Tensor Rank Estimation and Completion via CP-based Nuclear Norm

Qiquan Shi
Department of Computer Science,
Hong Kong Baptist University
Hong Kong, China
csqshi@comp.hkbu.edu.hk

Haiping Lu
Department of Computer Science,
University of Sheffield
Sheffield, U.K.
h.lu@sheffield.ac.uk

Yiu-ming Cheung
Department of Computer Science,
Hong Kong Baptist University
Hong Kong, China
ymc@comp.hkbu.edu.hk

ABSTRACT

Tensor completion (TC) is a challenging problem of recovering missing entries of a tensor from its partial observation. One main TC approach is based on CP/Tucker decomposition. However, this approach often requires the determination of a tensor rank a priori. This rank estimation problem is difficult in practice. Several Bayesian solutions have been proposed but they often under/overestimate the tensor rank while being quite slow. To address this problem of rank estimation with missing entries, we view the weight vector of the orthogonal CP decomposition of a tensor to be analogous to the vector of singular values of a matrix. Subsequently, we define a new CP-based tensor nuclear norm as the $L_1$-norm of this weight vector. We then propose Tensor Rank Estimation based on $L_1$-regularized orthogonal CP decomposition (TREL1) for both CP-rank and Tucker-rank. Specifically, we incorporate a regularization with CP-based tensor nuclear norm when minimizing the reconstruction error in TC to automatically determine the rank of an incomplete tensor. Experimental results on both synthetic and real data show that: 1) Given sufficient observed entries, TREL1 can estimate the true rank (both CP-rank and Tucker-rank) of incomplete tensors well; 2) The rank estimated by TREL1 can consistently improve recovery accuracy of decomposition-based TC methods; 3) TREL1 is not sensitive to its parameters in general and more efficient than existing rank estimation methods.

KEYWORDS

Tensor Rank Estimation, CP-based Tensor Nuclear Norm, CP Decomposition, Tensor Completion

ACM Reference format:

1 INTRODUCTION

Tensors are generalization of vectors (first-order tensors) and matrices (second-order tensors). They are ubiquitous (e.g., multichannel EEGs, images, videos, and social networks) and attract increasing interests [16]. Tensor completion, a task of recovering the missing entries based on partially observed entries, has drawn much attention recently in many applications of machine learning [24, 28, 32, 34] and data mining [29–31, 36].

One popular approach to solving tensor completion problems is tensor nuclear norm minimization, which is extended from matrix [5] to tensor case as a convex surrogate for rank minimization [18]. Although the nuclear norm approximation leads to good tensor completion performance under typical conditions [7, 11, 19], there is no theoretical guarantee that it is the tightest convex envelope of a tensor rank. Moreover, this approach is not efficient on large-scale tensors due to the heavy computation of singular value decomposition (SVD).

Another popular approach is based on tensor decompositions including CANDECOMP/PARAFAC (CP) [6, 12]) and Tucker decomposition [35], which is more promising for large-scale data. These two main decompositions lead to two common definitions for tensor rank: CP-rank and Tucker-rank respectively. This approach often requires a tensor rank as input. For example, a CP decomposition with weighted optimization method (CP-WOPT) [1] and an alternating minimization algorithm for tensors with a (fixed) low-rank orthogonal CP decomposition (TenALS) [14] can obtain good completion results for data with missing values under typical conditions. However, they need to manually choose a CP-rank as input, which is quite challenging because estimating the CP-rank is NP-hard [13], especially given incomplete information. On the other hand, by enforcing orthogonality into Tucker model, a generalized higher-order orthogonal iteration method (gHOI) [20] is developed to efficiently solve tensor completion problem, where the Tucker-rank for their model is obtained via a heuristic rank-increasing scheme. Furthermore, a simple Tucker decomposition-based approach (Tucker-WOPT) [10] fails to recover missing data accurately if the pre-specified rank is smaller than the true rank. Most recently, a Riemannian manifold optimization method (FRTC) [15] achieves good recovery performance on large-scale tensors, while still requiring a good rank value to be pre-specified. Moreover, its time cost increases exponentially with increasing input Tucker-rank.

Some studies attempt to estimate the CP/Tucker-rank of incomplete tensors automatically. Several Bayesian models have been proposed to automatically determine the CP-rank [3, 27, 42, 43]. For example, the CP rank of an incomplete tensor can be inferred by employing a multiplicative gamma process prior in [27], where the inference is performed by Gibbs sampler with slow convergence. Most recently, a Bayesian robust tensor factorization (BRTF) [43] employs a fully Bayesian generative model for automatic CP-rank
estimation. However, BRTF often under/over-estimates the true rank of incomplete tensors and has high computational cost.

To automatically estimate the Tucker-rank, an automatic relevance determination (ARD) algorithm is applied for sparse Tucker decomposition (ARD-Tucker) [23]. ARD is a hierarchical Bayesian approach widely used in many methods [26, 33, 37]. However, ARD-Tucker is not applicable to incomplete tensor data and its efficiency is quite low. Most recently, a robust Tucker-rank estimation method using Bayesian information criteria is proposed [39], while also only applicable to complete tensors.

In this paper, we view the weight vector of the orthogonal CP decomposition of a tensor as analogous to the vector of singular values of the SVD of a matrix. We then define a simple CP-based tensor nuclear norm as the $L_1$ norm of this weight vector. Based on this new tensor norm, we propose Tensor Rank Estimation based on $L_1$-regularized orthogonal CP decomposition, denoted as TREL1. TREL1 can automatically determine both CP-rank and Tucker-rank accurately given observed entries by removing the zero entries of the weight vector after optimization. We solve the optimization problem by block coordinate descent, where we optimize a block of variables while fixing the other blocks and update one variable while fixing the other variables in each block. In a nutshell, our contributions are fourfold:

- We propose TREL1 to automatically estimate the CP-rank of an incomplete tensor via a newly defined CP-based tensor nuclear norm.
- We automatically estimate the Tucker-rank in each mode by degenerating TREL1 to matrix case and applying it on the unfolded matrices of an incomplete tensor.
- We develop an efficient block coordinate descent algorithm for model optimization.
- We carry out extensive experiments to show that TREL1 is not sensitive to its parameters in general and more efficient than existing tensor rank estimation methods, and using TREL1 for rank estimation can improve the recovery accuracy of the state-of-the-art decomposition-based tensor completion methods.

This paper is organized as follows. We review preliminaries and backgrounds in Section 2. In Section 3, we define a CP-based tensor nuclear norm and propose two tensor rank estimation methods for both CP-rank and Tucker-rank estimation. We report empirical results in Section 4, and conclude this paper in Section 5.

2 PRELIMINARIES AND BACKGROUNDS

We first review the preliminaries and backgrounds [16, 22].

2.1 Notations and Operations

The number of dimensions of a tensor is the order and each dimension is a mode of it. A vector is denoted by a bold lower-case letter $x \in \mathbb{R}^d$ and a matrix is denoted by a bold capital letter $X \in \mathbb{R}^I \times \mathbb{R}^J$. A higher-order ($N \geq 3$) tensor is denoted by a calligraphic letter $X \in \mathbb{R}^I \times \cdots \times \mathbb{R}^N$. The $i$th entry of a vector $a \in \mathbb{R}^d$ is denoted by $a_i$, and the $(i,j)$th entry of a matrix $X \in \mathbb{R}^I \times \mathbb{R}^J$ is denoted by $X_{ij}$. The $(i_1, \cdots, i_N)$th entry of an $N$-th-order tensor $X$ is denoted by $X(i_1, \cdots, i_N)$, where $i_n \in \{1, \cdots, I_n\}$ and $n \in \{1, \cdots, N\}$. The Frobenius norm of a tensor $X$ is defined by $\|X\|_F = \langle X, X \rangle^{1/2}$. $\Omega$ is a binary index set: $\Omega(i_1, \cdots, i_N) = 1$ if $X(i_1, \cdots, i_N)$ is observed, and $\Omega(i_1, \cdots, i_N) = 0$ otherwise. $\mathcal{P}_\Omega$ is the associated sampling operator which acquires only the entries indexed by $\Omega$:

$$\mathcal{P}_\Omega(X)(i_1, \cdots, i_N) = \begin{cases} X(i_1, \cdots, i_N), & \text{if } (i_1, \cdots, i_N) \in \Omega \\ 0, & \text{if } (i_1, \cdots, i_N) \notin \Omega \end{cases},$$

where $\Omega^c$ is the complement of $\Omega$. We have $\mathcal{P}_\Omega(X) + \mathcal{P}_{\Omega^c}(X) = X$.

**Definition 1. Mode-$n$ Product.** A mode-$n$ product between a tensor $X \in \mathbb{R}^I \times \cdots \times \mathbb{R}^N$ and a matrix/vector $U \in \mathbb{R}^{I_n} \times \mathbb{R}^{R_n}$ is denoted by $Y = X \times_n U^T \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times I_{n+1} \times \cdots \times I_N}$, with entries given by $y_{i_1 \cdots i_{n-1} i_n i_{n+1} \cdots i_N} = \sum_{i_n} X_{i_1 \cdots i_{n-1} i_n i_{n+1} \cdots i_N} U_{i_n, J_n}$, and we have $Y_{(n)} = U^T X_{(n)}$ [22].

**Definition 2. Mode-$n$ Unfolding.** Unfolding, a.k.a., matricization or flattening, is the process of reordering the elements of a tensor into matrices along each mode [16]. A mode-$n$ unfolding matrix of a tensor $X \in \mathbb{R}^I \times \cdots \times \mathbb{R}^N$ is denoted as $X_{(n)} \in \mathbb{R}^{I_n \times \prod_{i \neq n} I_i}$.

2.2 CP and Tucker Decomposition

2.2.1 Tucker Decomposition and Tucker-rank. A tensor $X \in \mathbb{R}^I \times \cdots \times \mathbb{R}^N$ is represented as a core tensor with factor matrices in Tucker decomposition model [16]:

$$X = G \times_1 U^{(1)} \times_2 U^{(2)} \times_N U^{(N)},$$

where $U^{(n)} \in \mathbb{R}^{I_n \times R_n}$, $n = 1, 2, \ldots, N$, and $R_n < I_n$ are factor matrices with orthonormal columns and $G \in \mathbb{R}^{R_1 \times R_2 \times \cdots \times R_N}$ is the core tensor with smaller dimension. The Tucker-rank of an $N$th-order tensor $X$ is an $N$-dimensional vector, denoted as $r = (R_1, \ldots, R_N)$, whose $n$-th entry $R_n$ is the rank of the mode-$n$ unfolded matrix $X_{(n)}$ of $X$. $R_n$ is the mode-$n$ rank. Figure 1 illustrates this decomposition.

Figure 1: The Tucker decomposition of a third-order tensor.

![Figure 1](image1.png)

2.2.2 CP Decomposition

The CP decomposition of a third-order tensor $X$, where the core tensor $W$ is a super-diagonal tensor.

Figure 2: The CP decomposition of a third-order tensor $X$.
2.2.2 CP Decomposition and CP-rank. CP decomposition decomposes a tensor \( X \in \mathbb{R}^{l_1 \times \cdots \times l_N} \) as the weighted summation of a set of rank-one tensors:

\[
X = \sum_{r=1}^{R} w_r \, u_r^{(1)} \odot \cdots \odot u_r^{(N)} = \mathcal{W} \odot \mathbf{U}^{(1)} \odot \mathbf{U}^{(2)} \cdots \odot \mathbf{U}^{(N)},
\]

where each \( u_r^{(n)} \), \( n = 1, \cdots, N \) is a unit vector with the weight absorbed into the weight vector \( w = [w_1, \cdots, w_r, \cdots, w_R]^T \in \mathbb{R}^R \), and \( \odot \) denotes the outer product [16]. Figure 2 shows that CP decomposition is also could be reformulated as the Tucker decomposition where the core tensor \( \mathcal{W} \) is a super-diagonal tensor, i.e., \( \mathcal{W}(r, \cdots, r) = w_r \). \( R \) is CP-rank as the minimum number of rank-one components.

3 PROPOSED TENSOR RANK ESTIMATION METHODS

This section presents new Tensor Rank Estimation methods based on \( L_1 \)-regularized orthogonal CP decomposition, namely, TREL1. For simpler notations, we consider third order tensors \( X \in \mathbb{R}^{l_1 \times l_2 \times l_3} \), only while our methods generalize easily to higher-order tensors.

Orthogonal CP decomposition. In this paper, we consider the orthogonal CP decomposition, i.e., we enforce \( u_r^{(n)} \odot u_q^{(n)} = 0 \) for \( p \neq q \) and \( u_p^{(n)T} u_q^{(n)} = 1 \) otherwise. There are two motivations:

1. **CP decomposition** can be viewed as a generalization of SVD to tensors [8]. It seems natural to inherit the orthogonality of SVD in CP decomposition.
2. Although orthogonality is considered unnecessary in general or even impossible in certain cases in exact CP decomposition [4, 9, 40], some recent studies show that imposing orthogonality in the CP model can turn non-unique tensor decomposition into a unique one with guaranteed optimality [2, 14, 17, 40].

Tensor decomposition with missing data is more challenging than that with complete data in traditional problems. Furthermore, it is important to estimate a good rank from an incomplete tensor for accurate tensor completion. Therefore, we believe incorporating orthogonality into the CP model can help us determine the rank and further recover the tensor in the context of tensor completion. Our empirical studies to be presented later will show that the orthogonality constraint indeed gives better tensor rank estimation and completion results. Furthermore, we view the weight vector \( w \) of the orthogonal CP decomposition of a tensor \( X \) to be analogous to the vector of singular values of a matrix.

**Definition 3.** The CP-based Tensor Nuclear Norm of a tensor \( X \) is defined as the \( L_1 \) norm of the weight vector \( w \) of its orthogonal CP decomposition: \( \| X \|_{\text{CP}} = \| w \|_1 \).

In TREL1, we incorporate a regularization of CP-based tensor nuclear norm while minimizing the reconstruction error to obtain the estimated rank of an incomplete tensor and a low-rank recovery.

\[
\text{Then, our objective function is:}
\]

\[
\min_{X, w, u_r^{(n)}, R} \lambda \| w \|_1 + \frac{1}{2} \| X - \sum_{r=1}^{R} w_r u_r^{(1)} \odot u_r^{(2)} \odot u_r^{(3)} \|_F^2,
\]

\[
\text{s.t. } P_w(X) = P_w(T), u_r^{(n)T} u_q^{(n)} = 0, q = 1 \cdots r - 1, r = 1 \cdots R,
\]

where \( T \in \mathbb{R}^{l_1 \times l_2 \times l_3} \) is the given incomplete tensor with observed entries in \( \Omega \). \( w = [w_1, \cdots, w_r, \cdots, w_R] \) is the weight vector and \( R \) is the CP-rank of \( X \). \( \lambda \) is a regularization parameter.

3.1 Derivation of TREL1 by BCD

We employ the Block Coordinate Descent (BCD) (a.k.a., alternating minimization [14]) method for optimization. We divide the target variables into \( R + 1 \) blocks: \( \{ w_r, u_r^{(1)}, u_r^{(2)}, u_r^{(3)} \}, \cdots, \{ w_r, u_r^{(1)}, u_r^{(2)}, u_r^{(3)} \} \), \( X \). We optimize a block of variables while fixing the other blocks, and update one variable while fixing the other variables in each group. After updating the \( R + 1 \) blocks, we finally determine the tensor rank.

The Lagrangian function of (4) with respect to the \( r \)-th block \( \{ w_r, u_r^{(1)}, u_r^{(2)}, u_r^{(3)} \} \) is:

\[
\mathcal{L}_{w_r, u_r^{(n)}} = \lambda \| w_r \|_1 + \frac{1}{2} \| X_r - w_r u_r^{(1)} \odot u_r^{(2)} \odot u_r^{(3)} \|_F^2,
\]

\[
\text{s.t. } u_r^{(n)T} u_q^{(n)} = 0, q = 1 \cdots r - 1, r = 1 \cdots R,
\]

where \( X_r = X - \sum_{q=1}^{r-1} w_q u_q^{(1)} \odot u_q^{(2)} \odot u_q^{(3)} \) is the residual of the approximation. We use Lagrange multipliers to transform (5) to include all the constraints as:

\[
\mathcal{L}_{w_r, u_r^{(n)}} = \lambda \| w_r \|_1 + \frac{1}{2} \| X_r - w_r u_r^{(1)} \odot u_r^{(2)} \odot u_r^{(3)} \|_F^2
\]

\[
-\gamma (u_r^{(n)T} u_r^{(n)}) - 1 - \sum_{q=1}^{r-1} \mu_q u_q^{(1)} u_q^{(1)},
\]

where \( \gamma \) and \( \mu_q \) are the Lagrange multipliers.

3.1.1 Update \( u_r^{(n)} \). The function (6) with respect to \( u_r^{(1)} \) is:

\[
\mathcal{L}_{u_r^{(1)}} = \frac{1}{2} \| X_r - w_r u_r^{(1)} \odot u_r^{(2)} \odot u_r^{(3)} \|_F^2
\]

\[
-\gamma (u_r^{(1)T} u_r^{(1)}) - 1 - \sum_{q=1}^{r-1} \mu_q u_q^{(1)} u_q^{(1)}.
\]

Then we set the partial derivative of \( \mathcal{L}_{u_r^{(1)}} \) with respect to \( u_r^{(1)} \) to zero and eliminate the Lagrange multipliers, and get:

\[
u_r^{(1)} = (X_r \times_2 u_r^{(2)} \times_3 u_r^{(3)}) / w_r,
\]

\[
u_r^{(i)} = (X_r \times_2 u_r^{(2)} \times_3 u_r^{(3)}) / w_r,
\]

\[
\text{(8)}
\]
Algorithm 1 CP-rank Estimation Based on $L_1$-Regularized Orthogonal CP Decomposition (TREL1CP)

1. **Input:** $P_{\Omega}(T)$, $\Omega$, $\lambda$, initial rank $\hat{R}$, maximum iterations $K$, and stopping tolerance $tol$.
2. **Initialization:** Set $Z = \text{zeros}(l_1,l_2,l_3)$, $P_{\Omega}(X) = P_{\Omega}(T)$, $P_{\Omega^c}(X) = 0$; Initialize $[u^{(1)}_r, u^{(2)}_r, u^{(3)}_r, w_r]_{r=1}^\hat{R}$ of $X$ by RTPM [2].
3. for $k = 1, ..., K$ do
4. $X_k = X_k$;
5. for $r = 1, ..., \hat{R}$ do
6. if $w_r \neq 0$ then
7. Update $[u^{(1)}_r, u^{(2)}_r, u^{(3)}_r]$ by (8), (9), (10) respectively.
8. end if
9. end for
10. $X_r = X_r - w_r u^{(1)}_r \circ u^{(2)}_r \circ u^{(3)}_r$.
11. end for
12. **Update X:** Set $Z = X - X_r$ and update the missing entries by: $P_{\Omega^c}(X) = P_{\Omega^c}(Z)$.
13. If $\|w_{k+1} - w_k\|_2/\|w_{k+1}\|_2 < tol$, break; otherwise, continue.
14. end for
15. **CP-rank Estimation:** Only keep $w_r > 0$ in $w_r$ and then obtain the CP-rank $\hat{R} = \text{length}(w_r)$.
16. **Output:** $\hat{R}$.

and normalize $u^{(1)}_r = \frac{u^{(1)}_r}{\|u^{(1)}_r\|_2}$. Note that we only update the blocks with non-zero weights. Similarly, we update $u^{(2)}_r$ by

$$u^{(2)}_r = (X_r \times_1 u^{(1)}_r \times_3 u^{(3)}_r)/w_r - \left( \sum_{q=1}^{r-1} u^{(2)}_q (X_r \times_1 u^{(1)}_q \times_3 u^{(3)}_q) u^{(3)}_q \right)/w_r,$$

and normalize $u^{(2)}_r = \frac{u^{(2)}_r}{\|u^{(2)}_r\|_2}$, and update $u^{(3)}_r$ by

$$u^{(3)}_r = (X_r \times_1 u^{(1)}_r \times_2 u^{(2)}_r)/w_r - \left( \sum_{q=1}^{r-1} u^{(3)}_q (X_r \times_1 u^{(1)}_q \times_2 u^{(2)}_q) u^{(2)}_q \right)/w_r,$$

and normalize $u^{(3)}_r = \frac{u^{(3)}_r}{\|u^{(3)}_r\|_2}$.

3.1.2 **Update $w_r$.** The function (6) with respect to $w_r$ is:

$$L_{w_r} = \lambda |w_r|_1 + \frac{1}{2} \|X_r - w_r u^{(1)}_r \circ u^{(2)}_r \circ u^{(3)}_r\|_F^2.$$

Then we set the partial derivative $\frac{\partial L_{w_r}}{\partial w_r}$ to zero and obtain,

$$w_r = X_r \times_1 u^{(1)}_r \times_2 u^{(2)}_r \times_3 u^{(3)}_r - \frac{\lambda |w_r|_1}{\partial w_r}.$$

Based on the soft thresholding algorithm [25] for $L_1$ regularization, we update $w_r$ in (12) by:

$$w_r = \text{shrink}_\lambda \left( X_r \times_1 u^{(1)}_r \circ u^{(2)}_r \circ u^{(3)}_r \right),$$

where shrink is the soft thresholding operator [25], and we denote

$$S = (X_r \times_1 u^{(1)}_r \circ u^{(2)}_r \circ u^{(3)}_r).$$

3.2 TREL1 for CP-rank Estimation

Applying TREL1 directly for CP-rank estimation, we obtain a new CP-rank estimation method, namely, TREL1CP, summarized in Algorithm 1. Here we specify an initial medium rank value $\hat{R}$ for efficiency though we could also automatically set it to some high rank value, e.g., $\min(1,l_2,l_3)$. Besides, to obtain a good initialization, we initialize the CP decomposition of an incomplete tensor using Robust Tensor Power Method (RTPM) [2] following [14]. RTPM makes TREL1CP less sensitive to parameter $\lambda$ than using random initialization.

3.2.1 **Estimate CP-rank.** In Algorithm 1, after iteratively updating all the $R + 1$ blocks till convergence or reaching the maximum iterations, we finally determine the CP-rank: checking the weight vector $w$, we only keep the weights greater than zero. The number of the remaining weights in $w$ is the estimated CP-rank.

3.3 TREL1 for Tucker-rank Estimation

Since the Tucker-rank $r$ consists of the mode ranks of unfolded matrices of $X$ along each mode, we can compute the rank of each unfolded matrix $X_{(i)}$, $i = 1, 2, 3$, by degenerating TREL1 to matrix case. For the mode-$i$ unfolded matrix $X_{(i)}$ of a tensor $X \in \mathbb{R}^{l_1 \times l_2 \times l_3}$,
we have:
\[
\min_{X(i), w(u^r_n), R_i} \lambda \|w\|_1 + \frac{1}{2} \|X(i) - \sum_{r=1}^{R_i} w_r (u^r_1 \otimes u^r_2)\|^F, \\
\text{s.t. } P_{\Omega}(X) = P_{\Omega}(T), \quad u^r_n = 1, n = 1 \cdots 2, \\
\quad u^r_n \otimes u^r_q = 0, q = 1 \cdots r - 1, r = 1 \cdots R_i,
\]
where \(R_i\) is the rank \((i)-th\) entry of Tucker-rank) of mode-\(i\) unfolded matrix \(X(i)\) of \(X\). Here, each unfolded matrix is approximated by an orthogonal CP decomposition, which is actually the SVD of the unfolded matrix as the orthogonal CP decomposition is a generalization of SVD from matrices to tensors.

3.3.1 Estimate Tucker-rank. We degenerate the TREL1 to matrix case to estimate the mode ranks \([R_i]_{i=1}^3\) of unfolded matrices along each mode, and finally determine the Tucker-rank \(r = [R_1, R_2, R_3]\). We denote this TREL1 for Tucker-rank estimation as TREL1_Tucker and summarize it in Algorithm 2. Here, we use random initialization for weights and factors of \(X(i)\) because RTPM is only for third-order tensors.

Remark: This mode-wise estimation in TREL1_Tucker shares the same spirit as the Tucker-based nuclear norm (i.e., sum of the nuclear-norms of all the matricizations) and many other existing Tucker-based works. However, the key difference is that our TREL1 objective is to estimate the true Tucker-rank while Tucker-based nuclear norm is used to minimize the Tucker-rank. As to be shown in the empirical studies (e.g., Figs. 7 and 8), a smaller rank is not necessarily better and a rank smaller than the true rank often deteriorates the recovery performance.

3.4 Complexity Analysis

We analyze the complexity of TREL1 following [21], which mainly includes the shrinkage operator and some multiplications. At each iteration, the time complexity of performing the shrinkage operator in (13) is \(O(|R| \prod_{j=1}^3 I_j)\). This is also the time complexity of computing \((u^r_n)^T \otimes (u^r_q)^T = 0, q = 1 \cdots r - 1, r = 1 \cdots R_i\). The overall time complexity is \(O(KR|\prod_{j=1}^3 I_j|)\).

4 EXPERIMENTAL RESULTS

We implemented TREL1 in MATLAB to evaluate the rank estimation and tensor completion/recovery performance. All experiments were performed on a PC (Intel Xeon(R) 4.0GHz, 64GB).

4.1 Experimental Setup

4.1.1 Compared Methods. We mainly consider decomposition-based methods with two steps: (i) rank estimation, and (ii) tensor completion with the rank estimated in (i). In addition, we tested three popular baseline methods (SiLRTC, FaLRTC and HaLRTC) in [18, 19]. FaLRTC performs the best among the three, but inferior to gHOI with TREL1 on the whole. Thus, their results are not included below for more compact presentation.

(i) Rank estimation. We study four existing methods:
   - MGP-CP [27]: a Bayesian method for low-rank CP decomposition of incomplete tensors, which infers the CP-rank using a multiplicative gamma process.
   - BRTF [43]: a Bayesian robust tensor factorization which employs a fully Bayesian generative model for automatic CP-rank estimation.
   - ARD-Tucker [23]: an automatic relevance determination algorithm for sparse Tucker decomposition using the gradient based sparse coding algorithm.
   - SCORE [39]: a robust Tucker-rank estimation method using Bayesian information criteria for complete tensors.

Among the four methods, BRTF and ARD-Tucker performed much better than MGP-CP and SCORE, respectively. Thus, we only report the comparison of TREL1 against BRTF and ARD-Tucker to save space.

(ii) Tensor completion. We study two representative CP decomposition-based methods and three representative Tucker decomposition-based methods:
   - CP-WOPT [1]: CP decomposition with missing data is formulated as a weighted least squares problem and solved by a gradient descent optimization approach.
   - TenALS [14]: decomposing of incomplete tensors is formulated as a low-rank orthogonal CP decomposition problem, solved by an alternating minimization algorithm.
   - gHOI [20]: a generalized higher-order orthogonal iteration tensor completion method, based on orthogonal Tucker decomposition.
   - Tucker-WOPT [10]: a nonlinear conjugate gradient method that solves Tucker decomposition with missing data in a similar way as CP-WOPT.
   - FRTC [15]: a Riemannian manifold preconditioning approach for tensor completion with rank constraint.

\[\lambda \in [50 : 10 : 250]\]

Figure 3: (a-b) Estimated CP-ranks of two tensors \((100 \times 100 \times 100)\) with \(R = 5\) and \(200 \times 200 \times 200\) with \(R = 50\) by TREL1CP with \(\lambda \in [50 : 10 : 250]\); (c-d) the corresponding time costs.
Figure 4: (a-c) Estimated Tucker-ranks in each mode of a $200 \times 200 \times 200$ tensor by TREL1 with $\lambda \in [50 : 10 : 250]$; (d) The corresponding time costs.

Figure 5: (a) Estimated CP-ranks of the $100 \times 100 \times 100$ ($R = 5$) tensor with 50% missing entries by TREL1 given different initial ranks $\hat{R} \in [10 : 10 : 200]$ (b) Estimated Tucker-ranks of the $200 \times 200 \times 200$ ($r = [15, 18, 20]$) with 50% missing entries by TREL1 given different initial rank $\hat{R} \in [50 : 10 : 250]$.

Figure 6: Convergence curves of TREL1CP in terms of weights error: $\|w_{(k+1)} - w_k\|_2/\|w_k\|_2$ on two tensors.

4.2 Parameter Sensitivity

4.2.1 Rank Estimation Sensitivity on $\lambda$. Figures 3 and 4 show the rank estimation results on various synthetic tensors by TREL1 with different $\lambda$ values. As seen from Figs. 3(a) and 3(b), the rank estimation performance of TREL1CP is stable and not sensitive to the values of $\lambda$ in most cases. Only for very high missing ratios (e.g., $MR = 90\%$), a large $\lambda$ (e.g., $\lambda = 110$) can make the $L_1$ regularization dominate the whole objective function (4) and result in zero rank. In addition, the time costs of TREL1CP are stable with most of $\lambda$ values (e.g., $\lambda \in [70, 200]$), as shown in Figs. 3(c) and 3(d).

Figure 4 shows that TREL1 is not sensitive to $\lambda$ either on tensors with no more than half of data missing (i.e., $MR \leq 50\%$). However, for larger MRs, the rank estimation performance will deteriorate, which is not shown in the figures for clarity. Nevertheless, this is not surprising by noting that TREL1 is formulated based on CP decomposition so it suits the CP model better than the Tucker model. Nevertheless, for small to medium MRs, TREL1 can mostly produce an accurate estimate of the Tucker-rank for a wide range of $\lambda$. In addition, it is interesting to show that the time cost with a larger $\lambda$ is lower, as shown in Fig. 4(d).

In summary, CP/Tucker rank estimation performance does not require careful tuning of $\lambda$. The CP-rank estimation is accurate even for some challenging high MRs. Furthermore, the selection of $\lambda$ is largely insensitive to data. For example, good $\lambda$ values for synthetic tensors are good values for real tensors as well (to be shown in the following). Thus, we can fix $\lambda = 100$ in both CP/Tucker-rank estimation for both synthetic and real tensors. Note that in Tucker-rank estimation, we only need to set a single $\lambda$ and there is no need to set separate $\lambda$ values for each mode. Therefore, setting $\lambda$ is much more user-friendly than manually setting the rank values.

4.2.2 Rank Estimation Sensitivity on Initial Rank $\hat{R}$. Figures 5(a) and 5(b) study the sensitivity of rank estimation on $\hat{R}$. We can see the rank estimation results by TREL1 with different $\lambda$ values are
Figure 7: RSE of recovering a tensor (true CP-rank $R = 25$) via CP decomposition-based methods given (a) manually fixed ranks and (b) rank estimated via TREL1 with different $\lambda$.

(a) Using manually fixed CP-ranks
(b) Using estimated CP-ranks via $\lambda$

Figure 8: RSE of recovering a tensor (true Tucker-rank $r = [8, 10, 12]$) via Tucker decomposition-based methods given (a) manually fixed ranks and (b) rank estimated via TREL1 with different $\lambda$.

(a) Using manually fixed Tucker-ranks
(b) Using estimated Tucker-ranks via $\lambda$

not sensitive to $\hat{R}$ for both CP and Tucker models. Thus, we set $\hat{R} = \text{round}(1/2 \times \text{mean}(I_1, I_2, I_3))$ for all tests.

4.3 Convergence Study

Since TREL1_Tucker can be viewed as performing TREL1_CP multiple times on unfolded matrices, we only study the convergence of TREL1 for CP-rank estimation in terms of weights error: $\|w_{(k+1)} - w_k\|_2/\|w_{(k+1)}\|_2$. Figure 6 shows that TREL1_CP converges within 100 iterations except for a large MR ($> 70\%$), which needs more iterations to converge.

4.4 Effects of Rank Value on Completion Performance

Here, we present studies that investigate the effects of rank estimation accuracy on tensor completion performance of decomposition-based methods. All five decomposition-based tensor completion methods (i.e., CP-based CP-WOPT and TenALS, and Tucker-based gHOI, Tucker-WOPT and FRTC) listed in Sec. 4.1.1 (ii) are studied. We compare tensor completion performance with two ways of rank determination: (i) setting the rank manually; (ii) setting $\lambda$ in TREL1 to estimate the rank. We show the results of recovering two synthetic tensors with MR = $\{30\%, 50\%, 70\%\}$ in Figs. 7 and 8.

- As seen from Figs. 7(a) and 8(a), the completion performance (in RSE) of all five methods is highly sensitive to the manually set rank value. Even a slight error in the rank value (particularly lower-than-true ranks) can lead to serious performance degradation. Only given the true tensor ranks, all the five methods can achieve their best completion results in all cases. For CP-WOPT, TenALS and gHOI, setting any rank value different from the true rank gives much worse performance than their best results. Tucker-WOPT and FRTC can achieve good results given a higher-than-true rank although still worse than their best results.
- In contrast, Figs. 7(b) and 8(b) show the corresponding results with TREL1 rank estimation by setting $\lambda$ to a limited number of values only. We can see a wide range of $\lambda$ values lead to the best performance of all methods. Such range is particular wide for CP-based methods and narrower for Tucker-based methods, which is not surprising since TREL1 is designed based on a CP model.

This study shows the advantage of TREL1 in rank estimation, compared to manually specifying the rank. TREL1 greatly simplifies parameter tuning where a simple setting of $\lambda$ from a wide range of feasible values works for a wide range of methods and data. This not only improves the completion performance but also reduces the time cost in parameter tuning.

4.5 Tensor Rank Estimation and Completion Performance

Here, we report the results for MR = $\{30\%, 50\%, 70\%\}$ in four tables. We highlight the correctly estimated rank in italic and bold fonts, smallest RSE results in bold fonts and the second smallest RSE in underline. Here, the corresponding CP-rank estimated by TREL1_CP and BRTF are denoted as $\text{TREL1-}^r$ and $\text{BRTF-}^r$ respectively, and the corresponding Tucker-rank estimated by TREL1_Tucker and ARD-Tucker are denoted as $\text{TREL1-}^r$ and $\text{ARD-}^r$ respectively. Furthermore, the estimated tensor ranks are fed into decomposition-based tensor completion methods to compare the recovery performance.

4.5.1 CP-rank Estimation and Tensor Completion

On synthetic tensors: As shown in the left half of Table 1, TREL1_CP correctly determines the true CP-ranks of the synthetic tensors in all cases, while BRTF over-estimates the ranks (it only succeeds in one case). More importantly, with TREL1-$_r$, both CP-WOPT and TenALS achieve their best recovery results, as seen from the left half of Table 2. Moreover, TenALS outperforms CP-WOPT both given the true ranks ($\text{TREL1-}^r$), which demonstrates the benefits of orthogonal CP decomposition for tensor completion.

On real tensors: We cannot directly evaluate the estimated CP-ranks since we do not know the ground-truth of CP-ranks for real tensors. Thus, we alternatively compare the tensor completion results affected by the estimated CP-ranks. As seen from the right half of Table 1, TREL1-$_r$ improves the completion performance of CP-WOPT and TenALS with around 25% than that of using BRTF-$_r$. Moreover, with TREL1-$_r$, TenALS still achieves better results than CP-WOPT on the whole.

4.5.2 Tucker-rank Estimation and Tensor Completion

On synthetic tensors: As reported in the left half of Table 3, TREL1_Tucker can correctly determine the true Tucker-ranks of the synthetic tensors in all cases, while ARD-Tucker fails (over-estimates or under-estimates) in these cases. Furthermore, with our estimated true ranks ($\text{TREL1-}^r$), the Tucker decomposition-based tensor completion methods (gHOI, Tucker-WOPT and FRTC)
Table 1: CP-rank estimation on synthetic and real tensors with MR = \{30\%, 50\%, 70\%\} missing entries. Est.\(R\) is the estimated CP-ranks and Time in seconds.

<table>
<thead>
<tr>
<th>Data</th>
<th>Synthetic</th>
<th>Synthetic</th>
<th>Synthetic</th>
<th>Real</th>
<th>Real</th>
<th>Real</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50 × 50 × 50</td>
<td>100 × 100 × 100</td>
<td>200 × 200 × 200</td>
<td>144 × 256 × 40</td>
<td>256 × 256 × 24</td>
<td>144 × 176 × 300</td>
</tr>
<tr>
<td>R = 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR</td>
<td>30%</td>
<td>50%</td>
<td>70%</td>
<td>30%</td>
<td>50%</td>
<td>70%</td>
</tr>
<tr>
<td>TREL1(CP) Est.(R)</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>25</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Time</td>
<td>1.59</td>
<td>1.61</td>
<td>2.41</td>
<td>29.17</td>
<td>33.04</td>
<td>43.74</td>
</tr>
<tr>
<td>BRTF</td>
<td>9</td>
<td>6</td>
<td>5</td>
<td>36</td>
<td>42</td>
<td>32</td>
</tr>
<tr>
<td>Est.(R)</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Time</td>
<td>165.57</td>
<td>100.35</td>
<td>88.47</td>
<td>176.88</td>
<td>191.31</td>
<td>193.54</td>
</tr>
</tbody>
</table>

Table 2: Tensor completion results by CP-based methods given estimated CP-ranks on synthetic and real tensors with MR = \{30\%, 50\%, 70\%\} missing entries. TREL1-\(R\) and BRTF-\(R\) refer to the corresponding CP-ranks estimated by TREL1\(CP\) and BRTF.

<table>
<thead>
<tr>
<th>Data</th>
<th>Synthetic</th>
<th>Synthetic</th>
<th>Synthetic</th>
<th>Real</th>
<th>Real</th>
<th>Real</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50 × 50 × 50</td>
<td>100 × 100 × 100</td>
<td>200 × 200 × 200</td>
<td>144 × 256 × 40</td>
<td>256 × 256 × 24</td>
<td>144 × 176 × 300</td>
</tr>
<tr>
<td>R = 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR</td>
<td>30%</td>
<td>50%</td>
<td>70%</td>
<td>30%</td>
<td>50%</td>
<td>70%</td>
</tr>
<tr>
<td>CP-WOPT</td>
<td>RSE</td>
<td>1.26E-07</td>
<td>1.43E-07</td>
<td>6.12E-07</td>
<td>1.94E-07</td>
<td>3.42E-07</td>
</tr>
<tr>
<td></td>
<td>with TREL1-(R) Time</td>
<td>6.16</td>
<td>3.12</td>
<td>1.65</td>
<td>1011.99</td>
<td>636.24</td>
</tr>
<tr>
<td>CP-WOPT</td>
<td>RSE</td>
<td>9.96E-05</td>
<td>3.82E-05</td>
<td>4.12E-07</td>
<td>1.94E-04</td>
<td>1.87E-04</td>
</tr>
<tr>
<td></td>
<td>with BRTF-(R) Time</td>
<td>23.62</td>
<td>9.69</td>
<td>1.66</td>
<td>6149.26</td>
<td>4631.69</td>
</tr>
<tr>
<td>TenALS</td>
<td>7.11E-09</td>
<td>1.65E-09</td>
<td>9.74E-09</td>
<td>1.36E-09</td>
<td>2.81E-09</td>
<td>8.26E-09</td>
</tr>
<tr>
<td></td>
<td>with TREL1-(R) Time</td>
<td>1.19</td>
<td>1.22</td>
<td>1.26</td>
<td>97.59</td>
<td>104.84</td>
</tr>
<tr>
<td>TenALS</td>
<td>1.61E-09</td>
<td>3.12E-09</td>
<td>9.74E-09</td>
<td>5.12E-07</td>
<td>1.46E-08</td>
<td>9.06E-04</td>
</tr>
<tr>
<td></td>
<td>with BRTF-(R) Time</td>
<td>24.02</td>
<td>14.78</td>
<td>1.23</td>
<td>2290.32</td>
<td>1876.08</td>
</tr>
</tbody>
</table>

Outperform the cases of using Tucker-ranks estimated by ARD-Tucker (ARD-\(r\)) by several orders, as shown in the left half of Table 4. Besides, FRTC fails to recover the tensors with more than 39 hours time costs in five cases as its computational cost increases exponentially given over-estimated Tucker-ranks (ARD-\(r\)).

On real tensors: Unlike synthetic data with true Tucker-rank because we constructed them via QR decomposition and can control the dimensions of its core tensor (Tucker-rank), the real tensors naturally do not have exact low Tucker-ranks. We can unfold a real tensor along each mode and then truncate its mode rank (\(R_1, R_2\) and \(R_3\)) to be exact low-rank. However, because the unfolded matrices are interdependent, we can only control the low-rank in one mode exactly. Therefore, we studied the cases of truncating the unfolded matrices of a tensor into exact low-rank in one of the three modes, and report the results for the mode with the highest dimension. In this way, we can directly evaluate the estimated Tucker-rank exactly in one mode at least. As shown in the right half of Table 3, our method can correctly estimate the mode-1 rank (\(R_1 = 22\)) of Ocean video, the mode-2 rank (\(R_2 = 7\)) of Flow Injection and the mode-3 rank (\(R_3 = 4\)) of Amino Acid in all cases, while ARD-Tucker fails to get the exact mode ranks for these real tensors. In addition, the results shown in the right half of Table 4 demonstrate that: with TREL1-\(r\), the three Tucker decomposition-based tensor completion methods improves recovery performance than that of given ARD-\(r\). Nevertheless, with ARD-\(r\), Tucker-WOPT achieves the second best recovery results in two cases because it assumes the true ranks can be over-estimated.

4.5.3 Time Cost of Rank Estimation and Tensor Completion.

Time cost of TREL1-\(R\) estimation: As seen from Table 1: TREL1\(CP\) is much faster than BRTF and only needs less than 1% and 18% of BRTF’s time cost on synthetic and real tensors on average respectively. Especially on the larger tensors with higher ranks, for example, TREL1\(CP\) is about 300 times faster than BRTF on average on the 200 × 200 × 200 tensor with \(R = 50\). Table 3 shows that TREL1\(Tucker\) is about 9 times faster than ARD-Tucker on average on the synthetic and real tensors. Thus, due to the expensive time costs of the compared methods, it is not feasible to report results of the lager tensors here.
Table 3: Tucker-rank estimation results on synthetic and real tensors with MR = \{30\%, 50\%, 70\%\} missing entries. Time in seconds. Est.$R_1$, Est.$R_2$ and Est.$R_3$ are the estimated Tucker-ranks in mode-1, mode-2 and mode-3, respectively.

<table>
<thead>
<tr>
<th>Data</th>
<th>Synthetic</th>
<th>Synthetic</th>
<th>Synthetic</th>
<th>Real</th>
<th>Real</th>
<th>Real</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50 × 50</td>
<td>80 × 100</td>
<td>200 × 200</td>
<td>5 × 61</td>
<td>12 × 100</td>
<td>160 × 112</td>
</tr>
<tr>
<td>r = [5, 5, 5]</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>5</td>
<td>11</td>
<td>22</td>
</tr>
<tr>
<td>r = [8, 10, 12]</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>51</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td>r = [15, 18, 20]</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>23</td>
<td>32</td>
<td>32</td>
</tr>
</tbody>
</table>

Table 4: Tensor completion results by Tucker-based methods given estimated ranks on synthetic and real tensors with MR = \{30\%, 50\%, 70\%\} missing entries. Time in seconds and "-" indicates that the results cost more than 50 hours. TREL1-r and ARD-r refer to the corresponding Tucker-ranks estimated by TREL1-Tucker and ARD-Tucker.

<table>
<thead>
<tr>
<th>Data</th>
<th>Synthetic</th>
<th>Synthetic</th>
<th>Synthetic</th>
<th>Real</th>
<th>Real</th>
<th>Real</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50 × 50</td>
<td>80 × 100</td>
<td>200 × 200</td>
<td>5 × 61</td>
<td>12 × 100</td>
<td>160 × 112</td>
</tr>
<tr>
<td>r = [5, 5, 5]</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>5</td>
<td>11</td>
<td>22</td>
</tr>
<tr>
<td>r = [8, 10, 12]</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>51</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td>r = [15, 18, 20]</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>23</td>
<td>32</td>
<td>32</td>
</tr>
</tbody>
</table>

Time cost of tensor completion using TREL1: As shown in Table 1: unlike on synthetic tensors, CP-WOPT and TenALS with TREL1-R cost more time than those with BRTF-R on real tensors because TREL1-Rs are larger than BRTF-Rs, though leading to better accuracy. This increased time cost is inherent for the tensor completion algorithms rather than TREL1. On Tucker-rank estimation, FRTC with TREL1-r is much faster than FRTC given ARD-r in most cases, as observed in Table 4.

5 CONCLUSION

In this paper, we defined a simple CP-based tensor nuclear norm and proposed two novel tensor rank (both CP-rank and Tucker-rank) estimation methods, TREL1-CP and TREL1-Tucker, based on orthogonal CP decomposition. In the proposed methods, we impose an $L_1$ regularization on the weight vector of the orthogonal CP decomposition, served as the CP-based tensor nuclear norm, while minimizing the reconstruction error. This leads to automatic rank determination for incomplete tensors. As demonstrated in
our experimental results, TREL1 can correctly determine the true tensor ranks (both CP-rank and Tucker-rank) of synthetic tensors, and also can estimate the rank of real tensors well given sufficient observed entries. More importantly, our estimated tensor ranks consistently improved the recovery performance of decomposition-based tensor completion methods. Besides, TREL1 is not sensitive to its parameters in general and much more efficient than existing tensor rank estimation methods.

ACKNOWLEDGMENT
This work is supported by the NSFC Grant: 61672444 and 61272366, HKBU Faculty Research Grant: FRG2/16-17/051, the SZSTI Grant: JCYJ20160531119406833, the HKBU KTO grant (MPCF-004-2017/18), and Hong Kong PhD Fellowship Scheme. We thank Prof. Piyush Rai, Dr. Bamdev Mishra, and Dr. Hiroyuki KASAI for their code sharing and helpful discussion.

REFERENCES