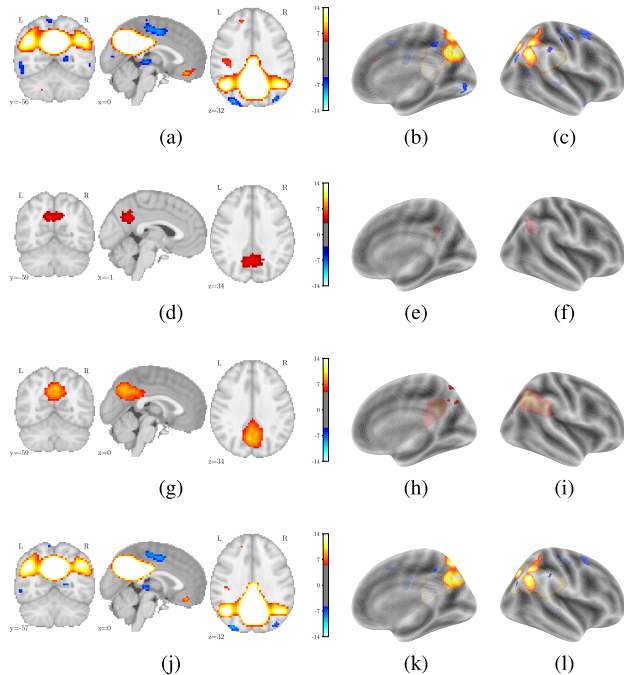


FIGURE 18. Estimation of the DMN component by GICA following different tensor completion algorithms. The group-level 2D and 3D T-maps for the reference DMN, CPD-SGD, Tucker-SGD, and RSCG-TT are shown. The threshold of the T-maps (one-sample t-test) is $p < 0.05$, and the T-maps are corrected for multiple comparisons by FWE. (a)-(c) Reference DMN component. (d)-(f) DMN component after CPD-SGD. (g)-(i) DMN component after Tucker-SGD. (j)-(l) DMN component after RSCG-TT.

is at the significance level of $p < 0.05$ (FWE-corrected). The T-maps in Fig. 19 and Supplementary Fig. S.2 show significant changes in sensitivity to voxel activation between

the reference RSN and the RSN component after each tensor completion method for the studied functional RSNs, except for RSCG-TT. We further studied the statistical significance of the T-map contrasts between the reference RSN component and RSN component after each tensor completion method by executing post hoc t-tests (FWE-corrected and $p < 0.05$). Supplementary Table S.1 shows that there are statistically significant group differences in sensitivity to voxel activation for both the reference RSN – CPD-SGD RSN and reference RSN – Tucker-SGD RSN contrasts ($p_{\text{corr}} < 0.0001$). Thus, compared to CPD-SGD and Tucker-SGD, RSCG-TT recovers RSN components much better. The good performance of RSCG-TT in preserving the signal of interest is most likely due to its efficiency in the estimation of missing voxels.

VI. DISCUSSIONS

A significant problem in fMRI research involving group studies is the loss of statistical power and the sensitivity of downstream analyses. We have demonstrated that the RSCG-TT algorithm can significantly increase the statistical power and enables the estimation of RSNs even in the presence of a large amount of missing voxels. As a result, RSCG-TT t -statistics had substantially larger the spatial extent of significant clusters. The cluster size of significant clusters was 58% (see Supplementary Fig. S.3) larger on average than the cluster sizes found by the-state-of-the-art completion methods. As a result of the loss of statistical power, the sizes of significant clusters were reduced after the CPD-SGD and Tucker-SGD methods. RSCG-TT has advantages over the-state-of-the-art methods because the TT manifold allows the algorithm to take full advantage of the global correlation structure between modes

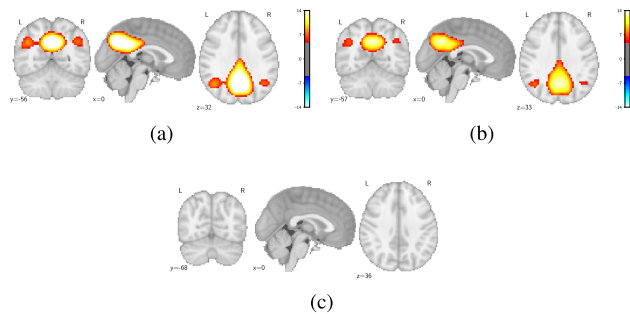


FIGURE 19. Comparison of the reference DMN with the DMN component that uses GICA following different tensor completion methods. The T-maps (the two-sample t-tests) between the reference DMN component and the DMN component after each tensor completion method are shown. The threshold of the T-maps is $p < 0.05$, and the T-maps are corrected for multiple comparisons by FWE. (a) Reference DMN - CPD-SGD DMN. (b) Reference DMN - Tucker-SGD DMN. (c) Reference DMN - RSCG-TT DMN.

of an fMRI tensor. Our study demonstrates that RSCG-TT method is a valuable approach for fMRI data completion that can increase the sensitivity of downstream analyses in identification of important biomarkers such as RSNs.

Detection of the statistically significant effect size in human neuroscience continues to be an ongoing field of active research. The empirical validity of the inferential methods in fMRI is still being studied [91]. To evaluate the significance of the RSNs, we employed the voxel-based inference, a parametric statistical principled correction for multiple comparisons to control the FWE rate. However, due to principled protection from false positives (Type I error), the voxel-based statistics may result in lower sensitivity to voxel activation [92]. Additional methods to control the FWE rate include those making use of cluster-level statistics: cluster-based thresholding [93] and threshold-free cluster enhancement (TFCE) [94]. However, they have their own drawbacks studied in [91], [95]. The voxel-based inference that is used in this work has been widely used, and is desirable due to its ability for rigorous control of false positives in fMRI [96].

A. NUMERICAL PERFORMANCE OF RSCG-TT

We investigated the performance of the RSCG-TT algorithm for the completion of different missing value patterns and as a function of tensor dimensionality. Unlike most works on tensor completion that primarily utilized the RMV pattern, we applied the proposed method to realistic 4D representations in which some selective voxels were missing in spatial ellipsoids across multiple timepoints. To the best of our knowledge, this is the first work applying tensor completion algorithm to the realistic scenario of missing voxels using SMV pattern. The experimental results demonstrate that the proposed RSCG-TT method exhibits rapid convergence rate with the order of multitude differences compared with the performance of the state-of-the-art methods and tensor decompositions such as CPD-SGD and Tucker-SGD. The fast convergence rate can be attributed to the

theoretical foundations of our RSCG-TT algorithm (see Section II-C1) since we minimize the spectral-scaling condition, which implies a lower condition number, and the number of iterations required for the algorithm to converge decreases. The results demonstrate the superior performance of RSCG-TT in terms of two accuracy measures compared with the state-of-the-art SGD-based algorithm such as Adaptive Moment Estimation using CPD and Tucker tensor decompositions. We observe that the mean TCS (see Fig. 14 and Table 10) is in the range of $3.289 - 9.407 \times 10^{-4}$ for the 4D fMRI tensor impacted by the SMV pattern with a realistic temporal missing value rate $MR_T(\%) \in \{5, 10\}\%$. The observed desirable numerical performance of the RSCG-TT can be attributed to several factors, such as the 4D data structure is fully utilized, TT decomposition is employed, and the use of Riemannian manifold. TT decomposition improves the accuracy by taking into account the global correlations across tensor modes. RSCG-TT exhibits the best numerical results in terms of both the TCS and RSE compared with Tucker-SGD, and CPD-SGD has the worst numerical accuracy. The poor performance of CPD-SGD can be explained by the fact that CPD decomposition is better suited for the recovery of unique factors, whereas TT and Tucker decompositions perform best in data completion scenarios.

B. EFFECT OF TENSOR DIMENSIONALITY ON fMRI DATA COMPLETION

We studied the numerical performance of the proposed algorithm with fMRI data represented as 4D, 3D, and 2D tensors. The experimental results suggest that the proposed algorithm performs best when we take the full 4D nature of fMRI data into account with a 4D tensor. Our experiments have shown a reasonable completion threshold confirmed by both quantitative and visual assessments in terms of TCS when it is below 10^{-2} . The 3D tensor representation has a worse performance than 4D tensor representation (see Fig. 8 and Fig. 14). The 2D tensor representation results in a high TCS above 10^{-2} for the majority of missing value rates for both the RMV and SMV patterns (see Fig. 8 and Fig. 14). Our results suggest that the 4D tensor is the natural structure of fMRI data represented as 3D space \times time and us allows to utilize the richness of both spatial and temporal information to its full potential.

C. ROLE OF A MISSING VALUE PATTERN ON THE CAPABILITY OF A FULL DATA RECOVERY

Based on the experimental results presented in Section V-B, we conclude that the RMV pattern is relatively easy to complete, which allows for a small TCS and consequently better numerical performance. RSCG-TT is able to successfully recover the data corrupted by the RMV pattern using 4D tensor format even with a very high missing ratio, 80% of the total number of voxels covered by a brain mask (see Fig. 8 and Fig. 9). As missing voxels are random, the neighborhood voxels contain enough information to infer them, which enables the successful completion of the data

tensor. On the other hand, as confirmed by earlier studies [13], [14], the data corrupted by the SMV pattern are more challenging to complete than those affected by the RMV pattern. The experiments indicate that a simultaneous increase in both $MR_T(\%)$ and $MR_{V_{e^{3D}}}(\%)$ results in a reduction in the numerical performance of RSCG-TT to complete the data tensor (see Fig. 14). In this case, since the missing voxels follow the SMV pattern, the neighborhood voxels do not have sufficient information for the inference, and therefore, the missing voxels for the SMV pattern are more difficult to recover.

D. COMPARISON OF THE COMPUTATIONAL COMPLEXITIES OF RSCG-TT AND THE STATE-OF-THE-ART METHODS

In this section, we compare the computational complexities of RSCG-TT and the state-of-the-art methods, as summarized in Table 13. Here, similar to Section III-B, we assume that R_{\max} is the upper bound on the tensor decompositions under consideration. The CPD-SGD, and Tucker-SGD methods are based on the distributed Adam's SGD, which averages the parameters after each iteration. The update for one entry of the SGD scheme requires $\mathcal{O}(1)$. The number of tensor entries for the CPD and Tucker models is $I = |\Omega|$. Therefore, the total computational complexity for one iteration w.r.t. the free parameters and the number of observed entries in the CPD model is $\mathcal{O}(|\Omega|NR_{\max})$, and the estimate of the Tucker model can be calculated in $\mathcal{O}(|\Omega|NR_{\max} + R_{\max}^N)$ [43], [97].

To complete the entire 4D fMRI tensor, RSCG-TT, on average, requires 3×10^3 ms of CPU time (see Fig. 10b) on a 2.7Hz Xeon processor with 8-16Gb of memory. In Fig. 10b, we compare RSCG-TT and the state-of-the-art completion algorithms w.r.t. the computational time. As we can note from Fig. 10b, the computational time is another key differentiator between the methods. It should be noted from Fig. 10b that RSCG-TT is at least two orders of magnitude faster than CPD-SGD, and Tucker-SGD per unit of time. Moreover, RSCG-TT demonstrates a sharp decrease in the value of the objective function within the same time interval compared with the state-of-the-art methods. The presented results demonstrate the ability of RSCG-TT to approach the solution quicker in terms of both computational measures, i.e., complexity per unit of iteration and the complexity per unit of time. In contrast, the state-of-the-art methods require higher computational time to achieve a reasonable reduction in the value of the model cost function. Therefore, the state-of-the-art algorithms require a significantly higher computational time to approach the solution. The results shown in Fig. 10 can be explained by the properties of the algorithms listed in Table 13. The rate of change in the objective function value of RSCG-TT method is superlinear since it belongs to the family of nonlinear conjugate gradient algorithms [55], [59]. The state-of-the-art methods exhibit a linear time complexity due to their design inherited from the gradient descent scheme [55], [59]. With respect to computational time, the CPD-SGD method is clearly less efficient among the three

methods, requiring the largest computational time, whereas RSCG-TT performs the best. The other notable state-of-the-art tensor completion algorithms are TT-ALS [98], [99], and TMac-TT [17]. The former is based on the Alternating Least Square (ALS) scheme, and the latter is based on parallel matrix factorization using adaptive rank estimation. In each TT-ALS step, all TT-cores but one is kept fixed, and the overall optimization problem is reduced to a small optimization problem of a single core. Compared with the complexity per iteration of RSCG-TT, the TT-ALS procedure uses the fourth power of rank instead of square. While the space complexity is similar to RSCG-TT, the TT-ALS possesses a slower theoretical [100] and practical [99] convergence property.

The advantages of TMac-TT over RSCG-TT include the adaptive rank estimation and bypassing the computationally expensive SVD. Despite the advantages such as adaptive rank estimation, the exponential computational complexity in tensor dimension $\mathcal{O}(3(N-1)I^N R_{\max})$ makes TMac-TT method impractical in large-scale tensor completion scenarios. Although as shown in Table 13, the CPD-SGD method possesses a better computational and space complexity (total space taken by the algorithm with respect to the input size), the determination of canonical tensor rank is NP-hard [101] problem. Second, the best rank- R CPD decomposition might not exist since the space of low-rank tensors is not closed [102]. It is worth noting that the Tucker-SGD model offers a more flexible tensor model. However, the space complexity of the Tucker format scales exponentially in the tensor dimension N , which makes the Tucker-based methods not practical for tensor dimensions $N \geq 4$ due to the curse of dimensionality.

From this perspective, our assessment suggests that TT-ALS, TMac-TT, CPD-SGD, and Tucker-SGD either provide suboptimal computational complexity or may not be able to deal with large-scale high-order tensors. Finally, as illustrated in Table 13, Fig. 10, and above considerations, RSCG-TT is theoretically and practically superior to the state-of-the-art algorithms, and it is computationally preferable for the data completion tasks.

E. SMV PATTERN IN PRACTICAL SCENARIOS

In this section, we discuss a practical application of the RSCG-TT method when fMRI data are analyzed in the complex-valued domain. fMRI data are natively acquired as complex-valued spatiotemporal images; however, most studies have only focused only on the magnitude data in fMRI data. Several studies have shown that employing fMRI in their native form with both the magnitude and phase improves the sensitivity of analyses [3], [3], [72], [74]–[77]. Since phase images pose a unique challenge, a typical solution is to remove noisy regions from fMRI data prior to analyses. Consequently, regions are eliminated through thresholding and edge reducing techniques or by using QMPD methods [3], [72], [74]–[77]. The QMPD maps are used to generate quality masks to exclude voxels that should not be further analyzed,

TABLE 13. Summary of computational complexities of RSCG-TT and the state-of-the-art tensor factorization algorithms. For simplicity we assume that the length of every mode is equal to I and R_{\max} is the upper bound on the considered tensor decompositions.

Algorithm	Computational complexity	Space complexity	Convergence rate	Rank computation complexity	Space of low-rank tensors
RSCG-TT	$\mathcal{O}(NIR_{\max}^3 + N \Omega R_{\max}^2)$	$\mathcal{O}(NIR_{\max}^2)$	Superlinear	Polynomial	Closed
CPD-SGD	$\mathcal{O}(\Omega NR_{\max})$	$\mathcal{O}(NIR_{\max})$	Linear	NP-hard	Not closed
Tucker-SGD	$\mathcal{O}(\Omega NR_{\max} + R_{\max}^N)$	$\mathcal{O}(NIR_{\max}) + R_{\max}^N$	Linear	Polynomial	Closed
TT-ALS	$\mathcal{O}(N \Omega R_{\max}^4)$	$\mathcal{O}(NIR_{\max}^2)$	Linear	Polynomial	Closed
TMac-TT	$\mathcal{O}(3(N-1)I^N R_{\max})$	$\mathcal{O}(NIR_{\max}^2)$	Linear	Polynomial	Closed

as described in Section IV-E. The SMV pattern studied in this paper aims to demonstrate one of the applications of the RSCG-TT method when 3D brain regions are excluded from the analysis. Here, we highlight the importance of our findings, demonstrating the viability of SMV pattern fMRI data completion, which can be used in the broader context of the MNAR voxel pattern. Finally, we would like to conclude that, indeed, taking the 4D structure of fMRI data into account provides significant performance gains and is very exciting on its own. We hope this will help encourage and accelerate research in the very promising area of tensors for fMRI data not only for data completion but for the development of other tools for the analyzing fMRI data when computational challenges might be even more significant.

VII. CONCLUSION

We introduce a novel approach to the fMRI data completion problem based on TT decomposition along with a corresponding algorithm for its solution. We develop Riemannian nonlinear spectral conjugate gradient method called RSCG-TT to solve a higher-order fMRI data completion in the form of a 4D tensor. The RSCG-TT algorithm exploits the second-order information and offers additional guardrails to ensure that computational performance is coupled with numerical stability. Since we deal with high-order datasets, we extend our method to a TT format, which provides scalable numerical algebra operations and hierarchical compressed representations of structurally diverse datasets. Exploiting the Riemannian structure and the efficient utilization of the covariance information by TT decomposition boosts the numerical performance significantly when the proposed method is compared with state-of-the-art SGD-based methods, such as Tucker-SGD and CPD-SGD. Parallelizing tensor computations can further improve the execution time of RSCG-TT. RSCG-TT can benefit from the use of sparse data structures, and graphical processing units (GPU) specialized in high-performance computing (HPC) to reduce the elapsed time. The topic of shortening the computational time of RSCG-TT is the subject of future research.

One of the significant advantages of our approach is that we suggest viewing fMRI data as 4D spatiotemporal hypervolumes observed during a scanning session. This view takes into account the natural representation of fMRI data in the form of a fourth-order tensor represented as 3D space \times time. As demonstrated by the experiments in Section V, representing a 4D tensor as a TT tensor allows us to fully exploit the

global structure of fMRI data and introduces a joint learning scheme, which captures the intrinsic relationship between the spatial and temporal modes. We consider the practical use of the RSCG-TT algorithm when some select brain volumes of a 4D fMRI scan are affected by the exclusion of brain regions that resemble the MNAR missing voxel pattern. Our experiments suggest that the RSCG-TT algorithm can estimate the missing brain voxels affected by the MNAR pattern and provides a reliable recovery even in the areas with very low activation thresholds and Z-scores below two times the standard deviation. Studies with other lower-dimensional representations demonstrate that taking the full 4D structure of fMRI data into account provides significant gains compared to when 3D and 2D representations of the data are considered. We evaluated the ability of RSCG-TT to recover the RSN components after data completion. Our results suggest that RSCG-TT provides advantages in retaining the RSN components compared to the state-of-the-art tensor completion methods.

Another promising direction in data completion research includes fMRI data completion in the native k -space. The k -space intrinsically possesses very desirable properties, such as fMRI image sparsity in wavelet bases and the incoherence between Fourier and inverse wavelet bases. These properties might allow one to significantly reduce the order of a model, i.e., the TT-rank, and to employ nonlinear data completion through a nonsmooth formulation of the optimization problem.

Future potential studies include a number of exciting directions, such as simultaneous estimations of an optimal multilinear TT-rank and robust tensor completion, complex-valued tensor completion, incorporating nonparametric cluster-level statistics, and working directly in the k -space with a nonsmooth objective function.

APPENDIX A LIST OF ABBREVIATIONS

2D	two-dimensional.
3D	three-dimensional.
4D	four-dimensional.
ARD	Automatic Relevance Determination.
BGFS	Broyden-Fletcher-Goldfarb-Shanno.
CPD	Canonical Polyadic Decomposition.
CPD-SGD	Stochastic Gradient Descent based on CPD.
EM	Expectation Minimization.

fMRI	Functional Magnetic Resonance Imaging.
MAR	Missing at Random.
MBFGS	Modified Broyden-Fletcher-Goldfarb-Shanno.
MNAR	Missing not at Random.
MRI	Magnetic Resonance Imaging.
RMV	Random Missing Values.
RCT	Riemannian Curvature Tensor.
RSCG-TT	Riemannian nonlinear Spectral Conjugate Gradient via Tensor Train.
RSE	Relative Square Error.
SCG	Scaled Conjugate Gradient.
SGD	Stochastic Gradient Descent.
SMV	Structural Missing Values.
TCS	Tensor Completion Score.
TCS_Z	Tensor Completion Score with Z-score above two times the standard deviation.
TT	Tensor Train.
Tucker-SGD	Stochastic Gradient Descent based on Tucker Decomposition.
RSN	Resting State Networks.

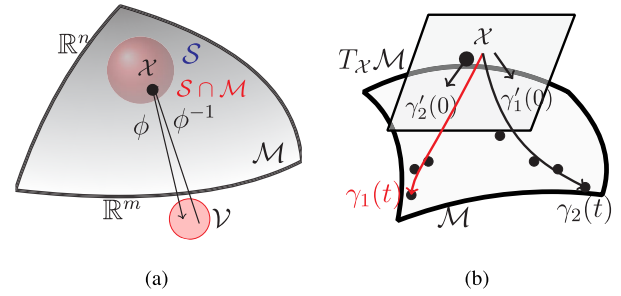


FIGURE 20. Differential manifold concepts. (a) A smooth m -dimensional submanifold of \mathbb{R}^n . A coordinate chart ϕ maps a neighborhood $\mathcal{S} \cap \mathcal{M}$ of a point $\mathcal{X} \in \mathcal{M}$ to a subset $\mathcal{V} \in \mathbb{R}^m$. (b) A tangent space $T_{\mathcal{X}}\mathcal{M}$ of an embedded submanifold \mathcal{M} , $\mathcal{X} \in \mathcal{M}$. The tangent vectors γ'_1 and γ'_2 are realizations of the two curves γ_1 and γ_2 .

**APPENDIX B
PRELIMINARIES ON RIEMANNIAN OPTIMIZATION**

Definition 1 (Manifold): A real n -dimensional manifold is a topological space \mathcal{M} , where every point $\mathcal{X} \in \mathcal{M}$ has a neighbourhood homeomorphic to Euclidean space \mathbb{R}^n .

Definition 2 (Smooth Manifold [103]): A subset $\mathcal{M} \in \mathbb{R}^n$ is an m -dimensional smooth manifold of \mathbb{R}^n if each $\mathcal{X} \in \mathcal{M}$ has an open neighbourhood $\mathcal{S} \in \mathbb{R}^n$ that $\mathcal{S} \cap \mathcal{M}$ is diffeomorphic $\mathcal{V} \in \mathbb{R}^m$. A diffeomorphism $\phi : \mathcal{S} \cap \mathcal{M} \rightarrow \mathcal{V}$ is called the coordinate chart of \mathcal{M} , while its inverse $\phi^{-1} : \mathcal{V} \rightarrow \mathcal{S} \cap \mathcal{M}$ is called a parametrization of $\mathcal{S} \cap \mathcal{M}$. We illustrate the concept of a smooth manifold in Fig. 20(a).

Definition 3 (A Riemannian Manifold $(\mathcal{M}, \mathfrak{g})$): (B.1) is a real smooth manifold \mathcal{M} equipped with a Riemannian metric \mathfrak{g} .

Definition 4 (A Riemannian Metric \mathfrak{g} (B.2)): on a smooth manifold \mathcal{M} is a smoothly chosen inner product $\mathfrak{g}_{\mathcal{X}} : T_{\mathcal{X}}\mathcal{M} \times T_{\mathcal{X}}\mathcal{M} \rightarrow \mathbb{R}$ on each of the tangent spaces $T_{\mathcal{X}}\mathcal{M}$ of \mathcal{M} such that:

- 1) $\mathfrak{g}(\xi, \eta) = \mathfrak{g}(\eta, \xi) \forall \xi, \eta \in T_{\mathcal{X}}\mathcal{M}$;
- 2) $\mathfrak{g}(\xi, \xi) \geq 0 \forall \xi \in T_{\mathcal{X}}\mathcal{M}$;
- 3) $\mathfrak{g}(\xi, \xi) = 0$ if and only if $\xi = 0$.

Definition 5 (Tangent Vector, Space, and Bundle): A vector ξ is called a tangent vector of \mathcal{M} at a point $\mathcal{X} \in \mathcal{M}$ if there is a small curve $\gamma : \mathbb{R} \rightarrow \mathcal{M}$ such that

$$\gamma(0) = x, \gamma' = \lim_{t \rightarrow 0} \frac{\gamma(t) - \gamma(0)}{t} = \xi$$

The set of tangent vectors of \mathcal{M} at x forms a tangent space of \mathcal{M} at \mathcal{X} : $T_{\mathcal{X}}\mathcal{M} := \{\gamma'(0) | \gamma : \mathbb{R} \rightarrow \mathcal{M} \text{ smooth}, \gamma(0) = x$ (B.3). The tangent bundle is a disjoint union of all tangent spaces,

$$T\mathcal{M} := \bigcup_{\mathcal{X} \in \mathcal{M}} \{\mathcal{X}\} \times T_{\mathcal{X}}\mathcal{M}, \quad (\text{B.4})$$

where $\{(\mathcal{X}, \xi) \in \mathbb{R}^{I_1 I_2 \dots I_N} \times \mathbb{R}^{I_1 I_2 \dots I_N} : \mathcal{X} \in \mathcal{M}, \xi \in T_{\mathcal{X}}\mathcal{M}\}$. The concept of a tangent space is illustrated in Fig. 20(b). We restrict the Euclidean inner product of two tensors to the tangent bundle $\mathfrak{g}_{\mathcal{X}} : T_{\mathcal{X}}\mathcal{M} \times T_{\mathcal{X}}\mathcal{M}$,

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \text{tr}(\mathcal{X}^T \mathcal{Y}) \text{ with } \mathcal{X}, \mathcal{Y} \in \mathbb{R}^{I_1 I_2 \dots I_N}$$

Therefore, we can convert a smooth manifold \mathcal{M} into a Riemannian manifold with a Riemannian metric

$$\mathfrak{g}_{\mathcal{X}} = \langle \xi, \eta \rangle = \text{tr}(\xi^T \eta) \quad (\text{B.5})$$

where $\mathcal{X} \in \mathcal{M}$ and $\xi, \eta \in T_{\mathcal{X}}\mathcal{M}$, and tangent vectors ξ, η are tensors $\in \mathbb{R}^{I_1 I_2 \dots I_N}$. The inner product defined by the Riemannian metric (B.2) induces the following norm on $T_{\mathcal{X}}\mathcal{M}$:

$$\|\xi\| = \sqrt{\mathfrak{g}_{\mathcal{X}}(\xi, \xi)}. \quad (\text{B.6})$$

We further define the orthogonal projection of tensor $\mathcal{X} \in \mathbb{R}^{I_1 I_2 \dots I_N}$ onto $T_{\mathcal{X}}\mathcal{M}$ as

$$P_{T_{\mathcal{X}}\mathcal{M}} : \mathbb{R}^{I_1 I_2 \dots I_N} \rightarrow T_{\mathcal{X}}\mathcal{M}$$

Definition 6 (Retraction): The retraction $R_{\mathcal{X}}$ is a smooth mapping from a tangent vector $\xi \in T_{\mathcal{X}}\mathcal{M}$ to a point $\mathcal{X} \in \mathcal{M}$ such that

$$R_{\mathcal{X}} : T_{\mathcal{X}}\mathcal{M} \rightarrow \mathcal{M}, \quad (\text{B.7})$$

and it satisfies the local rigidity condition [37].

By its properties, it guarantees a natural way to move on the manifold along a search direction. The retraction R can be computed by the projection operator with the rank truncation by TT-SVD [44] as follows:

$$R_{\mathcal{X}}(\xi) \rightarrow P_{\mathbf{r}}^{\text{TT-SVD}}(\mathcal{X} + \xi), \quad P_{\mathbf{r}}^{\text{TT-SVD}} \rightarrow \mathbb{R}^{I_1 \times \dots \times I_N}. \quad (\text{B.8})$$

Definition 7 (Vector Transport): A vector transport on manifold \mathcal{M} is a smooth mapping that transports a tangent vector $\xi \in T_{\mathcal{X}}\mathcal{M}$ at $\mathcal{X} \in \mathcal{M}$ (B.9) to a vector in the tangent space at a point $\mathcal{Y} \in T_{\mathcal{Y}}\mathcal{M}$.

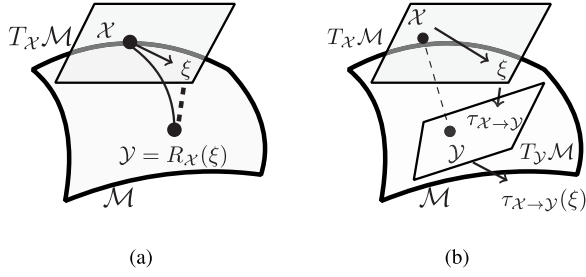


FIGURE 21. Geometrical concepts within the framework of Riemannian optimization. (a) Retraction mapping $R_{\mathcal{X}} : T_{\mathcal{X}}\mathcal{M} \rightarrow \mathcal{M}$. (b) Vector transport mapping $\tau_{\mathcal{X} \rightarrow \mathcal{Y}} : T_{\mathcal{X}}\mathcal{M} \rightarrow T_{\mathcal{Y}}\mathcal{M}$.

It was shown in [37] that for the embedded submanifold of $\mathbb{R}^{I_1 \times \dots \times I_N}$, the vector transport is given by the orthogonal projection of $P_{T_{\mathcal{Y}}\mathcal{M}}\xi$ onto the tangent space $T_{\mathcal{Y}}\mathcal{M}$ as follows:

$$\tau_{\mathcal{X} \rightarrow \mathcal{Y}} : T_{\mathcal{X}}\mathcal{M} \rightarrow T_{\mathcal{Y}}\mathcal{M}, \quad (\text{B.10})$$

$$\xi_{\mathcal{X}} \rightarrow P_{T_{\mathcal{Y}}\mathcal{M}}\xi. \quad (\text{B.11})$$

The concept of vector transport $\tau_{\mathcal{X} \rightarrow \mathcal{Y}}$ is depicted in Fig. 21(b).

Definition 8: (Riemannian gradient). Given a smooth real function on $f : \mathcal{M} \rightarrow \mathbb{R}^{I_1 \times \dots \times I_N}$, $\mathcal{X} \in \mathcal{M}$ on a Riemannian manifold \mathcal{M} , its gradient $\text{grad}f(x)$ is defined as a unique tangent vector in $T_{\mathcal{X}}\mathcal{M}$ satisfying

$$\langle \text{grad}f(x), \xi \rangle_x = Df(x)[\xi], \quad \forall \xi \in T_{\mathcal{X}}\mathcal{M}, \quad (\text{B.12})$$

where we denote the directional derivative as Df .

Since the TT-manifold \mathcal{M}_r is embedded in Euclidean space $\mathbb{R}^{I_1 \times \dots \times I_N}$ [47], the Riemannian gradient for TT decomposition is computed as the orthogonal projection of the Euclidean gradient onto tangent space (B.3) [37]

$$\text{grad}f(\mathcal{X}) = P_{T_{\mathcal{X}}\mathcal{M}_r}(\nabla f(\mathcal{X})), \quad (\text{B.13})$$

where $f : \mathbb{R}^{I_1 \times \dots \times I_N} \rightarrow \mathbb{R}$ is the cost function with Euclidean gradient $\nabla f(\mathcal{X})$ and $\mathcal{X} \in \mathcal{M}_r$.

Definition 9 (Geodesic): A smooth curve $\gamma : I \rightarrow \mathcal{M}$, where ($I \in \mathbb{R}$ is any interval), is a geodesic if $\gamma''(\alpha) = \nabla_{\gamma'}\gamma'(\alpha) = 0 \forall \alpha \in I$ (B.14).

A. RIEMANNIAN MANIFOLD LEARNING FOR TENSOR DECOMPOSITION IN A TT FORMAT

Riemannian optimization on tensor manifolds is a growing field, and it is a generalization of the theory and algorithms from unconstrained Euclidean optimization to problems on smooth manifolds [33], [37]. It consists of optimizing a tensor valued function on a curved manifold instead of Euclidean space. In the Riemannian setting, the Euclidean notion of straight lines and ordinary differentiation is replaced with geodesics and covariant differentiation. Mathematically speaking, Riemannian tensor optimization considers finding an optimum of a scalar cost function of a tensor variable $f(\mathcal{X})$ defined on a smooth differentiable manifold,

\mathcal{M} in (B.1), equipped with a Riemannian metric structure \mathfrak{g} , given in (B.2), that is both symmetric and positive definite [104], [105] as:

$$\underset{\mathcal{X} \in \mathcal{M}}{\text{argmin}} f(\mathcal{X}). \quad (\text{B.15})$$

1) RIEMANNIAN OPTIMIZATION CONCEPTS

The continuous optimization of Riemannian manifolds requires us to construct a descent search curve γ such that:

$$\gamma'(\alpha) = -\text{grad}f(\gamma(\alpha)) \quad \forall \alpha, \quad (\text{B.16})$$

$$\gamma(\alpha) := R_{\mathcal{X}}(\alpha\eta), \quad (\text{B.17})$$

where $\mathcal{X} \in \mathcal{M}$, a search direction $\eta \in T_{\mathcal{X}}\mathcal{M}$, $T_{\mathcal{X}}\mathcal{M}$ is the tangent space in (B.3) to the manifold at the point \mathcal{X} , and α is the scalar step size following the geodesics (B.14) along the curve γ . The interpretation of moving along the curve γ can be seen as moving in the search direction η , while every iteration \mathcal{X}_k is constrained to the manifold \mathcal{M} . In practice, since it is not feasible to compute geodesics in a closed form, the computation of geodesics is replaced with a smooth mapping, called retraction $R_{\mathcal{X}} : T_{\mathcal{X}}\mathcal{M} \rightarrow \mathcal{M}$ (B.7). The concept of retraction is illustrated in Fig. 21(a). Additionally, to compute the new search direction, we have to linearly combine the current gradient vector and the previous search direction. Technically, we need to transport the information from one tangent space $T_{\mathcal{X}}\mathcal{M}$ to another tangent space $T_{\mathcal{Y}}\mathcal{M}$ using the vector transport $\tau_{\mathcal{X} \rightarrow \mathcal{Y}}$ in (B.9). An intuitive drawing of a vector transport is depicted in Fig. 21(b). Finally, with the retraction R , the updating formula for the curvilinear search is given as

$$\mathcal{X}_{k+1} = R_{\mathcal{X}_k}(\alpha_k \eta_k), \quad (\text{B.18})$$

where the step size α_k is computed to satisfy the Wolfe conditions [59] and η_k is the search direction satisfying the descent condition [55].

2) MANIFOLD STRUCTURE FOR TT DECOMPOSITION

In [47], the authors show that the set \mathcal{M}_r of the tensors of fixed multilinear rank $\mathbf{r} = (R_0, R_1, R_2, \dots, R_N)$ in TT format form a smooth embedded submanifold of $\mathbb{R}^{I_1 I_2 \dots I_N}$.

$$\mathcal{M}_r = \{\mathcal{X} \in \mathbb{R}^{I_1 I_2 \dots I_N} \mid \text{rank}_{\text{TT}}(\mathcal{X}) = \mathbf{r}\}, \quad (\text{B.19})$$

where the dimension of \mathcal{M}_r is given by

$$\dim(\mathcal{M}_r) = \sum_{n=1}^N R_{n-1} I_n R_n - \sum_{n=1}^{N-1} R_n^2, \quad (\text{B.20})$$

and the tensor \mathcal{X} allows TT decomposition (1) with multilinear TT rank \mathbf{r} in (2).

We use the inner product in (4) defined for TT decomposition as a metric on \mathcal{M}_r . Therefore, when we equip \mathcal{M}_r with this metric, \mathcal{M}_r becomes a Riemannian manifold $(\mathcal{M}_r, \mathfrak{g})$. As a result, the Riemannian structure of the TT manifold allows us to use any method from the optimization on Riemannian manifolds with the Riemannian gradient defined in (B.13).

ACKNOWLEDGMENT

The authors would like to thank Dr. Vince Calhoun for providing the fMRI COBRE dataset used in this study and Dr. Evrim Acar for the useful comments and feedback.

REFERENCES

- [1] J. T. Devlin, R. P. Russell, M. H. Davis, C. J. Price, J. Wilson, H. E. Moss, P. M. Matthews, and L. K. Tyler, "Susceptibility-induced loss of signal: Comparing PET and fMRI on a semantic task," *NeuroImage*, vol. 11, no. 6, pp. 589–600, Jun. 2000.
- [2] K. I. Vaden, M. Gebregziabher, S. E. Kuchinsky, and M. A. Eckert, "Multiple imputation of missing fMRI data in whole brain analysis," *NeuroImage*, vol. 60, no. 3, pp. 1843–1855, Apr. 2012.
- [3] P. A. Rodriguez, N. M. Correa, T. Eichele, V. D. Calhoun, and T. Adali, "Quality map thresholding for de-noising of complex-valued fMRI data and its application to ICA of fMRI," *J. Signal Process. Syst.*, vol. 65, no. 3, pp. 497–508, Dec. 2009.
- [4] P. A. Rodriguez, V. D. Calhoun, and T. Adali, "De-noising, phase ambiguity correction and visualization techniques for complex-valued ICA of group fMRI data," *Pattern Recognit.*, vol. 45, no. 6, pp. 2050–2063, Jun. 2012.
- [5] T. Adali and V. D. Calhoun, "Complex ICA of brain imaging data," *IEEE Signal Process. Mag.*, vol. 24, no. 5, p. 136, Sep. 2007.
- [6] R. J. Little and D. B. Rubin, *Statistical Analysis With Missing Data*, vol. 333. Hoboken, NJ, USA: Wiley, 2014.
- [7] J. G. Ojemann, E. Akbudak, A. Z. Snyder, R. C. McKinstry, M. E. Raichle, and T. E. Conturo, "Anatomic localization and quantitative analysis of gradient refocused echo-planar fMRI susceptibility artifacts," *NeuroImage*, vol. 6, no. 3, pp. 156–167, Oct. 1997.
- [8] H. Togo, J. Rokicki, K. Yoshinaga, T. Hisatsune, H. Matsuda, N. Haga, and T. Hanakawa, "Effects of field-map distortion correction on resting state functional connectivity MRI," *Frontiers Neurosci.*, vol. 11, p. 656, Dec. 2017.
- [9] J. D. Power, K. A. Barnes, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen, "Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion," *NeuroImage*, vol. 59, no. 3, pp. 2142–2154, 2012.
- [10] J. D. Power, M. Plitt, P. Kundu, P. A. Bandettini, and A. Martin, "Temporal interpolation alters motion in fMRI scans: Magnitudes and consequences for artifact detection," *PLoS ONE*, vol. 12, no. 9, Sep. 2017, Art. no. e0182939.
- [11] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010.
- [12] M. Tao and X. Yuan, "Recovering low-rank and sparse components of matrices from incomplete and noisy observations," *SIAM J. Optim.*, vol. 21, no. 2, pp. 57–81, 2011.
- [13] G. Tomasi and R. Bro, "PARAFAC and missing values," *Chemometric Intell. Lab. Syst.*, vol. 75, no. 2, pp. 163–180, Feb. 2005.
- [14] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, "Scalable tensor factorizations for incomplete data," *Chemometrics Intell. Lab. Syst.*, vol. 106, no. 1, pp. 41–56, 2011.
- [15] B. Vandereycken, "Low-rank matrix completion by Riemannian optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1214–1236, Jun. 2013.
- [16] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2012.
- [17] J. A. Bengua, H. N. Phien, H. D. Tuan, and M. N. Do, "Efficient tensor completion for color image and video recovery: Low-rank tensor train," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2466–2479, May 2017.
- [18] S. Ferdowsi and V. Abolghasemi, "Simultaneous BOLD detection and incomplete fMRI data reconstruction," *Med. Biol. Eng. Comput.*, vol. 56, no. 4, pp. 599–610, Apr. 2018.
- [19] M. Chiew, S. M. Smith, P. J. Koopmans, N. N. Graedel, T. Blumensath, and K. L. Miller, "k-t FASTER: Acceleration of functional MRI data acquisition using low rank constraints," *Magn. Reson. Med.*, vol. 74, no. 2, pp. 353–364, 2015.
- [20] P. Aggarwal and A. Gupta, "Accelerated fMRI reconstruction using matrix completion with sparse recovery via split Bregman," *Neurocomputing*, vol. 216, pp. 319–330, Dec. 2016.
- [21] V. Singh, A. H. Tewfik, and D. B. Renshaw, "Under-sampled functional MRI using low-rank plus sparse matrix decomposition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 897–901.
- [22] T. Yokota, B. Erem, S. Guler, S. K. Warfield, and H. Hontani, "Missing slice recovery for tensors using a low-rank model in embedded space," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8251–8259.
- [23] D. Banco, S. Aeron, and W. S. Hoge, "Sampling and recovery of MRI data using low rank tensor models," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 448–452.
- [24] C. I. Kanatsoulis, N. D. Sidiropoulos, M. Akçakaya, and X. Fu, "Regular sampling of tensor signals: Theory and application to fMRI," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 2932–2936.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] M. Boussé, O. Debals, and L. De Lathauwer, "A tensor-based method for large-scale blind source separation using segmentation," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 346–358, Jan. 2016.
- [29] K. H. Madsen, N. W. Churchill, and M. Mørup, "Quantifying functional connectivity in multi-subject fMRI data using component models," *Hum. Brain Mapping*, vol. 38, no. 2, pp. 882–899, Feb. 2017.
- [30] I. Belyaeva, Q. Long, and T. Adali, "Inexact proximal conjugate sub-gradient algorithm for fMRI data completion," in *Proc. 28th Eur. Signal Process. Conf. (EUSIPCO)*, Jan. 2021, pp. 1025–1029.
- [31] P. T. Fletcher, S. Venkatasubramanian, and S. Joshi, "The geometric median on Riemannian manifolds with application to robust atlas estimation," *NeuroImage*, vol. 45, no. 1, pp. S143–S152, Mar. 2009.
- [32] S. Gerber, T. Tasdizen, P. T. Fletcher, S. Joshi, and R. Whitaker, "Manifold modeling for brain population analysis," *Med. Image Anal.*, vol. 14, no. 5, pp. 643–653, Oct. 2010.
- [33] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *Int. J. Comput. Vis.*, vol. 66, no. 1, pp. 41–66, 2006.
- [34] K. Slavakis, S. Salsabilian, D. S. Wack, S. F. Muldoon, H. E. Baidoo-Williams, J. M. Vettel, M. Cieslak, and S. T. Grafton, "Clustering brain-network time series by Riemannian geometry," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 3, pp. 519–533, Sep. 2017.
- [35] A. Qiu, A. Lee, M. Tan, and M. K. Chung, "Manifold learning on brain functional networks in aging," *Med. Image Anal.*, vol. 20, no. 1, pp. 52–60, Feb. 2015.
- [36] Q. Zhao, D. Kwon, and K. M. Pohl, "A Riemannian framework for longitudinal analysis of resting-state functional connectivity," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 145–153.
- [37] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton Univ. Press, 2009.
- [38] S. T. Smith, "Geometric optimization methods for adaptive filtering," Ph.D. dissertation, Division Appl. Sci., Harvard Univ., Cambridge, MA, USA, 1993.
- [39] S. I. Amari, "Natural gradient works efficiently in learning," *Neural Comput.*, vol. 10, no. 2, pp. 251–276, 1998.
- [40] T.-T. Yao, Z.-J. Bai, Z. Zhao, and W.-K. Ching, "A Riemannian fletcher-reeves conjugate gradient method for doubly stochastic inverse eigenvalue problems," *SIAM J. Matrix Anal. Appl.*, vol. 37, no. 1, pp. 215–234, Jan. 2016.
- [41] M. Steinlechner, "Riemannian optimization for high-dimensional tensor completion," *SIAM J. Sci. Comput.*, vol. 38, no. 5, pp. S461–S484, Jan. 2016.
- [42] COBRE. *The Center for Biomedical Research Excellence*. Accessed: Feb. 27, 2019. [Online]. Available: http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html
- [43] A. Cichocki, N. Lee, I. Oseledets, A.-H. Phan, Q. Zhao, and D. P. Mandic, "Tensor networks for dimensionality reduction and large-scale optimization: Part I low-rank tensor decompositions," *Found. Trends Mach. Learn.*, vol. 9, nos. 4–5, pp. 249–429, 2016.

- [44] I. V. Oseledets, "Tensor-train decomposition," *SIAM J. Sci. Comput.*, vol. 33, no. 5, pp. 2295–2317, Jan. 2011.
- [45] D. Savostyanov and I. Oseledets, "Fast adaptive interpolation of multi-dimensional arrays in tensor train format," in *Proc. Int. Workshop Multi-dimensional (nD) Syst.*, Sep. 2011, pp. 1–8.
- [46] A. Cichocki, S. Cruces, and S.-I. Amari, "Log-determinant divergences revisited: Alpha-beta and gamma log-det divergences," *Entropy*, vol. 17, no. 5, pp. 2988–3034, 2015.
- [47] S. Holtz, T. Rohwedder, and R. Schneider, "On manifolds of tensors of fixed TT-rank," *Numerische Math.*, vol. 120, no. 4, pp. 701–731, Apr. 2012.
- [48] L. Grasedyck, D. Kressner, and C. Tobler, "A literature survey of low-rank tensor approximation techniques," *GAMM-Mitteilungen*, vol. 36, no. 1, pp. 53–78, Aug. 2013.
- [49] A. Uschmajew and B. Vandereycken, "The geometry of algorithms using hierarchical tensors," *Linear Algebra Appl.*, vol. 439, no. 1, pp. 133–166, 2013.
- [50] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
- [51] A. Cichocki, R. Zdunek, A. H. Phan, and S.-I. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation*. Hoboken, NJ, USA: Wiley, 2009.
- [52] W. Y. Cheng and D. H. Li, "Spectral scaling BFGS method," *J. Optim. Theory Appl.*, vol. 146, no. 2, pp. 305–319, Aug. 2010.
- [53] D.-H. Li and M. Fukushima, "A modified BFGS method and its global convergence in nonconvex minimization," *J. Comput. Appl. Math.*, vol. 129, nos. 1–2, pp. 15–35, Apr. 2001.
- [54] E. G. Birgin and J. M. Martínez, "A spectral conjugate gradient method for unconstrained optimization," *Appl. Math. Optim.*, vol. 43, pp. 117–128, Jan. 2001.
- [55] W. W. Hager and H. Zhang, "A new conjugate gradient method with guaranteed descent and an efficient line search," *SIAM J. Optim.*, vol. 16, no. 1, pp. 170–192, Sep. 2005.
- [56] Y.-H. Dai and C.-X. Kou, "A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolfe line search," *SIAM J. Optim.*, vol. 23, no. 1, pp. 296–320, Jan. 2013.
- [57] D. F. Shanno, "Conjugate gradient methods with inexact searches," *Math. Oper. Res.*, vol. 3, no. 3, pp. 244–256, Aug. 1978.
- [58] L. Nazareth, "A relationship between the BFGS and conjugate gradient algorithms and its implications for new algorithms," *SIAM J. Numer. Anal.*, vol. 16, no. 5, pp. 794–800, 1979.
- [59] J. Nocedal and S. Wright, *Numerical Optimization*. New York, NY, USA: Springer, 2006.
- [60] S. Babaie-Kafaki and Z. Aminifard, "Two-parameter scaled memoryless BFGS methods with a nonmonotone choice for the initial step length," *Numer. Algorithms*, vol. 82, no. 4, pp. 1345–1357, Dec. 2019.
- [61] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA J. Numer. Anal.*, vol. 8, no. 1, pp. 141–148, 1988.
- [62] M. Raydan, "The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem," *SIAM J. Optim.*, vol. 7, no. 1, pp. 26–33, 1997.
- [63] S. Babaie-Kafaki, "On optimality of the parameters of self-scaling memoryless quasi-Newton updating formulae," *J. Optim. Theory Appl.*, vol. 167, no. 1, pp. 91–101, Oct. 2015.
- [64] W. Zhou and L. Zhang, "Global convergence of the nonmonotone MBFGS method for nonconvex unconstrained minimization," *J. Comput. Appl. Math.*, vol. 223, no. 1, pp. 40–47, Jan. 2009.
- [65] G. Yuan, Z. Sheng, B. Wang, W. Hu, and C. Li, "The global convergence of a modified BFGS method for nonconvex functions," *J. Comput. Appl. Math.*, vol. 327, pp. 274–294, Jan. 2018.
- [66] P. Wolfe, "Convergence conditions for ascent methods," *SIAM Rev.*, vol. 11, no. 2, pp. 226–235, 1969.
- [67] A. Novikov, P. Izmailov, V. Khilkin, M. Figurnov, and I. Oseledets, "Tensor train decomposition on TensorFlow (T3F)," *J. Mach. Learn. Res.*, vol. 21, no. 30, pp. 1–7, 2020. [Online]. Available: <http://jmlr.org/papers/v21/18-008.html>
- [68] A. Scott, W. Courtney, D. Wood, R. De la Garza, S. Lane, R. Wang, M. King, J. Roberts, J. A. Turner, and V. D. Calhoun, "COINS: An innovative informatics and neuroimaging tool suite built for large heterogeneous datasets," *Frontiers NeuroInform.*, vol. 5, p. 33, Dec. 2011.
- [69] W. W. Hager and H. Zhang, "Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranteed descent," *ACM Trans. Math. Softw.*, vol. 32, no. 1, pp. 113–137, 2006.
- [70] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [71] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*. [Online]. Available: <http://arxiv.org/abs/1609.04747>
- [72] C. Hutton, A. Bork, O. Josephs, R. Deichmann, J. Ashburner, and R. Turner, "Image distortion correction in fMRI: A quantitative evaluation," *NeuroImage*, vol. 16, no. 1, pp. 217–240, May 2002.
- [73] P. Jezzard and R. S. Balaban, "Correction for geometric distortion in echo planar images from B0 field variations," *Magn. Reson. Med.*, vol. 34, no. 1, pp. 65–73, Jul. 1995.
- [74] D. C. Ghiglia and M. D. Pritt, *Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software*, 1st ed. Hoboken, NJ, USA: Wiley, Apr. 1998.
- [75] S. Chavez, Q.-S. Xiang, and L. An, "Understanding phase maps in MRI: A new outline phase unwrapping method," *IEEE Trans. Med. Imag.*, vol. 21, no. 8, pp. 966–977, Aug. 2002.
- [76] R. Deichmann, J. A. Gottfried, C. Hutton, and R. Turner, "Optimized EPI for fMRI studies of the orbitofrontal cortex," *NeuroImage*, vol. 19, no. 2, pp. 430–441, Jun. 2003.
- [77] N. Weiskopf, C. Hutton, O. Josephs, R. Turner, and R. Deichmann, "Optimized EPI for fMRI studies of the orbitofrontal cortex: Compensation of susceptibility-induced gradients in the readout direction," *Magn. Reson. Mater. Phys., Biol. Med.*, vol. 20, no. 1, p. 39, Feb. 2007.
- [78] M. Mørup and L. K. Hansen, "Automatic relevance determination for multi-way models," *J. Chemometrics*, vol. 23, nos. 7–8, pp. 352–363, 2009.
- [79] J. L. Castellanos, S. Gómez, and V. Guerra, "The triangle method for finding the corner of the L-curve," *Appl. Numer. Math.*, vol. 43, no. 4, pp. 359–373, Dec. 2002.
- [80] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer. (2016). *Tensorlab 3.0*. [Online]. Available: <http://www.tensorlab.net>
- [81] J. Kossaifi, Y. Panagakis, A. Anandkumar, and M. Pantic, "TensorLy: Tensor learning in Python," *J. Mach. Learn. Res.*, vol. 20, no. 1, pp. 925–930, 2019.
- [82] V. D. Calhoun, T. Adali, G. Pearlson, and J. J. Pekar, "A method for making group inferences from functional MRI data using independent component analysis," *Hum. Brain Mapping*, vol. 16, no. 2, p. 131, 2002.
- [83] V. D. Calhoun and T. Adali, "Multisubject independent component analysis of fMRI: A decade of intrinsic networks, default mode, and neurodiagnostic discovery," *IEEE Rev. Biomed. Eng.*, vol. 5, pp. 60–73, 2012.
- [84] E. B. Erhardt, S. Rachakonda, E. J. Bedrick, E. A. Allen, T. Adali, and V. D. Calhoun, "Comparison of multi-subject ICA methods for analysis of fMRI data," *Hum. Brain Mapping*, vol. 32, no. 12, pp. 2075–2095, 2011.
- [85] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, 1995.
- [86] Y.-O. Li, T. Adali, and V. D. Calhoun, "Estimating the number of independent components for functional magnetic resonance imaging data," *Hum. Brain Mapping*, vol. 28, no. 11, pp. 1251–1266, 2007.
- [87] Q. Long, C. Jia, Z. Boukouvalas, B. Gabrielson, D. Emge, and T. Adali, "Consistent run selection for independent component analysis: Application to fMRI analysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 2581–2585.
- [88] E. A. Allen, E. B. Erhardt, E. Damaraju, W. Gruner, J. M. Segall, R. F. Silva, M. Havlicek, S. Rachakonda, J. Fries, R. Kalyanam, and A. M. Michael, "A baseline for the multivariate comparison of resting-state networks," *Frontiers Syst. Neurosci.*, vol. 5, p. 2, Feb. 2011.
- [89] M. D. Greicius, G. Srivastava, A. L. Reiss, and V. Menon, "Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 13, pp. 4637–4642, 2004.
- [90] W. Du, V. D. Calhoun, H. Li, S. Ma, T. Eichele, K. A. Kiehl, G. D. Pearlson, and T. Adali, "High classification accuracy for schizophrenia with rest and task fMRI data," *Frontiers Hum. Neurosci.*, vol. 6, p. 145, Jun. 2012.
- [91] S. Noble, D. Scheinost, and R. T. Constable, "Cluster failure or power failure? Evaluating sensitivity in cluster-level inference," *NeuroImage*, vol. 209, Apr. 2020, Art. no. 116468.
- [92] S. D. Forman, J. D. Cohen, M. Fitzgerald, W. F. Eddy, M. A. Mintun, and D. C. Noll, "Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold," *Magn. Reson. Med.*, vol. 33, no. 5, pp. 636–647, May 1995.

- [93] K. J. Friston, K. J. Worsley, R. S. J. Frackowiak, J. C. Mazziotta, and A. C. Evans, "Assessing the significance of focal activations using their spatial extent," *Hum. Brain Mapping*, vol. 1, no. 3, pp. 210–220, 1994.
- [94] S. M. Smith and T. E. Nichols, "Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference," *NeuroImage*, vol. 44, no. 1, pp. 83–98, 2009.
- [95] C.-W. Woo, A. Krishnan, and T. D. Wager, "Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations," *Neuroimage*, vol. 91, pp. 412–419, May 2014.
- [96] C. M. Bennett, G. L. Wolford, and M. B. Miller, "The principled control of false positives in neuroimaging," *Social Cognit. Affect. Neurosci.*, vol. 4, no. 4, pp. 417–422, Dec. 2009.
- [97] K. Shin and U. Kang, "Distributed methods for high-dimensional and large-scale tensor factorization," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2014, pp. 989–994.
- [98] L. Grasedyck, M. Kluge, and S. Krämer, *Alternating Directions Fitting (ADF) of Hierarchical Low Rank Tensors*. Marburg, Germany: Verlag nicht ermittelbar, 2013.
- [99] A. Uschmajew, "Local convergence of the alternating least squares algorithm for canonical tensor approximation," *SIAM J. Matrix Anal. Appl.*, vol. 33, no. 2, pp. 639–652, 2012.
- [100] T. Rohwedder and A. Uschmajew, "On local convergence of alternating schemes for optimization of convex problems in the tensor train format," *SIAM J. Numer. Anal.*, vol. 51, no. 2, pp. 1134–1162, 2013.
- [101] I. V. Oseledets and E. E. Tyrtyshnikov, "Breaking the curse of dimensionality, or how to use SVD in many dimensions," *SIAM J. Sci. Comput.*, vol. 31, no. 5, pp. 3744–3759, 2009.
- [102] V. de Silva and L.-H. Lim, "Tensor rank and the ill-posedness of the best low-rank approximation problem," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1084–1127, 2008.
- [103] J. M. Lee, *Introduction to Smooth Manifolds*. New York, NY, USA: Springer, 2001.
- [104] M. D. Spivak, "A comprehensive introduction to differential geometry," *Bull. Amer. Math. Soc.*, vol. 84, no. 1, pp. 27–32, 1978.
- [105] A. Cichocki, A.-H. Phan, Q. Zhao, N. Lee, I. Oseledets, M. Sugiyama, and D. P. Mandic, "Tensor networks for dimensionality reduction and large-scale optimization: Part 2 applications and future perspectives," *Found. Trends Mach. Learn.*, vol. 9, no. 6, pp. 431–673, 2017.



IRINA BELYAeva (Graduate Student Member, IEEE) received the M.S. degree in computer science from Johns Hopkins University, Baltimore, MD, USA, in 2013. She is currently pursuing the Ph.D. degree in computer science with the University of Maryland, Baltimore County, Baltimore. Her main research interests include matrix and tensor factorizations with applications in medical imaging, specifically working with fMRI and MEG imaging modalities.



SUCHITA BHINGHE received the B.S. degree in electronics and telecommunication from the University of Pune, India, in 2013, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, Baltimore County, Baltimore, MD, USA, in 2015 and 2020, respectively. Her research focuses on developing unsupervised machine learning models by integrating concepts from matrix and tensor factorization, information theory and optimization methods. Her research interests include matrix decomposition, machine learning, deep learning, medical imaging analysis, data fusion, and video processing.



QUNFANG LONG (Member, IEEE) received the B.S. and M.S. degrees from Wuhan University, Wuhan, China, in 2012 and 2015, respectively, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, Baltimore County, Baltimore, MD, USA, in 2018 and 2020, respectively. Her research interests include neuroimaging analysis, AI/ML for biomedical data analysis, matrix/tensor decomposition, blind source separation, and data fusion.



TÜLAY ADALI (Fellow, IEEE) received the Ph.D. degree in electrical engineering from North Carolina State University, Raleigh, NC, USA, in 1992. She joined the Faculty at the University of Maryland, Baltimore County (UMBC), Baltimore, MD, USA, in 1992. She is currently a Distinguished University Professor with the Department of Computer Science and Electrical Engineering, UMBC. Her current research interests include statistical signal processing, machine learning, and their applications with emphasis on applications in medical image analysis and fusion. She is a fellow of the AIMBE, a Fulbright Scholar, and an IEEE Signal Processing Society Distinguished Lecturer. She was a recipient of the 2021 Humboldt Research Award, the 2010 IEEE Signal Processing Society Best Paper Award, the 2013 University System of Maryland Regents' Award for Research, and an NSF CAREER Award. She has been active in conference and workshop organizations. She was the General or a Technical Co-Chair of the IEEE Machine Learning for Signal Processing (MLSP) and Neural Networks for Signal Processing Workshops 2001–2008, and helped organize a number of conferences, including the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). She has served or also serving on numerous editorial boards and technical committees of the IEEE Signal Processing Society. She was the Chair of the Technical Committee on MLSP, 2003–2005 and 2011–2013, the Technical Program Co-Chair for ICASSP 2017, and the Special Sessions Chair for ICASSP 2018. She is also serving as the Vice President for Technical Directions of the IEEE Signal Processing Society.

• • •