Inexact Proximal Conjugate Subgradient Algorithm for fMRI Data Completion

Irina Belyaeva
Department of CSEE, UMBC
Baltimore, MD 21250
irinbel1@umbc.edu

Qunfang Long
Department of CSEE, UMBC
Baltimore, MD 21250
qunfang1@umbc.edu

Tülay Adali
Department of CSEE, UMBC
Baltimore, MD 21250
adali@umbc.edu

Abstract—Tensor representations have proven useful for many problems, including data completion. A promising application for tensor completion is functional magnetic resonance imaging (fMRI) data that has an inherent four-dimensional (4D) structure and is prone to missing voxels and regions due to issues in acquisition. A key component of successful tensor completion is a rank estimation. While widely used as a convex relaxation of the tensor rank, tensor nuclear norm (TNN) imposes strong low-rank constraints on all tensor modes to be simultaneously low-rank and often leads to suboptimal solutions. We propose a novel tensor completion model in tensor train (TT) format with a proximal conjugate subgradient (PCS-TT) method for solving the nonconvex rank minimization problem by using properties of Moreau's decomposition. PCS-TT allows the use of a wide range of robust estimators and can be used for data completion and sparse signal recovery problems. We present experimental results for data completion in fMRI, where PCS-TT demonstrates significant improvements compared with competing methods. In addition, we present results that demonstrate the advantages of considering the 4D structure of the fMRI data. as opposed to using three- and two-dimensional representations that have dominated the work on fMRI analysis.

Index Terms—Tensor completion, Tensor train decomposition, fMRI missing data completion.

I. INTRODUCTION

Tensors allow effective description of multilinear relationships and have proven useful in a large array of problems, including tensor completion. FMRI data, which have been a vital tool in research and clinical settings, suffer from equipment instabilities, subject, and task-related noise. As a result, typically a significant number of brain voxels, sometimes regions, are excluded from fMRI analysis, which impacts the success of subsequent analysis steps. Therefore, the development of effective data completion methods for high-dimensional fMRI data is a critical path to improve the reliability of such analyses. FMRI data are inherently four dimensional (4D); hence, tensor representations, in particular TT decomposition, offer particular advantages in terms of scalability and stability.

In data completion problems, the tensor rank captures the correlation information in different tensor modes in a TT representation, and therefore, to be able to learn the most significant interactions across the modes, the tensor rank

has to be correctly estimated. Recently, there have been many contributions that use the nonconvex surrogate low-rank penalty functions to better estimate the rank. These approaches can be categorized as convex relaxation of nuclear norm [1] or general nonconvex methods such as generalized singular value thresholding (GSVT) [2] and proximal gradient methods [3]. It has been shown in [4] that convex relaxation applies the same penalty to all singular values and leads to a biased solution. To alleviate the bias introduced by the convex relaxation of nuclear norms, many nonconvex regularizers have been proposed [5]. The surrogate rank minimization functions assign weighted smaller coefficients to penalize the larger singular values. This results in better empirical performance and statistical guarantees on recovery bounds [4]. To solve the fMRI data completion problem, we propose an inexact proximal conjugate subgradient algorithm based on the generalized Moreau-Yosida gradient to regularize a nonconvex surrogate rank function for tensor completion. Thus, by using a nonconvex objective, we can use surrogate penalties, which implicitly promote sparsity and ensure a low-rank property even when the rank of the original problem is overestimated. We propose a new inexact proximal conjugate subgradient (PCS-TT) method for tensor completion in the TT format. By its derivation, PCS-TT allows the use of robust low-rank estimators and can be used for data completion as well as sparse signal recovery problems. Experiments on real fMRI data demonstrate that PCS-TT provides significant improvements in terms of tensor completion metrics such as tensor completion score (TCS) [6] over competing methods, primarily due to the use of TT decomposition in conjunction with adaptive conjugate subgradients in computation. In particular, we demonstrate that taking the 4D structure of fMRI data promises significant gains by presenting comparisons with three-dimensional (3D) and two-dimensional (2D) representations of the same data, which have dominated the field when working with fMRI data.

II. PRELIMINARIES

We denote scalar, vectors, matrices and tensors by x, $\mathbf{x}$, $\mathbf{X}$, and $\mathcal{X}$. The Frobenius norm is defined by $\|\cdot\|_F$. The n-mode product is represented by $\times_n$. For a matrix $\mathbf{X}$, $\|\mathbf{X}\|_s := \sum_j \sigma_j(\mathbf{X})$ defines the Schatten-1 norm, where σ_j is a j-th singular value of $\mathbf{X}$.
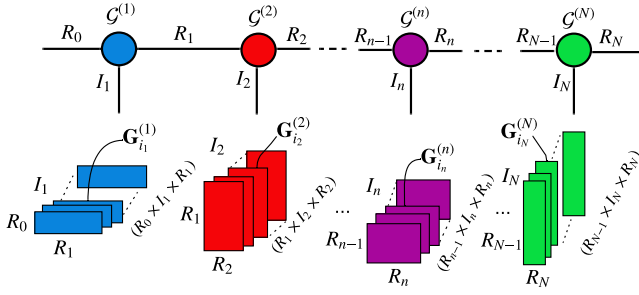
Fig. 1. Conceptual representation of the TT decomposition for an Nth multidimensional tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. TT decomposition of an Nth order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ with TT rank $\mathbf{r} = [r_0, r_1, r_2, \cdots, r_n, \cdots, r_{N-1}, r_N]$, where $r_0 = r_N = 1$. Analytically, the TT tensor can be described in the form of the slicewise representation as $x_{i_1, i_2, \cdots, i_N} = \mathbf{G}_{i_1}^{(1)} \mathbf{G}_{i_2}^{(2)} \cdots \mathbf{G}_{i_N}^{(N)}$, where $\mathbf{G}_{i_n}^{(n)} = \mathcal{G}^{(n)}(:, i_n, :) \in \mathbb{R}^{r_{n-1} \times r_n}$ are the slice matrices of TT-cores $\mathcal{G}^{(n)} \in \mathbb{R}^{r_{n-1} \times I_n \times r_n}$

A. Tensor Train Decomposition

A tensor train decomposition [7], shown in Fig. 1, represents an N-way tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_N}$ defined as

$$\mathcal{X} = \mathcal{G}^0 \times^1 \mathcal{G}^1 \times^1 \mathcal{G}^2 \times^1 \cdots \times^1 \mathcal{G}^N, \quad (1)$$

where $\mathcal{G}^n \in \mathbb{R}^{r_{n-1} \times I_n \times r_n}$ is the n-th TT core, and $\mathcal{G}^0, \mathcal{G}^1$ are (2D) matrices. The rank of TT decomposition (TT rank) [7] is an $N + 1$ tuple of ranks: $\operatorname{rank}_{TT}(\mathcal{X}) = \mathbf{r} = [r_0, r_1, r_2, \cdots, r_{N-1}, r_N]$, which controls the complexity of TT decomposition. The $\mathbf{r}_n$ component of TT rank is the rank of matrix $\mathbf{X}_{<n>}$ obtained by splitting operator [7] of tensor $\mathcal{X}$ as follows:

$$\Phi(\mathcal{X})_n = \mathbf{X}_{<n>} = \left(\mathcal{X}, \prod_{i=1}^{n} I_i, \prod_{i=n+1}^{N} I_i \right). \quad (2)$$

It was shown that the matrix rank of $\Phi(\mathcal{X})_n$ is the upper bound of the n-th TT rank of $\mathcal{X}$ and that $r_0 = r_n = 1$ [7].

B. Low-Rank Tensor Learning

The algorithm using robust low-rank tensor learning can be stated as the following optimization problem:

$$\min_{\mathcal{X}} F(\mathcal{X}) := f(\mathcal{X}) \quad (3)$$
$$s.t \quad \|\mathrm{P}_\Omega \mathcal{X} - \mathrm{P}_\Omega \mathcal{T}\|_F^2 \leq \epsilon$$

where ϵ is the noise tolerance. The constrained formulation in (3) can be converted into an unconstrained problem as

$$\min_{\mathcal{X}} F(\mathcal{X}) := \frac{1}{2} \|\mathrm{P}_\Omega \mathcal{X} - \mathrm{P}_\Omega \mathcal{T}_F\|^2 + \bar{F}_\lambda^a(\mathcal{X}), \quad (4)$$

where $\lambda > 0$ is the regularization parameter and $\bar{F}_\lambda^a$ is a nonconvex low-rank regularizer. P_Ω denotes projection onto sampling set Ω, and $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$.

$$\bar{F}_\lambda^a(\mathcal{X}) = \sum_{i=1}^{N-1} F_\lambda(\mathcal{X}), \quad (5)$$

, where F_λ is the generalized surrogate low-rank function satisfying Assumption 2 given below. Throughout the paper, we assume the following:

Assumption 1: $f(\mathcal{X})$ is convex differentiable, with L-Lipschitz gradient ∇f for some $L > 0$.

Assumption 2: $F_\lambda(\mathcal{X})$ is a concave, nondecreasing and a Lipschitz function on $\mathcal{X} \in [0, \infty)$ with $F_\lambda(0) = 0$. For a comprehensive list of nonconvex functions, we refer the reader to [5]. In Section III, we derive the update rules to solve the problem in (4) using the proximal algorithm framework [8].

III. GENERALIZED NONCONVEX LOW-RANK TENSOR COMPLETION

To pose TT decomposition as a low-rank learning problem (4), we need the surrogate relaxation of the TT rank. Hence, we introduce the overlapped Schatten TT norm [9]:

$$\|\mathcal{X}\|_{s_{TT}} := \frac{1}{N-1} \sum_{n=1}^{N-1} \|\Phi(\mathcal{X})_n\|_{s_{TT}}. \quad (6)$$

Since $\Phi(\mathcal{X})_n$ bounds the n-th TT rank of $\mathcal{X}$, the TT norm in (6) replaces the sum of the TT rank [4]. When we appply a nonconvex function (5) to the Schatten TT norm in (6), the tensor completion problem (4) becomes

$$\min_{\mathcal{X}} F(\mathcal{X}) := f(\mathcal{X}) + \frac{1}{N-1} \sum_{n=1}^{N-1} F_\lambda(\|\Phi(\mathcal{X})_n\|_{s_{TT}}) \quad (7)$$
$$= f(\mathcal{X}) + \bar{F}_\lambda^a(\mathcal{X}).$$

When $F_\lambda(\cdot)$ is a general low-rank promoting function, the problem in (7) can be regarded as a low-tensor rank minimization problem. The problem in (7) can be solved via simple fixed-point iterations, which can be derived from the optimality conditions. We record the first-order optimality conditions for problem (7) for any $\tau > 0$ as follows:

$$0 \in \nabla f(\mathcal{X}_*) + \partial \bar{F}_\lambda^a(\mathcal{X}_*) \quad (8)$$
$$0 \in \tau \nabla f(\mathcal{X}_*) - \mathcal{X}_* + \mathcal{X}_* + \partial \bar{F}_\lambda^a(\mathcal{X}_*)$$
$$(\mathcal{I} + \tau \partial \bar{F}_\lambda^a)\mathcal{X}_* \in (\mathcal{I} - \tau \nabla f)\mathcal{X}_*$$
$$\mathcal{X}_* = (\mathcal{I} + \tau \partial \bar{F}_\lambda^a)^{-1}(\mathcal{I} - \tau \nabla f)\mathcal{X}_*.$$

By definition $(\mathcal{I} + \tau \partial \bar{F}_\lambda^a)^{-1}\mathcal{X} \equiv \operatorname{prox}_{\frac{\lambda}{\tau} \bar{F}_\lambda^a}(\mathcal{X})$ [8], the fixed-point iterative scheme is given as

$$\mathcal{X}_k = (\mathcal{I} + \tau_k \partial \bar{F})^{-1}(\mathcal{I} - \tau_k \nabla f)\mathcal{X}_{k-1}, \quad (9)$$

, where the proximity operator $\operatorname{prox}_{\frac{\lambda}{\tau} \bar{F}_\lambda^a}(\mathcal{X})$ of $\bar{F}_\lambda^a$ is given in the context of Moreau's decomposition [8]:

$$\operatorname{prox}_{\frac{\lambda}{\tau} \bar{F}_\lambda^a}(\mathcal{X}) = \operatorname{argmin}_{\mathcal{Z}} \frac{1}{2\tau} \|\mathcal{Z} - \mathcal{X}\|_F^2 + \bar{F}_\lambda^a(\mathcal{Z}). \quad (10)$$

Using the definition of the proximity operator in (10), the proximal step can be written as follows:

$$\mathcal{X}_k = \operatorname{prox}_{\frac{\lambda}{\tau_k} \bar{F}_\lambda^a} \left(\mathcal{X}_{k-1} - \tau_k \nabla f(\mathcal{X}_{k-1}) \right) \quad (11)$$

$$= \operatorname{argmin}_{\mathcal{X}} \left\{ \frac{1}{2\tau_k} \left[\|\mathcal{Z} - (\mathcal{X}_{k-1} - \tau_k \nabla f(\mathcal{X}_{k-1}))\|_F^2 \right] + \bar{F}_\lambda^a(\mathcal{X}) \right\}$$

The proximal mapping in (11) can be obtained via singular value thresholding [2], which can be computed using subgradients of the components functions $F_\lambda(\mathcal{X})$ in an elementwise manner.

1026

IV. PROXIMAL CONJUGATE SUBGRADIENT METHOD FOR TENSOR COMPLETION

In this section, we present a conjugate subgradient algorithm to solve the problem in (7). The proximal step (11) can be regarded as the Moreau envelope [8] of the function $\bar{F}_\lambda^a(\mathcal{X})$, which is a smoothed version of $f(\mathcal{X})$ [8]. Therefore, we can use inexact gradient-based methods designed for nonconvex differential problems, such as the variable metric inexact proximal methods. We develop the inexact conjugate subgradient method, which uses the variable metric $\mathcal{H}_k$ to scale the gradient step, and the quasi-Newton update to smooth the proximal step. The scaling and smoothing operations are in place to enforce sufficient descent for the search direction. The conjugate search direction $\mathcal{D}_k$ is computed using the Polak–Ribière–Polyak (PRP) method for nonsmooth problems in [10], which utilizes the modified secant equation with the gradient and function values so that the sufficient descent property is satisfied [10].

A. PCS-TT Algorithm

With the prerequisite concepts introduced in Section III, we define the PCS-TT update rule

$$\mathcal{P}_k^+ = \mathcal{X}_{k-1} - \tau_k \mathcal{H}_k \nabla f(\mathcal{X}_{k-1}) \tag{12a}$$

$$\mathcal{X}_k^{(+,\epsilon_k)} = \text{prox}_{\lambda\tau_k}^{\mathcal{H}_k}(\mathcal{P}_k^+) \tag{12b}$$

$$\mathcal{G}(\mathcal{X}_{k+1}) = (\mathcal{X}_k - \mathcal{X}_k^{(+,\epsilon_k)})/\tau_k \tag{12c}$$

$$\mathcal{X}_{k+1} = \mathcal{X}_k + \alpha_k \mathcal{D}_k, \tag{12d}$$

, where α_k is a step length, and $\mathcal{D}_k$ is a search direction at the k-th iteration, and $\mathcal{H}_k = \tau_k \mathcal{I}$, which serves as an approximation of the inverse Hessian, the step length τ_k obtained by the Barzilai-Borwein (BB) [11] rule $\tau_k = (\mathcal{S}_{k-1}\mathcal{S}_{k-1})/(\mathcal{S}_{k-1}^T \mathcal{Z}_{k-1})$, $\mathcal{S}_{k-1} = \mathcal{X}_k - \mathcal{X}_{k-1}$, $\mathcal{Z}_{k-1} = \nabla f(\mathcal{X}_k) - \nabla f(\mathcal{X}_{k-1})$.

The search direction in (12d) is computed as

$$\mathcal{D}_{k+1} = \begin{cases} -\mathcal{G}_{k+1} & \text{if } k = 0 \\ -\mathcal{G}_{k+1} + \dfrac{\mathcal{G}_{k+1}^T \mathcal{Y}^* \mathcal{D}_k - \mathcal{D}_k^T \mathcal{G}_{k+1} \mathcal{Y}^*}{\lambda_k}, & \text{if } k \geq 1, \end{cases} \tag{13}$$

where $\lambda_k = \max\{2\mu\|\mathcal{D}_k\|_F \|\mathcal{Y}_k^*\|_F, \|\mathcal{G}_{k+1}\|_F^2\}$, $\mathcal{Y}^* = \mathcal{Y}_k + \gamma_k \mathcal{S}_k$, $\mathcal{Y} = \mathcal{G}_{k+1} - \mathcal{G}_k$, $\mathcal{S}_k = \mathcal{X}_{k+1} - \mathcal{X}_k$, $\mu > 0$, and

$$\gamma_k = \frac{\mathcal{G}_{k+1} + \mathcal{G}_k^T \mathcal{S}_k + 2(F(\mathcal{X}_k) - F(\mathcal{X}_{k+1}))}{\|\mathcal{S}_k\|_F^2}.$$

We find step length $\alpha_k > 0$ satisfying nonmonotone Armijo rules proposed in [12]

$$F(\mathcal{X}_k + \alpha_k \mathcal{D}_k) - c_k \leq \sigma \alpha_k \|\mathcal{D}_k\|_F^2 \tag{14a}$$

$$c_{k+1} = \frac{\eta q_k + F(\mathcal{X}_k)}{q_{k+1}}, q_{k+1} = \eta q_k + 1, \tag{14b}$$

where $\sigma > 0, \mu > 0, \rho > 0, \eta > 0$, $\alpha_k = \rho 2^{-i_k}, i_k = \{0, 1, 2, \cdots\}$, and the parameter η defines the level of non-momotonicity for the Armijo line search. Next, we present our PCS-TT algorithm for solving (7), which is stated in

---

Algorithm 1 PCS-TT: Algorithm for solving (7)

Input: The observed tensor $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, sampling set Ω, prescribed tolerance $\epsilon > 0$, $\sigma > 0, \eta > 0, \rho > 0$.
Output: The completed tensor $\mathcal{X}$ as approximation of $\mathcal{T}$.
    Initialize: $k = 0$, $\mathcal{X}_0 \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is a random tensor, $c_0 = F(\mathcal{X}_0), q_0 = 1, \mathcal{D}_0 = -\mathcal{G}_0(\mathcal{X}_0)$.
1: **while** $\|\mathcal{G}_k\|_F \geq \epsilon$ **do**
2:     Compute $\tau_k = (\mathcal{S}_{k-1}\mathcal{S}_{k-1})/(\mathcal{S}_{k-1}^T \mathcal{Y}_{k-1})$.
3:     Compute $\mathcal{X}_k^{(+,\epsilon_k)}$ by (12b).
4:     Compute gradient $\mathcal{G}_{k+1}$ by (12c).
5:     Compute the step length α_k by (14)
6:     $i_k = 0$
7:     Run the backtracking loop until (14) is satisfied
8:     **while** $F(\mathcal{X}_k + \alpha_k \mathcal{D}_k) - c_k \geq \sigma \alpha_k \|\mathcal{D}_k\|_F^2$ **do**
9:         $\alpha_k = \rho 2^{-i_k}$
10:        $i_k = i_k + 1$
11:        Update c_k and q_k by (14b)
12:     **end while**
13:     $\mathcal{X}_{k+1} = \mathcal{X}_k + \alpha_k \mathcal{D}_k$
14:     Compute search direction $\mathcal{D}_{k+1}$ by (13).
15: **end while**

Algorithm 1.

At each iteration, we check if the condition (14) is satisfied and run the backtracking loop until we find a suitable step length. The stopping criterion is defined as the norm of the composite gradient $\|\mathcal{G}\|_F$ being below a prescribed tolerance ϵ or the difference of the iterates $\frac{|F(\mathcal{X}_{k+1}) - F(\mathcal{X}_k)|}{|F(\mathcal{X}_k)|}$ being less than the convergence tolerance ϵ_c.

V. EXPERIMENTS

In this section, we present experiments on real fMRI data and show the effectiveness of the PCS-TT algorithm. We compare our method with low-rank tensor completion solvers based on the sum of nuclear norms (ADMM-SNN) [1] and the TMac-TT algorithm [9]. We select ADMM-SNN and TMac-TT for comparison because they belong to a family of tensor completion solvers based on the convex relaxation of a nuclear norm similar to PCS-TT. We conduct experiments in different tensor dimensions and study the effectiveness of PCS-TT using 2D, 3D, and 4D representations of fMRI data. There are many alternatives that exist for the choice of low-rank function $\bar{F}_\lambda^a(\cdot)$ [5]. However, we test our algorithm by solving problem (7) on nonconvex surrogate low-rank functions such as Geman [13], logarithm [14], and convex formulation of the Schatten TT norm in (6). The selected regularizers satisfy Assumption 2, and the problem in (7) can be solved using Algorithm 1.

A. Methods

In our experiments, the input is a sparse tensor $P_\Omega \mathcal{X} * \mathcal{T}$, where $\mathcal{T}$ is the fully observed true tensor, $P_\Omega \mathcal{X}$ is the projected tensor onto the sampling set Ω, and $*$ is the Hadamard (elementwise) product. The completed tensor $\hat{\mathcal{X}}$ is computed as $\hat{\mathcal{X}} = (1 - P_\Omega \mathcal{X} * \mathcal{T}) * \mathcal{X} + P_\Omega \mathcal{X} * \mathcal{T}$, where $\mathcal{X}$ is the output tensor estimated by each experiment. The following
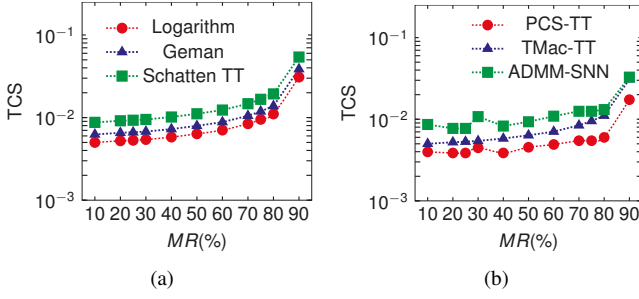
Fig. 2. Tensor completion performance. (a) TCS comparison by penalty type (b) TCS comparison between the state-of-the-art algorithms and PCS-TT for the logarithm penalty.
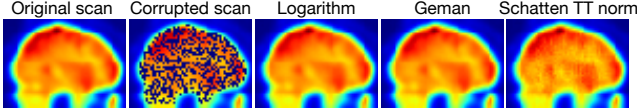

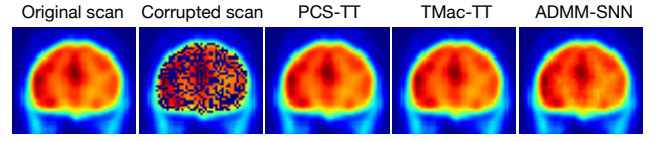Fig. 3. 4D fMRI tensor completion with MR = 50% at 10th timepoint using different low-rank penalty functions.

Fig. 4. Tensor completion with MR = 50% at the 10th timepoint using the proposed and state-of-the-art methods.

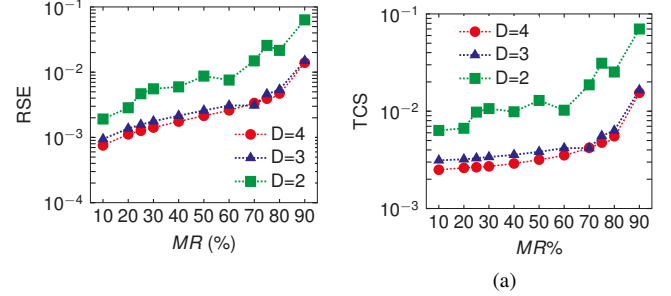Fig. 5. High-dimensional fMRI data completion by MR for tensor dimensions: $D = 2, 3, 4$.

parameters were used in the implementation of the proposed method: $\epsilon = 10^{-8}$, $\sigma = 10^{-4}, \rho = 0.5, \eta = 0.001, \mu = 0.8$. We evaluate the performance of the algorithm using the residual square error (RSE) between the completed tensor $\hat{\mathcal{X}}$ and the true tensor $\mathcal{T}$ that is defined as RSE $= (\|\hat{\mathcal{X}} - \mathcal{T}\|_F)/(\|\mathcal{T}\|_F)$, and the reconstruction quality in missing values is measured using the tensor completion score (TCS) TCS $= (\|(1 - P_\Omega \mathcal{X}) * (\hat{\mathcal{X}} - \mathcal{T})\|_F)/(\|(1 - P_\Omega \mathcal{X}) * \mathcal{T}\|_F)$ [6]. The convergence criterion of the competing algorithms is defined when the relative tolerance for RSE_Ω is below a tolerance, $\epsilon = 10^{-6}$ or exceeds the maximum number of iterations $\max_{\text{iter}} = 500$. The missing elements were sampled at random from a uniform distribution over [0,1]. We study the numerical performance with respect to different missing values rate MR defined as $MR = m/(\prod_{k=1}^{N} I_k)$, where m is the number of missing entries, and I_k is the number elements in each tensor mode. Each experiment is repeated ten times, and we present averages across ten runs.

B. Image Acquisition and Processing

We assessed the numerical performance of the PCS-TT method using real fMRI data from the Center of Biomedical Research Excellence (COBRE) data exchange repository [15]. The resting state fMRI data consisted of 149 volumes of $T2^*$-weighted functional images each, acquired using a gradient-echo EPI sequence: TR = 2s, TE = 29 ms, flip angle = 75 degree. The data were spatially normalized to the standard Montreal Neurological Institute space, resampled to 3 mm×3 mm×3mm voxels, and smoothed using a Gaussian kernel with a full-width at half-maximum of 5 mm.

VI. RESULTS

A. Performance Comparison by Missing Value Rate

In this section, we investigate the performance of PCS-TT using 4D spatiotemporal fMRI scans with $10\% - 90\%$

missing values with respect to different low-rank penalty functions and two selected state-of-the-art tensor completion methods. We randomly sample voxels of the 4D fMRI tensor $\mathcal{X} \in \mathbb{R}^{53 \times 63 \times 46 \times 144}$, which represents a specific subject from the COBRE study described in Section V-B. The initial value of TT rank is set such that each $\mathbf{r}_n \le R_{\max}$, where $R_{\max}$ is the maximal theoretical tensor rank that corresponds to canonical polyadic decomposition (CPD) [7].

We compare the performance of convex and nonconvex regularizers as described in Section V with different MR values. The results are shown in Fig. 2 and Fig. 3. From Fig. 2, we can see that the lower values of MR result in better data completion in terms of TCS. As seen in Fig. 2a, the nonconvex rank penalty functions resulted in a lower TCS than the convex Schatten TT norm since the nonconvex low-rank penalty approximates the TT rank better than the convex relaxation of the Schatten TT norm. Fig. 2b and Fig. 4 show the comparison of PCS-TT with the logarithm penalty versus the state-of-the-art algorithms. PCS-TT outperforms ADMM-SNN and TMac-TT. The reason is that PCS-TT-logarithm has a tighter rank bound than methods based on the convex relaxation of the tensor norm.

B. Evaluation of PCS-TT in High-Dimensional Settings

FMRI data are typically analyzed as a 2D matrix where 3D brain volumes are flattened as rows, thus ignoring the 3D and 4D nature of the data. To assess the algorithmic performance using natural 4D representation of fMRI data, we evaluate tensor completion methods across 4D, 3D, and 2D representations. We study the numerical performance of PCS-TT using the logarithm low-rank penalty function in tensor different dimensions, where MR changes from 10-90%. We run fMRI tensor completion experiments for 2D, 3D, and 4D representation of the fMRI data tensor. We reshape the original 4D fMRI data tensor to 2D and 3D forms yielding
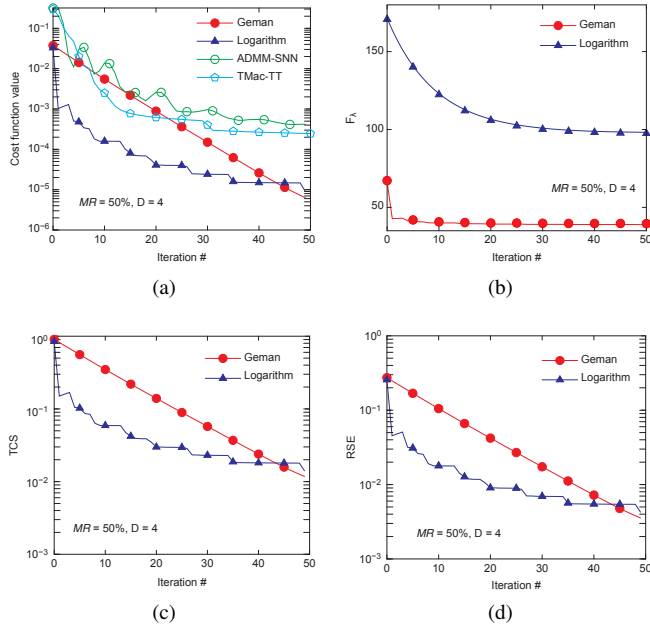
Fig. 6. Convergence curves for different surrogate low-rank functions using the PCS-TT algorithm with $MR = 50\%$ and tensor dimension $D = 4$. (a) Cost function value. (b) Surrogate rank penalty function by iteration. (c) TCS by iteration. (d) RSE by iteration.

$153,254 \times 144$ (2D) and $3,339 \times 46 \times 144$ (3D) and keep 4D in the original form $53 \times 63 \times 46 \times 144$. We access the impact of tensor dimensionality using numerical results in terms of RSE and TCS and present the results in Fig. 5. In Fig. 5, we can see that PCS-TT has the lowest RSE and TCS when the data are represented as the original spatiotemporal 4D tensor. The performance of PCS-TT slightly degrades when we cast the tensor in 3D form, and there is sharp degradation in terms of both RSE and TCS when we complete missing voxels using the 2D representation of the fMRI tensor.

C. Computational Cost and Convergence Rate

In this section, we present the results of the computational cost model for PCS-TT. Fig. 6 shows convergence curves of the objective function (7) w.r.t the number of iterations using different surrogate low-rank penalty functions. As we can see from Fig. 6a, the surrogate penalties have a similar impact on the convergence rate of PCS-TT w.r.t the number of iterations. Fig. 6a shows that the logarithm penalty decreases the objective function faster than the other low-rank penalties. In addition, Fig. 6a demonstrates the computational superiority of PCS-TT versus the state-of-the-art methods by providing a lower value of the objective function and faster convergence rate. In Fig. 6b, we examine how fast the surrogate low-rank penalty shrinks w.r.t the number of iterations by presenting convergence curves of logarithm and Geman functions we used in our experiments.

VII. CONCLUSION

We present a new conjugate subgradient method for tensor completion using a nonconvex generalized formulation of the nonsmooth objective function. Our method provides automatic low-rank tensor learning even when the original problem is overcomplete, which is especially important in high-dimensional tensor completion. PCS-TT demonstrates excellent numerical performance when fMRI data are in the form of a 4D tensor. The 4D representation of fMRI data takes into account the natural formation of the original data as a combination of 3D spatial brain volumes by time and therefore fully exploits spatiotemporal interactions. We conduct experiments using a number of robust estimators and compare PCS-TT with state-of-the-art methods. The experimental results demonstrate the advantages of nonconvex low-rank penalties over convex formulation, which can be explained by the fact that the robust estimators have intrinsic ability to be error prone against noise and outliers. We can also observe from the experiments the superiority of PCS-TT compared with the nuclear norm-based tensor completion methods. PCS-TT is more numerically efficient than TMac-TT and ADMM-SNN due to the use of robust adaptive low-rank penalty functions, which have tighter rank bounds.

REFERENCES

[1] C. Lu, J. Feng, S. Yan, and Z. Lin, "A unified alternating direction method of multipliers by majorization minimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 40, no. 3, pp. 527–541, 2018.

[2] C. Lu, C. Zhu, C. Xu, S. Yan, and Z. Lin, "Generalized singular value thresholding," in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[3] H. Li and Z. Lin, "Accelerated proximal gradient methods for nonconvex programming," in *Advances in neural information processing systems*, 2015, pp. 379–387.

[4] M. Imaizumi, T. Maehara, and K. Hayashi, "On tensor train rank minimization: Statistical efficiency and scalable algorithm," in *Advances in Neural Information Processing Systems*, 2017, pp. 3930–3939.

[5] C. Lu, J. Tang, S. Yan, and Z. Lin, "Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm," *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 829–839, 2015.

[6] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, "Scalable tensor factorizations for incomplete data," *Chemometrics and Intelligent Laboratory Systems*, vol. 106, no. 1, pp. 41–56, 2011.

[7] I. V. Oseledets, "Tensor-train decomposition," *SIAM Journal on Scientific Computing*, vol. 33, no. 5, pp. 2295–2317, 2011.

[8] N. Parikh, S. Boyd *et al.*, "Proximal algorithms," *Foundations and Trends® in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[9] J. A. Bengua *et al.*, "Efficient tensor completion for color image and video recovery: Low-rank tensor train," *IEEE Transactions on Image Processing*, vol. 26, no. 5, pp. 2466–2479, 2017.

[10] G. Yuan, Z. Wei, and G. Li, "A modified polak–ribière–polyak conjugate gradient algorithm for nonsmooth convex programs," *Journal of Computational and Applied mathematics*, vol. 255, pp. 86–96, 2014.

[11] E. G. Birgin, J. M. Martínez, and M. Raydan, "Inexact spectral projected gradient methods on convex sets," *IMA Journal of Numerical Analysis*, vol. 23, no. 4, pp. 539–559, 2003.

[12] H. Zhang and W. W. Hager, "A nonmonotone line search technique and its application to unconstrained optimization," *SIAM journal on Optimization*, vol. 14, no. 4, pp. 1043–1056, 2004.

[13] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE transactions on Image Processing*, vol. 4, no. 7, pp. 932–946, 1995.

[14] J. H. Friedman, "Fast sparse regression and classification," *International Journal of Forecasting*, vol. 28, no. 3, pp. 722–738, 2012.

[15] A. Scott *et al.*, "COINS: an innovative informatics and neuroimaging tool suite built for large heterogeneous datasets," *Frontiers in neuroinformatics*, vol. 5, p. 33, 2011.