Received January 18, 2018, accepted February 15, 2018, date of publication March 2, 2018, date of current version March 16, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2811396

Tensor Completion Using Spectral (k, p)-Support Norm

DONGXU WEI[1,2,3], ANDONG WANG[4], BO WANG[5], AND XIAOQIN FENG[1]

[1]School of Physics and Electronic Electrical Engineering, Huaiyin Normal University, Huai'an 223300, China
[2]School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[3]Jiangsu Province Key Construction Laboratory of Modern Measurement Technology and Intelligent System, Huai'an 223300, China
[4]School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[5]Jiangsu Shuoshi Welding Technology Company, Ltd., Nanjing 211103, China

Corresponding author: Dongxu Wei (weidx2003@126.com)

This work was supported in part by Huai 'an Key Research And Development Plan (industry and informatization) under Grant HAG2015038, in part by the National Natural Science Foundation of China under Grant 61373063, Grant 61601235, Grant 61703209, and Grant 91420201, and in part by the Pre-Research Area Foundation under Grant 6140312010101.

ABSTRACT In this paper, the goal is to reconstruct a tensor, i.e., a multi-dimensional array, when only subsets of its entries are observed. For well-posedness, the tensor is assumed to have a low-Tucker-rank structure. To estimate the underlying tensor from its partial observations, we first propose an estimator based on a newly defined balanced spectral (k, p)-support norm. To efficiently compute the estimator, we come up with a scalable algorithm for the minimization of the spectral (k, p)-support norm. Instead of directly solving the primal problem which involves full SVD in each iteration, the proposed algorithm benefits from the Lagrangian dual through minimizing the dual norm of the (k, p)-support norm which only computes the first k leading singular values and singular vectors in each iteration. To explore the statistical performance of the proposed estimator, upper bounds on the sample complexity and estimation error are then established. Simulation studies confirm that the error bounds can predict the scalable behavior of the estimation error. Experimental results on synthetic and real datasets demonstrate that the spectral (k, p)-support norm based method outperforms the nuclear norm based ones.

INDEX TERMS Tensor completion, square deal, atomic norm, sample complexity, APG.

I. INTRODUCTION

Benefiting from its strong multi-linear algebraic background [1], tensor, or multi-dimensional array, has become a powerful tool in modeling multidimensional data like videos [2], hyper-spectral images [3], electroencephalography (EEG) [4], etc, and has attracted great attentions in computer vision, signal processing, data mining, machine learning and bio-informatics in recent years [5]–[11]. In some real tasks, like video in-painting, only a small part of the data tensor is available whereas we want to know the whole tensor. This motivates us to study the tensor completion problem [2] which concerns predicting the missing values of a tensor from its partial observations.

Typically, we consider recovering an underlying K-th order tensor $\mathcal{L}^* \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ from a subset of its entries. Generally speaking, the tensor completion problem is ill-posed, since the unobserved entries can take arbitrary values. However, if we impose some structural assumptions, like low-rankness [2], on the underlying tensor $\mathcal{L}^*$, the problem may be well-posed. Extending the low-rank matrix completion [12], the key idea of low-rank tensor completion [2], [13] is that the underlying tensor is assumed to be the one which has the smallest 'rank' among all tensors having the same observed entries. A general model is to solve a constrained tensor rank minimization problem and use one solution in the solution set to estimate the underlying tensor, i.e.,

$$\begin{aligned} &\min_{\mathcal{L}} \ \operatorname{rank}(\mathcal{L}) \\ &\text{s.t. } \mathcal{L}_{i_1 \cdots i_K} = \mathcal{L}^*_{i_1 \cdots i_K}, \quad \forall (i_1, \cdots, i_K) \in \Omega, \end{aligned} \tag{1}$$

where Ω denotes the set of observed indexes. Since a tensor has many different definitions of rank functions, such as the CP rank [14], Tucker rank [15], Tubal-rank [16], [17], etc, one has many choices of $\operatorname{rank}(\mathcal{L})$ in Problem (1). However, no matter which rank function we choose, the general constrained rank minimization problem is NP hard [18]. In this paper, we assume that the underlying tensor is

low-Tucker-rank, i.e., it is simultaneously low rank along all of its modes [1], [19].

Due to the hardness of directly minimizing the rank function, many convex or non-convex surrogates of the rank function are proposed to replace the rank function in Problem (1) for computational tractability. Among different rank surrogates, the tensor nuclear norm (see §II-B.2) should be the most popular and it has been successfully adopted in image and video in-painting, recommender systems, brain MRI data completion, etc [2], [20]. However the tensor nuclear norm based models have some drawbacks (see §II-B.2) and models based on many new metrics, i.e., the latent nuclear norm [21], Romera-Paredes Pontil relaxation [22], tensor square norm [23], tubal nuclear norm [24], TT-norm [25], etc, are proposed to get better low-rank recovery performances. For different models based on the above rank surrogates, various optimization algorithms are proposed [2], [13]. The related convex methods will be introduced in section II-B. Since this paper only concerns convex tensor completion methods, non-convex methods like [26] and [27], Alt will not be covered. For a summary of some fundamental tensor completion approaches, see §2 in [20].

In this paper, the goal is to complete a partially observed low-Tucker-rank tensor. A novel estimator based on a newly defined 'balanced spectral (k,p)-support norm' will be proposed. An efficient optimization algorithm and the statistical performances will be provided on both the computational side and the statistical side of the proposed estimator. The effectiveness of the proposed estimator will be demonstrated through experiments on synthetic and real datasets.

The remaining of this paper is organized as follows. Notations and related work are introduced in §II. The problem will be formulated in §III where a estimator based on a newly defined 'balanced spectral (k,p)-support norm' will be proposed. In §IV, the optimization algorithm is presented. The statistical performance of the proposed estimator is analyzed in §V. The experimental results are shown in §VI. This work is concluded in §VII.

II. NOTATIONS AND RELATED WORK

A. NOTATIONS AND PRELIMINARIES

The notations in this paper follows that of [1]. For any integer n, we use $[n]$ to denote the set of integers from 1 to n. Vectors are denoted in bold lower case (e.g. $\boldsymbol{x}$); Matrices are denoted in bold upper case (e.g. $\boldsymbol{M}$); Tensors are denoted by bold calligraphic letters (e.g. $\boldsymbol{\mathcal{T}}$). Let $\mathbf{1}(\cdot)$ be the indicator function whose value is 1 if the condition is true and 0 otherwise.

Given a vector $\boldsymbol{x} \in \mathbb{R}^m$, its l_0-(semi-)norm and l_p-norm ($p \in [1,\infty]$) are respectively defined as

$$\|\boldsymbol{x}\|_0 := \sum_{i=1}^{m} \mathbf{1}(x_i \neq 0) \quad \text{and } \|\boldsymbol{x}\|_p := (\sum_{i=1}^{m} |x_i|^p)^{1/p}.$$

Given a matrix $\boldsymbol{M} \in \mathbb{R}_r^{m\times n}$, suppose that it has a singular value decomposition (SVD) as

$$\boldsymbol{M} = \boldsymbol{U}\mathrm{diag}(\boldsymbol{\sigma})\boldsymbol{V}^\top,$$

where $\boldsymbol{U} \in \mathbb{R}^{m\times r}$ and $\boldsymbol{V} \in \mathbb{R}^{n\times r}$ are orthogonal matrices and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \cdots, \sigma_r)^\top \in \mathbb{R}^r$ is the vector of singular values of $\boldsymbol{M}$. The spectral norm, nuclear norm and Schatten-p norm ($p \in [1,\infty]$) of $\boldsymbol{M}$ are respectively defined as

$$\|\boldsymbol{M}\|_2 := \max_i \sigma_i, \quad \|\boldsymbol{M}\|_* := \sum_{i=1}^{m} \sigma_i,$$

and

$$\|\boldsymbol{M}\|_{S_p} := (\sum_{i=1}^{m} \sigma_i^p)^{1/p}.$$

Given a $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{d_1\times\cdots\times d_K}$, its mode-$k$ fibers are the vectors in $\mathbb{R}^{d_k}$ obtained by fixing all indices except for the k-th one. The mode-k *matricization* or *unfolding* of $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{d_1\times\cdots\times d_K}$ is the *matrix* having the mode-k fibers of $\boldsymbol{\mathcal{T}}$ for columns and is denoted by $\boldsymbol{T}_{(k)} \in \mathbb{R}^{d_k\times d_k'}$, where $d_k' = \prod_{j\neq k} d_j$ [1]. The inner product between two tensors $\boldsymbol{\mathcal{T}}_1 \in \mathbb{R}^{d_1\times\cdots\times d_K}$ and $\boldsymbol{\mathcal{T}}_2 \in \mathbb{R}^{d_1\times\cdots\times d_K}$ is defined as

$$\langle \boldsymbol{\mathcal{T}}_1, \boldsymbol{\mathcal{T}}_2 \rangle := \mathrm{vec}(\boldsymbol{\mathcal{T}}_1)^\top \mathrm{vec}(\boldsymbol{\mathcal{T}}_2),$$

where $\mathrm{vec}(\cdot)$ is the vectorization operation of a tensor [1]. The l_1-norm, Frobenius norm and l_∞-norm of a tensor are respectively defined as

$$\|\boldsymbol{\mathcal{T}}\|_1 := \|\mathrm{vec}(\boldsymbol{\mathcal{T}})\|_1, \quad \|\boldsymbol{\mathcal{T}}\|_\mathrm{F} := \|\mathrm{vec}(\boldsymbol{\mathcal{T}})\|_2,$$

and

$$\|\boldsymbol{\mathcal{T}}\|_\infty := \|\mathrm{vec}(\boldsymbol{\mathcal{T}})\|_\infty.$$

The Tucker rank of a K-th order tensor $\boldsymbol{\mathcal{T}}$ is the vector

$$\Big(\mathrm{rank}(\boldsymbol{T}_{(1)}), \cdots, \mathrm{rank}(\boldsymbol{T}_{(K)})\Big)^\top \in \mathbb{R}^K,$$

where $\mathrm{rank}(\boldsymbol{T}_{(k)})$ denotes the rank of the mode-k unfolding. Given $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{d_1\times\cdots\times d_K}$, let $D = \prod_{k=1}^K d_k$ be the number of its entries. For simplicity, other notations are defined when necessary.

B. RELATED WORK

In this subsection, some directly related work will be listed. Before introducing the K-th order tensor completion models, we first introduce the low-rank matrix completion as a special case where $K = 2$.

1) LOW-RANK MATRIX COMPLETION

The low-rank matrix completion problem [12] concerns imputing missing values of a matrix $\boldsymbol{M} \in \mathbb{R}_r^{m\times n}$ from partial observations, supposing that its rank $r \ll \min\{m,n\}$. The main idea to reconstruct a partially observed matrix is the assumption that it has the smallest rank among all matrices with the same observations. This idea results in the following

model that finds a solution of the rank minimization problem to estimate the unknown $\boldsymbol{M}$:

$$\begin{aligned} \min_{\boldsymbol{X}\in\mathbb{R}^{m\times n}} \quad & \operatorname{rank}(\boldsymbol{X}) \\ \text{s.t. } & \boldsymbol{X}_{ij} = \boldsymbol{M}_{ij}, \quad \forall (i,j) \in \Omega, \end{aligned} \tag{2}$$

where Ω is the set of the observed indices. Due to the combinatorial nature of the rank function, Problem (2) is proved to be NP hard [28]. To get a solution in polynomial time, the nuclear norm is proposed as the convex envelope, i.e., the tightest convex relaxation, of the rank function for the rank minimization problem [28], [29]. The nuclear norm based matrix completion model is formulated as follows:

$$\begin{aligned} \min_{\boldsymbol{X}\in\mathbb{R}^{m\times n}} \quad & \|\boldsymbol{X}\|_* \\ \text{s.t. } & \boldsymbol{X}_{ij} = \boldsymbol{M}_{ij}, \quad \forall (i,j) \in \Omega. \end{aligned} \tag{3}$$

It has been proved that when the unknown matrix $\boldsymbol{M}$ satisfies the matrix incoherent conditions and the observation number is of order $O(r\max\{m,n\}\log^2(m+n))$ under uniform sampling, the solution of Problem (3) is $\boldsymbol{M}$ itself with high probability [12], [30]. The nuclear norm based model has been extensively used in different fields, see an overview in [31].

2) TENSOR NUCLEAR NORM BASED TENSOR COMPLETION

To complete a K-th ($K \geq 3$) order tensor $\boldsymbol{\mathcal{L}}^* \in \mathbb{R}^{d_1\times\cdots\times d_K}$, the following low-Tucker-rank tensor completion model is considered:

$$\begin{aligned} \min_{\boldsymbol{\mathcal{L}}} \quad & \sum_{k=1}^{K} \operatorname{rank}(\boldsymbol{L}_{(k)}) \\ \text{s.t. } & \boldsymbol{\mathcal{L}}_{i_1\cdots i_K} = \boldsymbol{\mathcal{L}}^*_{i_1\cdots i_K}, \quad \forall (i_1,\cdots,i_K) \in \Omega, \end{aligned} \tag{4}$$

where $\boldsymbol{L}_{(k)} \in \mathbb{R}^{d_k\times\prod_{j\neq k} d_j}$ is the mode-k unfolding of $\boldsymbol{\mathcal{L}}$. Since Problem (4) is NP hard, the tensor nuclear norm, defined as

$$\|\boldsymbol{\mathcal{L}}\|_* := \sum_{k=1}^{K} \|\boldsymbol{L}_{(k)}\|_*,$$

is proposed to replace item $\sum_{k=1}^{K}\operatorname{rank}(\boldsymbol{L}_{(k)})$ in Problem (4) as a convex relaxation for computational tractability [2], [19]. The tensor nuclear norm based tensor completion model is as follows:

$$\begin{aligned} \min_{\boldsymbol{\mathcal{L}}} \quad & \|\boldsymbol{\mathcal{L}}\|_* \\ \text{s.t. } & \boldsymbol{\mathcal{L}}_{i_1\cdots i_K} = \boldsymbol{\mathcal{L}}^*_{i_1\cdots i_K}, \quad \forall (i_1,\cdots,i_K) \in \Omega, \end{aligned} \tag{5}$$

The tensor nuclear norm provides a natural regularization to encourage low-Tucker-rankness and has been broadly adopted in various applications [2], [19], [32], [33]. However, the tensor nuclear norm is not a convex envelope of $\sum_{k=1}^{K}\operatorname{rank}(\boldsymbol{L}_{(k)})$ [22], [34]. Moreover, as pointed out in [23], to encourage the low-Tucker-rank structure in tensor $\boldsymbol{\mathcal{L}}$, the tensor nuclear norm, i.e., the direct summation of nuclear norms of the unfolding matrices along each mode, is not often significantly powerful than the nuclear norm of a 'square' reshaping of $\boldsymbol{\mathcal{L}}$, which leads to the well-known concept *Square Deal*.

3) SQUARE DEAL

Square deal, proposed by Mu *et al.* [23] seeks a balanced reshaping of a tensor to a matrix which also preserves low-rankness. Given $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{d_1\times\cdots\times d_K}$, let $\overline{\boldsymbol{\mathcal{T}}} \in \mathbb{R}^{d_{l_1}\times d_{l_2}\times d_{l_K}}$ be the tensor obtained by relabeling the l_k-th mode of $\boldsymbol{\mathcal{T}}$ to the k-th mode, where $k \in [K]$ and $(l_1, l_2, \cdots, l_K)$ is a permutation of $(1, 2, \cdots, K)$. Selecting $j \in [K]$, the balanced reshaping matrix $\boldsymbol{T}_{[j]}$ of $\boldsymbol{\mathcal{T}}$ is defined through its permuted version $\overline{\boldsymbol{\mathcal{T}}}$ as follows:

$$\boldsymbol{T}_{[j]} := \operatorname{reshape}\Big(\overline{\boldsymbol{T}}_{(1)}, \prod_{k=1}^{j} d_{l_k}, \prod_{k=j+1}^{K} d_{l_k}\Big) \in \mathbb{R}^{\prod_{k=1}^{j} d_{l_k}\times\prod_{k=j+1}^{K} d_{l_k}}, \tag{6}$$

where reshape$(\cdot)$ is the matrix reshaping operator (see [23] for more details) and $\overline{\boldsymbol{T}}_{(1)}$ is the mode-1 unfolding of $\overline{\boldsymbol{\mathcal{T}}}$. According to Eq. (6), if $j = 1$, then $\boldsymbol{T}_{[j]} = \boldsymbol{T}_{(l_1)}$. If $j > 1$,$\boldsymbol{T}_{[j]}$ becomes a more balanced matrix. Based on the balanced reshaping, a low-rank tensor recovery model based on the minimization the nuclear norm of $\boldsymbol{T}_{[j]}$ is proposed in [23]. For the exact recovery of a K-th order tensor of size $d\times\cdots\times d$ with Tucker rank $(r,\cdots,r)$ under Gaussian measurements, the sample complexity drops from $O(rd^{K-1})$ for the tensor nuclear norm minimization problem [19] to $O(r^{\lfloor\frac{K}{2}\rfloor}d^{\lfloor\frac{K}{2}\rfloor})$ for a square norm minimization problem (see [23], Th. 6).

4) SPECTRAL (k,p)-SUPPORT NORM

The recently proposed spectral (k,p)-support norm [35] has been shown to have flexibility in fitting the decaying pattern of matrix singular values. Given $\boldsymbol{M} \in \mathbb{R}_r^{m\times n}$, suppose that it has an singular value decomposition (SVD) as $\boldsymbol{M} = \boldsymbol{U}\operatorname{diag}(\boldsymbol{\sigma})\boldsymbol{V}^\top$, where $\boldsymbol{U} \in \mathbb{R}^{m\times r}$ and $\boldsymbol{V} \in \mathbb{R}^{n\times r}$ are orthogonal matrices and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \cdots, \sigma_r)^\top \in \mathbb{R}^r$ is the vector of singular values of $\boldsymbol{M}$. The spectral (k,p)-support norm [35] of $\boldsymbol{M}$, denoted by $\|\boldsymbol{M}\|_k^{sp}$, is the gauge function whose unit ball is the convex set

$$\operatorname{conv}\Big\{\boldsymbol{\sigma}\,\Big|\,\|\boldsymbol{\sigma}\|_0 \leq k, \|\boldsymbol{\sigma}\|_p \leq 1\Big\}, \quad k \in \mathbb{N}_+,\ p \geq 1, \tag{7}$$

where conv$(\cdot)$ denotes the convex hall of a set [36]. Thus, the unit ball of the spectral (k,p)-support norm is the convex hull of matrices whose rank is no larger than k and Schatten-p norm no larger than 1. It can be computed as follows [35]:

$$\|\boldsymbol{M}\|_k^{sp} := \|\boldsymbol{\sigma}\|_{(k,p)} = \left(\sum_{i=1}^{k-l-1}\sigma_i^p + \left(\frac{\sum_{i=k-l}^{\min(m,n)}\sigma_i}{(l+1)^{\frac{1}{q}}}\right)^p\right)^{\frac{1}{p}}, \tag{8}$$

where $\frac{1}{p}+\frac{1}{q}=1$, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(m,n)}$ and l is the unique integer in $\{0,\ldots,k-1\}$ satisfying

$$\sigma_{k-l-1} > \frac{1}{l+1}\sum_{i=k-l}^{\min\{m,n\}} \sigma_i \geq \sigma_{k-l}. \tag{9}$$

Note that when $k=1$, the spectral (k,p)-support norm degenerates to the matrix nuclear norm. The spectral (k,p)-support norm based model has shown superior performance over the nuclear norm based models in applications like recommender systems and moving object detection [35], [37]. For matrices of size $m \times n$, the dual norm of the spectral (k,p)-support norm ($k \in \mathbb{N}_+$, $p>1$) is the Ky-Fan (k,q)-norm defined as follows:

$$\|\boldsymbol{Y}\|^{sp*}_{(k,p)} := \sup_{\|\boldsymbol{X}\|^{sp}_{(k,p)}\leq 1} \langle \boldsymbol{X},\boldsymbol{Y}\rangle = \Big(\sum_{i=1}^{k}\sigma_i(\boldsymbol{Y})^p\Big)^{\frac{1}{p}},$$

where $\sigma_1(\boldsymbol{Y}) \geq \sigma_2(\boldsymbol{Y}) \geq \cdots \geq \sigma_k(\boldsymbol{Y})$ are the first k largest singular values of $\boldsymbol{Y}$ [35].

III. PROBLEM FORMULATION

In this section, the observation model will be first introduced and then followed by the proposed estimator. In the remaining of this paper, let $\boldsymbol{\mathcal{L}}^*$ be the underlying tensor. To describe the low-rankness of $\boldsymbol{\mathcal{L}}^*$, its Tucker rank is denoted by $\boldsymbol{r}=(r_1,\cdots,r_K)^\top$, where $r_l=\text{rank}(\boldsymbol{L}^*_{(l)}) \ll \min\{n_l,n_l'\}$, for all $l\in[K]$.

A. THE OBSERVATION MODEL

Suppose that we have N scalar observations $y_1,\cdots,y_N$ generated through the following observation model:

$$y_i = \langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{L}}^*\rangle + \varsigma\xi_i, \quad \forall i\in[N], \tag{10}$$

where $\varsigma>0$ is a known standard deviation parameter, $\{\boldsymbol{\mathcal{X}}_i\}_{i=1}^N$ are known random design tensors and $\{\xi_i\}_{i=1}^m$ are unknown normalized random noise variables. In this paper, we make the following tow assumptions:

Assumption 1 (Uniform Sampling): The design tensors $\boldsymbol{\mathcal{X}}_i$ are i.i.d. sampled with replacement from a uniform distribution on the set tensor standard basis, i.e.,

$$\boldsymbol{\mathcal{X}}_i \sim \textit{Uniform}\Big(\big\{\boldsymbol{e}_{j_1}\circ\cdots\circ\boldsymbol{e}_{j_K} : \ \forall l\in[K],\ \forall j_l\in[d_l]\big\}\Big).$$

Assumption 2 (Sub-Gaussian Noise): The noise variables ξ_i are i.i.d. centered Sub-Gaussian variables with unit variance, i.e., there exists a positive K_0 such that

$$\mathbb{E}[\exp(K_0\xi_i^2)] \leq e, \quad \forall i\in[N].$$

Let $\boldsymbol{y}=(y_1,\cdots,y_N)^\top\in\mathbb{R}^N$ and $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_N)^\top\in\mathbb{R}^N$ be the vector of observations and noise, respectively. Define the observation operator $\mathfrak{X}:\mathbb{R}^{d_1\times\cdots\times d_K}\to\mathbb{R}^N$ as

$$\mathfrak{X}(\boldsymbol{\mathcal{T}}) := \Big(\langle\boldsymbol{\mathcal{T}},\boldsymbol{\mathcal{X}}_1\rangle,\cdots,\langle\boldsymbol{\mathcal{T}},\boldsymbol{\mathcal{X}}_N\rangle\Big)^\top, \tag{11}$$

for any tensor $\boldsymbol{\mathcal{T}}\in\mathbb{R}^{d_1\times\cdots\times d_K}$. Then the observation model (10) can be rewritten in a compact form

$$\boldsymbol{y} = \mathfrak{X}(\boldsymbol{\mathcal{L}}^*) + \varsigma\boldsymbol{\xi}.$$

In addition, the adjoint operator of $\mathfrak{X}(\cdot)$ denoted by $\mathfrak{X}^*(\cdot):\mathbb{R}^N\to\mathbb{R}^{d_1\times\cdots\times d_K}$ is defined as

$$\mathfrak{X}^*(z) := \sum_{i=1}^N z_i\boldsymbol{\mathcal{X}}_i, \tag{12}$$

for any vector $z\in\mathbb{R}^N$.

B. THE PROPOSED ESTIMATOR

The goal is to estimate the underlying tensor $\boldsymbol{\mathcal{L}}^*\in\mathbb{R}^{d_1\times\cdots\times d_K}$ given incomplete noisy measurements $\boldsymbol{y}$ from the observation model (10).

Motivated by the superiority of the spectral (k,p)-support norm based model over the nuclear norm model for low rank matrix recovery, we will adopt the spectral (k,p)-support norm in our low rank tensor completion problem. Following the spirit of square deal, we consider *the most squared reshaping matrix* of a given tensor and minimize the spectral (k,p)-support norm of it. A balanced reshaping matrix $\boldsymbol{T}_{[j]}$ of $\boldsymbol{\mathcal{T}}\in\mathbb{R}^{d_1\times\cdots\times d_K}$ is said to belong to the set of most squared reshaping matrices if the difference between its row and column is the smallest over all $j\in[K]$ and all permutations $(l_1,l_2,\cdots,l_K)$ of $(1,2,\cdots,K)$. Given $\boldsymbol{\mathcal{T}}\in\mathbb{R}^{d_1\times\cdots\times d_K}$, we denote one of its most balanced unfolding by $\boldsymbol{T}_\diamond\in\mathbb{R}^{D_1\times D_2}$, where $D_1\geq D_2$ and $D_1D_2=\prod_{l=1}^K d_l$.

In order to estimate the true but known $\boldsymbol{\mathcal{L}}^*$, the following estimator is defined:

$$\begin{aligned}\hat{\boldsymbol{\mathcal{L}}} \in \ &\underset{\|\boldsymbol{\mathcal{L}}\|_\infty\leq\alpha}{\text{argmin}} \ \|\boldsymbol{L}_\diamond\|^{sp}_{(k,p)}\\ &\text{s.t. } \|\boldsymbol{y}-\mathfrak{X}(\boldsymbol{\mathcal{L}})\|_2 \leq B_{cn}.\end{aligned} \tag{13}$$

In Eq. (13), $\alpha>0$ is a constant that upper estimates the l_∞-norm of $\boldsymbol{\mathcal{L}}^*$, the constraint $\|\boldsymbol{\mathcal{L}}\|_\infty\leq\alpha$ is used to exclude the 'over-spiky' tensors (see [5], [38]), and $B_{cn}>0$ serves as a tolerance for noise.

IV. OPTIMIZATION ALGORITHM

In this section, we will first formulate the Lagrangian dual of Problem (13) and then solve the Lagrangian dual using accelerated proximal (sub-)gradient (APG) algorithm. The complexity and convergence of the proposed algorithm will also be analyzed.

A. THE LAGRANGIAN DUAL

Problem (13) can be solved by the popular alternative direction methods of multipliers (ADMM) [39] which involves a full SVD of $\boldsymbol{L}_\diamond$ to compute the spectral (k,p)-support norm with time cost $O(D_1D_2\min\{D_1,D_2\})$ in each iteration and is thus very time consuming. To avoid the expensive full SVD in each iteration, we consider the dual problem of Problem (13) which can get rid of computing the spectral (k,p)-support norm. The same strategy has been used in [37] for spectral

k-support norm minimization and in [40] for nuclear norm minimization.

By introducing auxiliary variables $\tau > 0$, $\boldsymbol{g} \in \mathbb{R}^N$ and $\boldsymbol{\mathcal{K}} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$, we obtain an equivalent formulation of Problem (13):

$$\begin{aligned} \min_{\tau,\boldsymbol{g},\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{K}}} \ & \tau \\ \text{s.t. } & \|\boldsymbol{L}_\diamond\|_{(k,p)}^{sp} \le \tau \le B_{sp}, \\ & \|\boldsymbol{\mathcal{K}}\|_\infty \le \alpha, \quad \|\boldsymbol{g}\|_2 \le B_{cn}, \\ & \boldsymbol{\mathcal{L}} = \boldsymbol{\mathcal{K}}, \quad \mathfrak{X}(\boldsymbol{\mathcal{L}}) + \boldsymbol{g} = \boldsymbol{y}. \end{aligned} \tag{14}$$

where B_{sp} is a constant estimation of the upper bound of $\|\boldsymbol{L}_\diamond\|_{(k,p)}^{sp}$.

Denoting the Lagrangian multipliers by $\boldsymbol{\lambda} \in \mathbb{R}^N$ and $\boldsymbol{\Lambda} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$, the Lagrangian of Problem (14) is formulated as

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{\Lambda}} \min_{\tau,\boldsymbol{g},\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{K}}} & L(\tau, \boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{K}}, \boldsymbol{g}, \boldsymbol{\Lambda}, \boldsymbol{\lambda}) \\ & = \tau + \langle \boldsymbol{\Lambda}, \boldsymbol{\mathcal{L}} - \boldsymbol{\mathcal{K}} \rangle + \langle \boldsymbol{\lambda}, \boldsymbol{y} - \mathfrak{X}(\boldsymbol{\mathcal{L}}) - \boldsymbol{g} \rangle, \\ \text{s.t. } & \|\boldsymbol{L}_\diamond\|_{(k,p)}^{sp} \le \tau \le B_{sp}, \quad \|\boldsymbol{\mathcal{K}}\|_\infty \le \alpha, \ \|\boldsymbol{g}\|_2 \le B_{cn}. \end{aligned} \tag{15}$$

Next, we will solve the maximization problem in Problem (15). Note that since $\|\boldsymbol{L}_\diamond\|_{(k,p)}^{sp} \le \tau \le B_{sp}$, $\|\boldsymbol{\mathcal{K}}\|_\infty \le \alpha$, $\|\boldsymbol{g}\|_2 \le B_{cn}$, we have

$$\begin{aligned} & L(\tau, \boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{K}}, \boldsymbol{g}, \boldsymbol{\Lambda}, \boldsymbol{\lambda}) \\ & \overset{(i1)}{=} \tau + \left\langle \boldsymbol{\Lambda} - \mathfrak{X}^*(\boldsymbol{\lambda}), \boldsymbol{\mathcal{L}} \right\rangle - \langle \boldsymbol{\Lambda}, \boldsymbol{\mathcal{K}} \rangle - \langle \boldsymbol{\lambda}, \boldsymbol{g} \rangle + \langle \boldsymbol{\lambda}, \boldsymbol{y} \rangle \\ & \overset{(i2)}{\ge} \|\boldsymbol{L}_\diamond\|_{(k,p)}^{sp} - \|\left(\boldsymbol{\Lambda} - \mathfrak{X}^*(\boldsymbol{\lambda})\right)_\diamond\|_{(k,p)}^{sp*} \|\boldsymbol{L}_\diamond\|_{(k,p)}^{sp} \\ & \quad - \|\boldsymbol{\Lambda}\|_1 \|\boldsymbol{\mathcal{K}}\|_\infty - \|\boldsymbol{\lambda}\|_2 \|\boldsymbol{g}\|_2 + \langle \boldsymbol{\lambda}, \boldsymbol{y} \rangle \\ & \overset{(i3)}{\ge} \|\boldsymbol{L}_\diamond\|_{(k,p)}^{sp} \left(1 - \|\left(\boldsymbol{\Lambda} - \mathfrak{X}^*(\boldsymbol{\lambda})\right)_\diamond\|_{(k,p)}^{sp*}\right) + \langle \boldsymbol{\lambda}, \boldsymbol{y} \rangle \\ & \quad - \alpha \|\boldsymbol{\Lambda}\|_1 - B_{cn} \|\boldsymbol{\lambda}\|_2 \\ & \overset{(i4)}{\ge} \underbrace{\langle \boldsymbol{\lambda}, \boldsymbol{y} \rangle - B_{sp} \max\left\{0, \|\left(\boldsymbol{\Lambda} - \mathfrak{X}^*(\boldsymbol{\lambda})\right)_\diamond\|_{(k,p)}^{sp*} - 1\right\}}_{=:-f(\boldsymbol{\Lambda},\boldsymbol{\lambda})} \\ & \quad - \underbrace{\alpha \|\boldsymbol{\Lambda}\|_1}_{=:r_1(\boldsymbol{\Lambda})} - \underbrace{B_{cn} \|\boldsymbol{\lambda}\|_2}_{=:r_2(\boldsymbol{\lambda})}, \end{aligned}$$

where equality $(i1)$ holds because of the definitions of $\mathfrak{X}(\cdot)$ $\mathfrak{X}^*(\cdot)$ in Eqs. (11) (12), inequality $(i2)$ holds because of the Holder-like inequality of dual norms and $\|\boldsymbol{L}_\diamond\|_{(k,p)}^{sp} \le \tau$, inequality $(i3)$ holds because that $\|\boldsymbol{\mathcal{K}}\|_\infty \le \alpha$ and $\|\boldsymbol{g}\|_2 \le B_{cn}$, inequality $(i4)$ holds because of $\|\boldsymbol{L}_\diamond\|_{(k,p)}^{sp} \le B_{cn}$.

Thus, we have

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{\Lambda}} \min_{\tau,\boldsymbol{g},\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{K}}} & L(\tau, \boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{K}}, \boldsymbol{g}, \boldsymbol{\Lambda}, \boldsymbol{\lambda}) \\ & = \max_{\boldsymbol{\lambda},\boldsymbol{\Lambda}} -f(\boldsymbol{\Lambda}, \boldsymbol{\lambda}) - r_1(\boldsymbol{\Lambda}) - r_2(\boldsymbol{\lambda}), \end{aligned}$$

which is equivalent to

$$\min_{\boldsymbol{\lambda},\boldsymbol{\Lambda}} f(\boldsymbol{\Lambda}, \boldsymbol{\lambda}) + r_1(\boldsymbol{\Lambda}) + r_2(\boldsymbol{\lambda}), \tag{16}$$

where

$$\begin{aligned} f(\boldsymbol{\Lambda}, \boldsymbol{\lambda}) = \max\left\{0, B_{sp}(\|\left(\boldsymbol{\Lambda} + \mathfrak{X}^*(\boldsymbol{\lambda})\right)_\diamond\|_{(k,p)}^{sp*} - 1)\right\} - \langle \boldsymbol{\lambda}, \boldsymbol{y} \rangle, \\ \text{and } r_1(\boldsymbol{\Lambda}) = \alpha \|\boldsymbol{\Lambda}\|_1, \quad r_2(\boldsymbol{\lambda}) = B_{cn} \|\boldsymbol{\lambda}\|_2. \end{aligned}$$

B. APG SOLVER FOR THE LAGRANGIAN DUAL

We will solve Problem (16) with the accelerated proximal gradient descent (APG) algorithm summarized in Algorithm 1. In detail, we keep two interpolation sequences $(\hat{\boldsymbol{\Lambda}}_t, \hat{\boldsymbol{\lambda}}_t)$ and $(\boldsymbol{\Lambda}_t, \boldsymbol{\lambda}_t)$, which is typical for APG-type methods. Specifically, in each iteration, we iteratively solve the $(\boldsymbol{\Lambda}, \boldsymbol{\lambda})$-subproblem and the $(\hat{\boldsymbol{\Lambda}}, \hat{\boldsymbol{\lambda}})$-subproblem.

1) THE $(\boldsymbol{\Lambda}, \boldsymbol{\lambda})$-SUBPROBLEM

In APG, the dual variables $\boldsymbol{\Lambda}$ and $\boldsymbol{\lambda}$ are updated iteratively through approximating $f(\boldsymbol{\Lambda}, \boldsymbol{\lambda})$ with a quadratic function $Q_{H_{t+1}}\left((\boldsymbol{\Lambda}, \boldsymbol{\lambda}); (\hat{\boldsymbol{\Lambda}}_t, \hat{\boldsymbol{\lambda}}_t)\right)$ which is defined as follows:

$$\begin{aligned} & Q_{H_{t+1}}\left((\boldsymbol{\Lambda}, \boldsymbol{\lambda}); (\hat{\boldsymbol{\Lambda}}_t, \hat{\boldsymbol{\lambda}}_t)\right) \\ & := f(\hat{\boldsymbol{\Lambda}}_t, \hat{\boldsymbol{\lambda}}_t) + \left\langle g(\hat{\boldsymbol{\Lambda}}_t, \hat{\boldsymbol{\lambda}}_t), (\boldsymbol{\Lambda}, \hat{\boldsymbol{\lambda}}) - (\hat{\boldsymbol{\Lambda}}_t, \hat{\boldsymbol{\lambda}}_t) \right\rangle \\ & \quad + \frac{H_{t+1}}{2} \|(\boldsymbol{\Lambda}, \hat{\boldsymbol{\lambda}}) - (\hat{\boldsymbol{\Lambda}}_t, \hat{\boldsymbol{\lambda}}_t)\|_F^2. \end{aligned} \tag{17}$$

where H_t denotes the reciprocal of the step size which should be carefully tuned via line-search, and the (sub-)gradient $g(\hat{\boldsymbol{\Lambda}}_t, \hat{\boldsymbol{\lambda}}_t)$ can be computed by Lemmas 1 and 2.

Lemma 1: A particular sub-gradient of $f(\boldsymbol{\Lambda}, \boldsymbol{\lambda})$ can be given by:

$$g(\boldsymbol{\Lambda}, \boldsymbol{\lambda}) = \begin{cases} (\boldsymbol{0}, -\boldsymbol{y}), & \|\left(\boldsymbol{\Lambda} + \mathfrak{X}^*(\boldsymbol{\lambda})\right)_\diamond\|_{(k,p)}^{sp*} \le 1, \\ g^\#, & \text{otherwise}, \end{cases} \tag{18}$$

where

$$g^\# = \left(B_{sp} \boldsymbol{\mathcal{L}}^\#, B_{sp} \mathfrak{X}(\boldsymbol{\mathcal{L}}^\#) - \boldsymbol{y}\right),$$

and

$$\boldsymbol{L}^\# := (\boldsymbol{\mathcal{L}}^\#)_\diamond = argmax_{\|\boldsymbol{A}\|_{(k,p)}^{sp*} \le 1} \left\langle \left(\boldsymbol{\Lambda} + \mathfrak{X}^*(\boldsymbol{\lambda})\right)_\diamond, \boldsymbol{A} \right\rangle. \tag{19}$$

Lemma 2 [37]: A closed-form solution of Problem (19) can be given by $\boldsymbol{U} diag(\boldsymbol{s}) \boldsymbol{V}^\top$, where

$$s_i = \begin{cases} \dfrac{\sigma_i}{\|\left(\boldsymbol{\Lambda} + \mathfrak{X}^*(\boldsymbol{\lambda})\right)_\diamond\|_{(k,p)}^{sp*}}, & i = 1, \ldots, k, \\ 0, & i = k+1, \cdots, \min(m, n). \end{cases} \tag{20}$$

and $\left(\boldsymbol{\Lambda} + \mathfrak{X}^*(\boldsymbol{\lambda})\right)_\diamond = \boldsymbol{U} diag(\sigma) \boldsymbol{V}^\top$ is a particular SVD of matrix $\left(\boldsymbol{\Lambda} + \mathfrak{X}^*(\boldsymbol{\lambda})\right)_\diamond$.

According to Eq. (17) and Lemmas 1 and 2, the dual variables $\boldsymbol{\Lambda}_{t+1}$ and $\boldsymbol{\lambda}_{t+1}$ can be updated in closed-form as

$$\begin{aligned} \boldsymbol{\Lambda}_{t+1} &= \operatorname*{argmin}_{\boldsymbol{\Lambda}} Q_{H_{t+1}}((\boldsymbol{\Lambda}, \boldsymbol{\lambda}); (\hat{\boldsymbol{\Lambda}}_t, \hat{\boldsymbol{\lambda}}_t)) + r_1(\boldsymbol{\Lambda}) \\ &= \text{sign}(\tilde{\boldsymbol{\Lambda}}) \odot \max(|\tilde{\boldsymbol{\Lambda}}| - \bar{\tau}, 0). \end{aligned} \tag{21}$$

and

$$\begin{aligned} \boldsymbol{\lambda}_{t+1} &= \operatorname*{argmin}_{\boldsymbol{\lambda}} Q_{H_{t+1}}((\boldsymbol{\Lambda}, \boldsymbol{\lambda}); (\hat{\boldsymbol{\Lambda}}_t, \hat{\boldsymbol{\lambda}}_t)) + r_2(\boldsymbol{\lambda}), \\ &= \begin{cases} \max\{1 - \tilde{\tau}/\|\tilde{\boldsymbol{\lambda}}\|_2, 0\} \tilde{\boldsymbol{\lambda}}, & \|\tilde{\boldsymbol{\lambda}}\|_2 \ne 0, \\ \boldsymbol{0}, & \text{otherwise}, \end{cases} \end{aligned} \tag{22}$$

where $\odot$ denotes the element-wise multiplication and $g(\hat{\boldsymbol{\Lambda}}_t, \hat{\boldsymbol{\lambda}}_t) = (g_{\boldsymbol{\Lambda}_t}, g_{\boldsymbol{\lambda}_t})$, $\tilde{\boldsymbol{\Lambda}} = \hat{\boldsymbol{\Lambda}}_t - g_{\boldsymbol{\Lambda}_t}/H_{t+1}$, $\bar{\tau} = \alpha/(NH_{t+1})$, $\tilde{\boldsymbol{\lambda}} = \hat{\boldsymbol{\lambda}}_t - g_{\boldsymbol{\lambda}_t}/H_{t+1}$, $\tilde{\tau} = B_{cn}/H_{t+1}$.

2) THE $(\hat{\Lambda},\hat{\lambda})$-SUBPROBLEM

According to [41], the variable $(\hat{\Lambda}_{t+1},\hat{\lambda}_{t+1})$ is updated as an extrapolation of (Λ_t,λ_t) and $(\Lambda_{t+1},\lambda_{t+1})$, the $(\hat{\Lambda},\hat{\lambda})$-subproblem can be solved by

$$\hat{\Lambda}_{t+1}=\Lambda_{t+1}+\frac{\beta_t-1}{\beta_{t+1}}(\Lambda_{t+1}-\Lambda_t), \tag{23}$$

and

$$\hat{\lambda}_{t+1}=\lambda_{t+1}+\frac{\beta_t-1}{\beta_{t+1}}(\lambda_{t+1}-\lambda_t), \tag{24}$$

where $\{\beta_t\}$ is a scalar sequence which can be updated iteratively as follows

$$\beta_{t+1}=\frac{1+\sqrt{1+4\beta_t^2}}{2}, \tag{25}$$

with an initial value 1.

Algorithm 1 APG Solver for Problem (16)

Require: $\mathcal{L}_0=\mathcal{O},\lambda_0,\hat{\lambda}_0,\beta_0=1,\nu,B_{cn},B_{sp},\epsilon>0,T_{\max}$.
1: Set $t=0$.
2: **while** $t\le T_{\max}$ **do**
3: Compute $\mathcal{L}^{\#}$ by Eq. (20) at $(\hat{\Lambda}_t,\hat{\lambda}_t)$;
4: Compute $g(\hat{\Lambda}_t,\hat{\lambda}_t)$ by Eq. (18);
5: Compute H_{t+1} by **line-search** $((\hat{\Lambda}_t,\hat{\lambda}_t),g(\hat{\Lambda}_t,\hat{\lambda}_t),H_t,\epsilon,\beta_t)$;
6: Update Λ_{t+1} by Eq. (21);
7: Update λ_{t+1} by Eq. (22);
8: Update $\mathcal{L}_{t+1}$ by Eq. (26);
9: Update $\hat{\Lambda}_{t+1}$ and $\hat{\lambda}_{t+1}$ by Eqs. (23) and (24);
10: $t=t+1$.
11: **end while**
Ensure: $\mathcal{L}_t$.

3) LINE-SEARCH SUBROUTINE

The line search subroutine aims at finding the smallest H_{t+1} which satisfies

$$f(\Lambda_{t+1},\lambda_{t+1})\le Q_{H_{t+1}}((\Lambda,\lambda);(\hat{\Lambda}_t,\hat{\lambda}_t))+\frac{\epsilon}{2\beta_t}.$$

The line-search step is summarized in Algorithm 2 which follows [37].

4) PRIMAL VARIABLE RECOVERY

While carrying out a dual iteration, we simultaneously recover the primal variable $\mathcal{L}$ as follows:

$$\mathcal{L}_{t+1}=(1-\gamma_t)\mathcal{L}_t+\gamma_t B_{sp}\mathcal{L}^{\#}, \tag{26}$$

where $\mathcal{L}^{\#}$ is computed through Eq. (19) and γ_t is a weighting parameter which can be updated as follows:

$$\gamma_t=\frac{\beta_t/H_t}{\sum_{i=1}^{t}\beta_i/H_i}. \tag{27}$$

Algorithm 2 Line-Search Subroutine

Require: $(\hat{\Lambda},\hat{\lambda}),g(\hat{\Lambda},\hat{\lambda}),H_0,\epsilon,\beta$.
1: **for** $i=1,\cdots,I_{max}$ **do**
2: Compute new point $(\Lambda_{i+1},\lambda_{i+1})$ using (21) and (22) at $(\hat{\Lambda}_t,\hat{\lambda}_t)$ with penalty parameter H_i;
3: **if** $f(\Lambda_{i+1},\lambda_{i+1})\le Q_{H_i}(\Lambda_{i+1},\lambda_{i+1};\hat{\Lambda}_t,\hat{\lambda}_t)+\frac{\epsilon}{2\beta}$ **then**
4: break;
5: **else**
6: $H_{i+1}=2H_i$.
7: **end if**
8: **end for**
Ensure: $(\hat{\Lambda}_t,\hat{\lambda}_t),H_i$.

C. COMPLEXITY AND CONVERGENCE ANALYSIS

Note that in Algorithm 1, computation of $\mathcal{L}^{\#}$ costs $O(kD)$, computation of $g(\hat{\Lambda}_t,\hat{\lambda}_t)$ costs $O(D+N)$, computation of H_{t+1} costs at most $O(I_{max}kD)$ according to Algorithm 2, computations of $\Lambda_{t+1},\lambda_{t+1},\mathcal{L}_{t+1},\hat{\Lambda}_{t+1}$ and $\hat{\lambda}_{t+1}$ cost $O(D)$, $O(N)$, $O(D)$, $O(D)$ and $O(N)$, respectively. Recall that D denotes $\prod_{i=1}^{K}d_i$. Thus, the per-iteration time complexity is dominated by

$$O\left((I_{max}+1)k\prod_{i=1}^{K}d_i\right),$$

which indicates that the proposed algorithm has scalability in high dimensionality.

The convergence of Algorithm 1 is a direct consequence of the analysis in [41].

Theorem 1: The sequence of primal variables $\{\mathcal{L}_t\}$ generated by Algorithm 1 converges to a stationary point $\hat{\mathcal{L}}$, and the iteration number in the worst case to achieve an ϵ-solution is

$$T_{\max}=O\left(\inf_{0\le\nu\le1}\left(\frac{M_\nu}{\epsilon}\right)^{\frac{2}{1+\nu}}\right),$$

where M_ν is defined as

$$M_\nu:=\sup_{(\Lambda,\lambda)\neq(\tilde{\Lambda},\tilde{\lambda})\in\mathrm{Dom}(f)}\frac{\|\nabla f(\Lambda,\lambda)-\nabla f(\tilde{\Lambda},\tilde{\lambda})\|}{\|(\Lambda,\lambda)-(\tilde{\Lambda},\tilde{\lambda})\|_F^{\nu}}.$$

V. STATISTICAL PERFORMANCE OF THE PROPOSED ESTIMATOR

In this section, we study the statistical performance of the proposed estimator. Specifically, we establish the upper bounds on the sample complexity and the estimation error. For the ease of notation, we consider the K-th order tensor of size $d\times\cdots\times d$ with Tucker rank $(r,\cdots,r)^\top\in\mathbb{R}^K$. Given $\mathcal{L}\in\mathbb{R}^{d\times\cdots\times d}$, its most squared reshaping matrix $\boldsymbol{L}_\diamond$ is of size $d^{\lceil K/2\rceil}\times d^{\lfloor K/2\rfloor}$.

Our analysis follows the unified framework of analysis of matrix completion under general low dimensional constraints induced by any norm regularization provided by Gunasekar *et al.* in [42]. In the framework, the bounds on the sample complexity and the estimation error are given in term of a well understood complexity measure of Gaussian width

of a restricted error set associated with $\mathcal{L}^*$. First, the Gaussian width is defined as follows.

Definition 1 (Gaussian Width): Gaussian width of a set $\mathcal{S} \subset \mathbb{R}^{d_1 \times \cdots \times d_K}$ is defined as:

$$w_g(\mathcal{S}) = \mathbb{E}_{\mathcal{G}} \sup_{\mathcal{X} \in \mathcal{S}} \langle \mathcal{X}, \mathcal{G} \rangle,$$

where $\mathcal{G} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is a tensor of independent standard Gaussian random variables.

We then define the restricted error cone and its subset

$$\mathcal{T}(\mathcal{L}^*) = \text{cone}\Big\{\Delta \in \mathbb{R}^{d_1 \times \cdots \times d_K} : \|\boldsymbol{L}_\diamond^* + \Delta_\diamond\|_{(k,p)}^{sp} \le \|\boldsymbol{L}_\diamond^*\|_{(k,p)}^{sp}\Big\},$$

and

$$\mathcal{E}(\mathcal{L}^*) = \mathcal{T}(\mathcal{L}^*) \cap \Big\{\mathcal{X} \in \mathbb{R}^{d_1 \times \cdots \times d_K} : \|\mathcal{X}\|_{\text{F}} = 1\Big\}.$$

With a slight abuse of notation, we use $w_g(\mathcal{E}(\boldsymbol{L}_\diamond^*))$ to denote $w_g(\mathcal{E}(\mathcal{L}^*))$, since $w_g(\mathcal{E}(\boldsymbol{L}_\diamond^*))$ can be upper bounded in the manner of [42, Lemma 3] . We then have the bound of $w_g(\mathcal{E}(\boldsymbol{L}_\diamond^*))$ as follows:

Lemma 3 (Bound of $w_g^2(\mathcal{E}(\boldsymbol{L}_\diamond^*))$): Let $\boldsymbol{\sigma}^* \in \mathbb{R}^{\lfloor K/2 \rfloor}$ denote the vector of singular values of $\boldsymbol{L}_\diamond^*$ which are sorted in non-descending order. Let $l \in \{0, 1, \cdots, k-1\}$ be the unique integer that satisfies:

$$\sigma_{k-l-1}^* > \frac{1}{l+1} \sum_{i=k-1}^{\lfloor K/2 \rfloor} \sigma_i^* \ge \sigma_{k-l}^*,$$

then we have

$$w_g^2(\mathcal{E}(\boldsymbol{L}_\diamond^*)) \le \Big(A + 2r^{\lfloor K/2 \rfloor} + l - k + 1\Big)\big(2d^{\lceil K/2 \rceil} - r^{\lfloor K/2 \rfloor}\big), \tag{28}$$

where

$$A = \frac{(l+1)^{\frac{2p}{q}} \sum_{i=1}^{k-l-1} (\sigma_i^*)^{2(p-1)}}{(\sum_{i=k-l}^{r^{\lfloor K/2 \rfloor}} \sigma_i^*)^{2(p-1)}}.$$

To prove the above lemma, we need the following two lemmas:

Lemma 4 [42]: For a non-empty convex cone $\mathcal{C} \in \mathbb{R}^{m \times n}$, we have the following relationship:

$$w_g(\mathcal{C} \cap \mathbb{S}^{mn-1}) \le \mathbb{E}_{\boldsymbol{G}}[\inf_{\boldsymbol{X} \in \mathcal{C}^\circ} \|\boldsymbol{G} - \boldsymbol{X}\|_F],$$

where $\mathbb{S}^{mn-1} = \{\boldsymbol{X} \in \mathbb{R}^{m \times n} : \|\boldsymbol{X}\|_F = 1\}$, $\boldsymbol{G} \in \mathbb{R}^{m \times n}$ is a matrix of independent standard Gaussian random variables and $\mathcal{C}^\circ$ is the polar cone of $\mathcal{C}$ defined as $\mathcal{C}^\circ := \{\boldsymbol{Y} : \langle \boldsymbol{X}, \boldsymbol{Y} \rangle \le 0, \forall \boldsymbol{X} \in \mathcal{C}\}$.

The sub-gradient of the spectral (k,p)-support norm is provided in the following lemma that can be achieved according to [43].

Lemma 5: For any matrix $\boldsymbol{X} \in \mathbb{R}_{\tilde{r}}^{m \times n}$ having a singular value decomposition $\boldsymbol{X} = \boldsymbol{U}_X diag(\tilde{\boldsymbol{\sigma}}) \boldsymbol{V}_X{}^\top$, where the singular values are sorted in non-descending order. Let $l \in \{0, 1, \cdots, k-1\}$ be the unique integer satisfying: $\tilde{\sigma}_{k-l-1} > \frac{1}{l+1} \sum_{i=k-1}^{\lfloor K/2 \rfloor} \tilde{\sigma}_i \ge \tilde{\sigma}_{k-l}$, then we have

$$\partial \|\boldsymbol{X}\|_{(k,p)}^{sp} = \left\{ \frac{\boldsymbol{U}_X diag(\boldsymbol{\sigma}) \boldsymbol{V}_X^\top}{(\|\boldsymbol{X}\|_{(k,p)}^{sp})^{1/q}} \right\}, \tag{29}$$

where $\boldsymbol{\sigma}$ in Eq. (29) satisfies

$$\sigma_i \begin{cases} = \tilde{\sigma}_i^{p-1}, & 1 \le i \le k-l-1 \\ = \tilde{\varsigma}, & k-l \le i \le \tilde{r} \\ \le \tilde{\varsigma}, & \tilde{r} \le i \le \min\{d_1, d_2\} \end{cases}$$

and $\tilde{\varsigma} = (k-l)^{-p/q} (\sum_{j=k-l}^{\tilde{r}} \tilde{\sigma}_j)^{p-1}$.

Proof of Lemma 3: The error cone for $\|\cdot\|_{(k,p)}^{sp}$ at $\boldsymbol{L}_\diamond^*$ is given by the tangent cone

$$\mathcal{T} = \text{cone}\Big\{\|\Delta + \boldsymbol{L}_\diamond^*\|_{(k,p)}^{sp} \le \|\boldsymbol{L}_\diamond^*\|_{(k,p)}^{sp}\Big\},$$

and the polar of the tangent cone, i.e., the normal cone is given by

$$\begin{aligned} \mathcal{T}^* = \mathcal{N}(\boldsymbol{L}_\diamond^*) &= \Big\{\boldsymbol{Y} : \langle \boldsymbol{X}, \boldsymbol{Y} \rangle \le 0, \forall \boldsymbol{X} \in \mathcal{T}\Big\} \\ &= \text{cone}(\partial \|\boldsymbol{L}_\diamond^*\|_{(k,p)}^{sp}). \end{aligned}$$

Let $\boldsymbol{L}_\diamond^* = \boldsymbol{U}_1^* \Sigma^* \boldsymbol{V}^{*\top}$ be the full singular value decomposition of $\boldsymbol{L}_\diamond^*$, such that $\boldsymbol{\sigma}^* = \text{diag}(\Sigma^*) \in \mathbb{R}^{d^{\lfloor K/2 \rfloor}}$, and let $\sigma_1^* \ge \sigma_2^* \ge \cdots \sigma_{d^{\lfloor K/2 \rfloor}}^*$. Let $\boldsymbol{u}_i^*$ and $\boldsymbol{v}_i^*$ for $i \in [d^{\lfloor K/2 \rfloor}]$ denote the i-th column of $\boldsymbol{U}^*$ and $\boldsymbol{V}^*$, respectively. Further, the rank of $\boldsymbol{L}_\diamond^*$ is $\|\boldsymbol{\sigma}^*\|_0 \le r^{\lfloor K/2 \rfloor}$, according to [23, Lemma 4].

Denote $I_2 = \{1, \cdots, k-l-1\}$, $I_1 = \{k-r, \cdots, r^{\lfloor K/2 \rfloor}\}$, and $I_0 = \{r^{\lfloor K/2 \rfloor} + 1, \cdots, d^{\lfloor K/2 \rfloor}\}$. Also define the subspace:

$$\begin{aligned} \mathcal{M} = \text{span}\Big\{\boldsymbol{u}_i^* \boldsymbol{x}^\top : i \in I_2 \cup I_1, \forall \boldsymbol{x} \in \mathbb{R}^{\lceil K/2 \rceil}\Big\} \\ \cup \text{span}\Big\{\boldsymbol{y} \boldsymbol{v}_i^{*\top} : i \in I_2 \cup I_1, \forall \boldsymbol{y} \in \mathbb{R}^{\lfloor K/2 \rfloor}\Big\}. \end{aligned}$$

Let $\mathcal{M}^\perp$ be the subspace orthogonal to $\mathcal{M}$ and let $P_{\mathcal{M}}$ and $P_{\mathcal{M}^\perp}$ be the projections onto $\mathcal{M}$ and $\mathcal{M}^\perp$,respectively. From Eq. (29) we have

$$\mathcal{N}(\boldsymbol{L}_\diamond^*) = \Big\{t \boldsymbol{U}^* \text{diag}(\boldsymbol{\sigma}) \boldsymbol{V}^{*\top}\Big\},$$

where $\boldsymbol{\sigma}$ satisfies

$$\sigma_i \begin{cases} = \dfrac{\tilde{\sigma}_i^{p-1}}{\tilde{\varsigma}}, & 1 \le i \le k-l-1 \\ = 1, & k-l \le i \le \tilde{r} \\ \le 1, & \tilde{r} \le i \le \min\{d_1, d_2\} \end{cases}$$

and $\tilde{\varsigma} = (k-l)^{-p/q} (\sum_{j=k-l}^{\tilde{r}} \tilde{\sigma}_j)^{p-1}$.

Finally, we obtain

$$w_g^2(\mathcal{E}(\boldsymbol{L}_\diamond^*)) \le \mathbb{E}_{\boldsymbol{G}} \inf_{\boldsymbol{X} \in \mathcal{N}(\boldsymbol{L}_\diamond^*)} \|\boldsymbol{G} - \boldsymbol{X}\|_{\text{F}}^2.$$

Note that

$$\begin{aligned} \boldsymbol{G} - \boldsymbol{X} = P_{\mathcal{M}}(\boldsymbol{G}) - \frac{t}{\tilde{\varsigma}} \sum_{i \in I_2} (\sigma_i^*)^{p-1} \boldsymbol{u}_i^* \boldsymbol{v}_i^{*\top} \\ - t \sum_{i \in I_1} \boldsymbol{u}_i^* \boldsymbol{v}_i^{*\top} + P_{\mathcal{M}^\perp}(\boldsymbol{G}) - t \sum_{i \in I_0} h_i \boldsymbol{u}_i^* \boldsymbol{v}_i^{*\top}. \end{aligned}$$

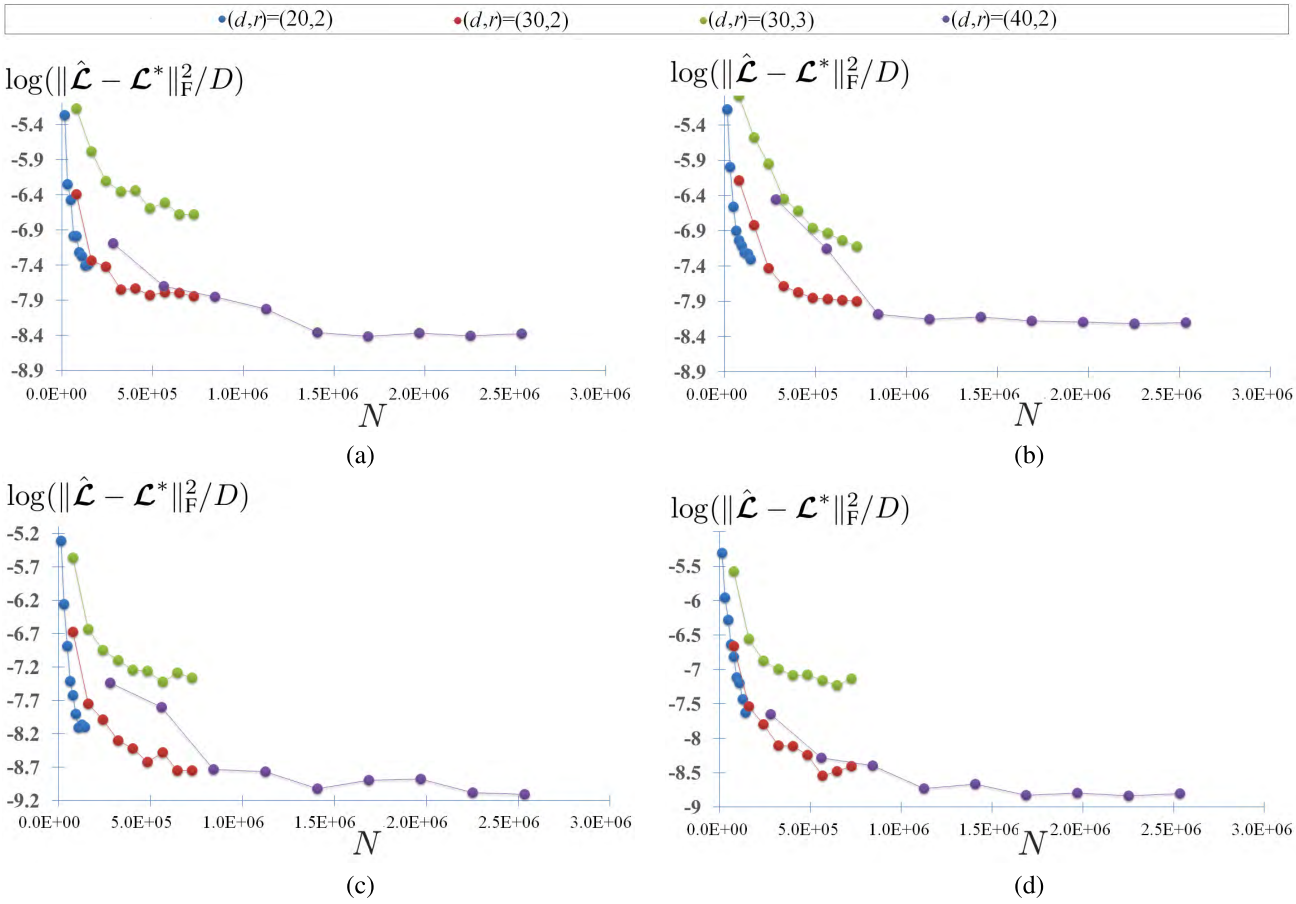

FIGURE 1. Plots of $\log(\|\hat{\mathcal{L}}-\mathcal{L}^*\|_F^2/D)$ versus the sample size N for 4-th order tensors of size $d \times d \times d \times d$ and Tucker rank (r, r, r, r) with different balanced spectral (k, p)-support norm regularizers. We consider tensors with 4 size and rank settings $(d, r) \in \{(20, 2), (30, 2), (30, 3), (40, 2)\}$. Subplots (a)-(d) show the results of different choices of (k, p) pairs: $(k, p) = (1, 2), (2, 1.5), (2, 2)$ and $(2, 4)$, respectively.

Let $P_{\mathcal{M}^\perp}(\boldsymbol{G}) = \sum_{i\in I_0} \sigma_i(P_{\mathcal{M}^\perp}(\boldsymbol{G}))\boldsymbol{u}_i^*\boldsymbol{v}_i^{*\top}$ be the decomposition of $P_{\mathcal{M}^\perp}(\boldsymbol{G})$ in the basis of $\{\boldsymbol{u}_i^*\boldsymbol{v}_i^{*\top}\}_{i\in I_0}$. Taking $t = \|P_{\mathcal{M}^\perp}(\boldsymbol{G})\|_2 = \max_{i\in I_0}\sigma_i(P_{\mathcal{M}^\perp}(\boldsymbol{G}))$ and $h_i = \sigma_i(P_{\mathcal{M}^\perp}(\boldsymbol{G}))/\|P_{\mathcal{M}^\perp}(\boldsymbol{G})\|_2 \le 1$, we have

$$w_g^2(\mathcal{E}(\boldsymbol{L}_\diamond^*)) \le \mathbb{E}_{\boldsymbol{G}}\|P_{\mathcal{M}}(\boldsymbol{G})\|_F^2 + \mathbb{E}_{\boldsymbol{G}}\|P_{\mathcal{M}}(\boldsymbol{G})\|_2^2\Big(\frac{1}{\varsigma^2}+r^{\lfloor K/2\rfloor}-k+l+1\Big).$$

Since

$$\mathbb{E}_{\boldsymbol{G}}\|P_{\mathcal{M}}(\boldsymbol{G})\|_F^2 \le r^{\lfloor K/2\rfloor}(2d^{\lfloor K/2\rfloor} - r^{\lfloor K/2\rfloor}),$$

and

$$\mathbb{E}_{\boldsymbol{G}}\|P_{\mathcal{M}}(\boldsymbol{G})\|_2^2 \le 2(2d^{\lfloor K/2\rfloor} - r^{\lfloor K/2\rfloor}),$$

the statement of Lemma 3 is proved. □

Equipped with Lemma 3, the main theorem is given as follows:

Theorem 2: If the sample size

$$N \ge c_0 w_g^2(\mathcal{E}(\boldsymbol{L}_\diamond^*))K\log d,$$

for some sufficiently large c_0 and B_{cn} is chosen by $B_{cn} = 2\varsigma\sqrt{N}$ then there exists parameter $\kappa_{c_0} \approx 1 - o(d^{-\lfloor K/4\rfloor})$, and constants c_1 and c_2 such that, the following inequality holds with probability at least $1 - \exp(-c_1 w_G^2) - 2\exp(-c_2 K w_G^2 \log d)$,

$$\frac{\|\hat{\mathcal{L}}-\mathcal{L}^*\|_F^2}{D} \le \max\big\{B_\varsigma, B_{w_g}\big\}, \tag{30}$$

with

$$B_{noise} = \frac{4\varsigma^2}{\kappa_{c_0}}, \quad \text{and } B_{w_g} = \frac{4c_0 K\alpha^2 w_g^2(\mathcal{E}(\boldsymbol{L}_\diamond^*))\log d}{N},$$

where $w_g(\mathcal{E}(\boldsymbol{L}_\diamond^*))$ is upper bounded in Eq. (28).

The proof of this theorem directly follows that of [42, Th. 1a] and is thus omitted.

Remark 1 (Sample Complexity): According to Theorem 2, the sample complexity is of order

$$O\Big(Kw_g^2(\mathcal{E}(\boldsymbol{L}_\diamond^*))\log d\Big).$$

Remark 2 (No Exact Recovery Guarantee): When the standard deviation of the noise $\varsigma = 0$, i.e., the observations are noiseless, then from the right hand side of Eq. (30), the estimation error will be dominated by the second term, which indicates that exact recovery can not be guaranteed. This can

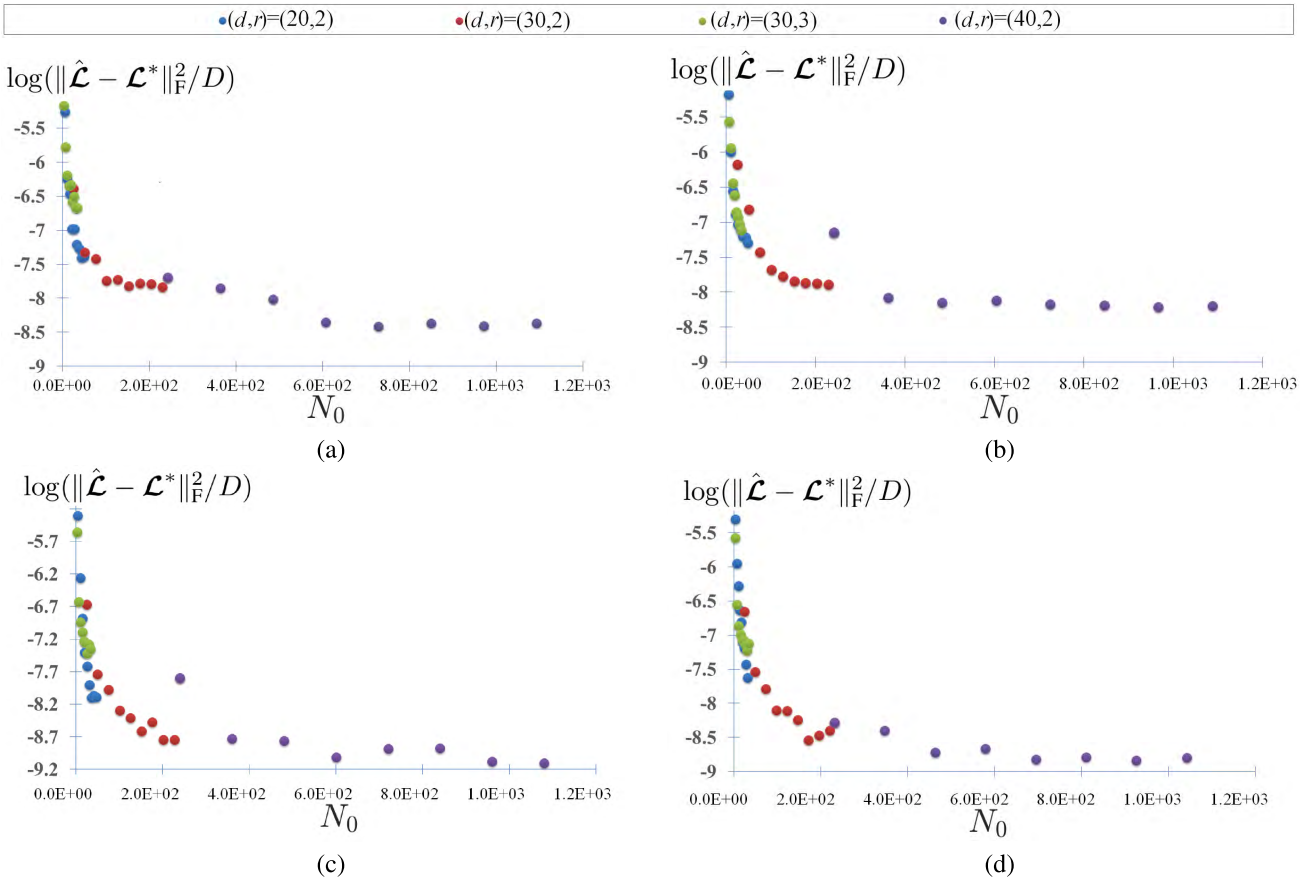

FIGURE 2. Plots of $\log(\|\hat{\mathcal{L}}-\mathcal{L}^*\|_F^2/D)$ versus the rescaled sample size $N_0 = N/(K\alpha^2 w_g^2(\mathcal{E}(L_\diamond^*))\log d)$ for 4-th order tensors of size $d \times d \times d \times d$ and Tucker rank (r,r,r,r) with different balanced spectral (k,p)-support norm regularizers. We consider tensors with 4 size and rank settings $(d,r) \in \{(20,2),(30,2),(30,3),(40,2)\}$. Subplots (a)-(d) show the results of different choices of (k,p) pairs: $(k,p) = (1,2),(2,1.5),(2,2)$ and $(2,4)$, respectively.

be seen as a trade-off that we don't assume strict incoherent conditions on $\mathcal{L}^*$ [12], [44].

Remark 3 (The Nuclear Norm Case): When $k = 1$ the results on the sample complexity and error bound degenerate to case of nuclear norm. By using Lemma 3, we get that

$$w_g^2(\mathcal{E}(\boldsymbol{L}_\diamond^*)) \le 4r^{\lfloor K/2\rfloor}\left(2d^{\lceil K/2\rceil} - r^{\lfloor K/2\rfloor}\right),$$

which indicates that the sample complexity is

$$O\left(Kr^{\lfloor K/2\rfloor}d^{\lceil K/2\rceil}\log d\right),$$

and the estimation error $\|\hat{\mathcal{L}}-\mathcal{L}^*\|_F^2/D$ is

$$O\left(\frac{Kr^{\lfloor K/2\rfloor}d^{\lceil K/2\rceil}\log d}{N}\right).$$

VI. EXPERIMENTS

In this section, we will first corroborate the correctness of Theorem 2 through an experiment on synthetic data. We will then evaluate the effectiveness of the proposed algorithm with experiment on synthetic data and video data. The algorithms are implemented in Matlab and the experiments are conducted on a PC with Intel i5 CPU and 12GB RAM.

A. CORROBORATION OF THEOREM 2

In this subsection, we will corroborate the correctness of Theorem 2. Specifically, we will show that the proposed upper bound can predict the scaling behavior through numerical simulations. First, the ground truth tensor $\mathcal{L}^* \in \mathbb{R}^{d_1\times\cdots\times d_K}$ with Tucker rank $(r_1, r_2, \cdots, r_K)$ is generated by

$$\mathcal{L}^* = \mathcal{C} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \times_3 \cdots \times_K \boldsymbol{U}_K, \tag{31}$$

where the core tensor $\mathcal{C} \in \mathbb{R}^{r_1\times\cdots\times r_K}$ and the factor matrices $\boldsymbol{U}_k \in \mathbb{R}^{r_k\times d_k}$ are all generated by element-wisely i.i.d. standard Gaussian. Then, given observed entries of $\mathcal{L}^*$, the additive noise $\{\varepsilon_i\}_{i=1}^N$ isc drown from i.i.d. standard Gaussian and the noise level are controlled by a parameter $c \ge 0$. The noise setting is specialized as follows:

$$\varepsilon_i \sim \mathcal{N}(0,1), \quad \varsigma = \frac{c\|\mathcal{L}^*\|_F}{\sqrt{D}}, \tag{32}$$

with c controlling the noise to signal ratio.

In this experiment, we consider 4-th order tensors of size (d,d,d,d) and Tucker rank (r,r,r,r) chosen from $(d,r) \in \{(20,2),(30,2),(30,3),(40,2)\}$. The balanced spectral (k,p)-support norm regularizers are chosen from $(k,p) \in \{(1,2),(2,1.5),(2,2),(2,4)\}$. The sampling ratio

TABLE 1. RSE values in noisy tensor completion result on synthetic datasets. The proposed balanced spectral (k,p)-support norm based method (with different choices of (k,p) pairs) is compared with the nuclear norm based methods SNN [19] and SquareNN [23]. ('SR' is short for 'sampling ratio', i.e., N/D.)

Tensor Size	Tucker Rank	SR	Noise Level	SNN	SquareNN	Proposed model with different (k,p)		
						RSE,(k,p)	RSE,(k,p)	RSE,(k,p)
$20\times20\times20\times20$	$(2,2,2,2)$	15%	0.05	0.213	0.195	**0.165**, $(2,2.05)$	0.177, $(2,2.00)$	0.197, $(2,4.00)$
$20\times20\times20\times20$	$(2,2,2,2)$	15%	0.15	0.364	0.273	**0.246**, $(2,2.05)$	0.259, $(2,4.00)$	0.261, $(2,2.00)$
$30\times30\times30\times30$	$(2,2,2,2)$	12.5%	0.05	0.077	0.118	**0.075**, $(2,1.10)$	0.081, $(2,1.50)$	0.085, $(2,2.05)$
$30\times30\times30\times30$	$(2,2,2,2)$	12.5%	0.15	0.213	0.175	**0.146**, $(2,2.05)$	0.151, $(2,2.10)$	0.152, $(2,2.40)$
$30\times30\times30\times30$	$(3,3,3,3)$	12.5%	0.05	0.39	0.188	**0.138**, $(3,2.00)$	0.140, $(4,2.00)$	0.144, $(3,2.05)$
$30\times30\times30\times30$	$(3,3,3,3)$	12.5%	0.15	0.457	0.285	**0.233**, $(4,2.05)$	0.235, $(3,2.10)$	0.237, $(5,2.00)$

N/D varies from 0.1 to 0.9 with step size 0.1. For fixed tensor size, Tucker rank, sampling ratio and specific spectral (k,p)-support norm, the experiment is repeated in 10 runs.

Note that according to Theorem 2, for a specific spectral (k,p)-support norm, when the noise is very small, the estimation error will be upper bounded by B_{w_g}, i.e.,

$$\frac{\|\hat{\mathcal{L}}-\mathcal{L}^*\|_{\mathrm{F}}^2}{D}\leq\frac{4c_0K\alpha^2w_g^2(\mathcal{E}(\boldsymbol{L}_\diamond^*))\log d}{N},$$

with high probability. Define the rescaled sample size

$$N_0:=\frac{N}{K\alpha^2w_g^2(\mathcal{E}(\boldsymbol{L}_\diamond^*))\log d}. \tag{33}$$

and let $c=0$ in Eq. (32). Thus it is expected that if we plot the estimation error against the rescaled sample size, then plots of different tensor sizes and ranks will align relatively well.

Fig. 1 shows the plots of the estimation error in log scale, i.e., $\log(\|\hat{\mathcal{L}}-\mathcal{L}^*\|_{\mathrm{F}}^2/D)$, versus the sample size for different tensor sizes, ranks and different spectral (k,p)-norm regularizers. It can be seen from Fig. 1 that, given a specific spectral (k,p)-norm regularizer, larger tensors with larger ranks require more observations to obtain a small estimation error which is consistent with our intuition. Fig. 2 shows the plots of the estimation error versus the rescaled sample size N_0 defined in Eq. (33). One can see from Fig. 2 that, given a certain spectral (k,p)-support norm regularizer, the estimation error of tensors (with different sizes and ranks) align relatively well with respect to N_0. This phenomenon is consistent with our expectation that the proposed upper bound can predict the scaling behavior of the estimation error. Thus, the correctness of Theorem 2 is corroborated.

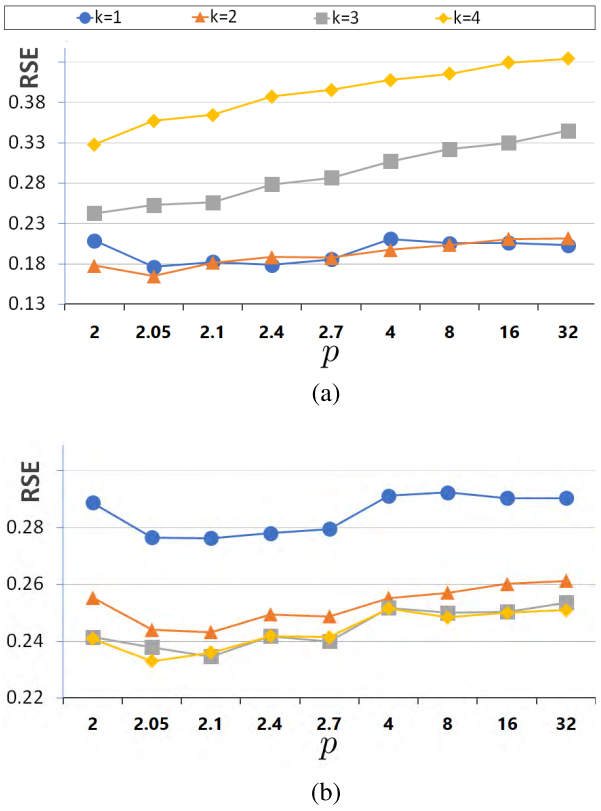

FIGURE 3. The impact of different choices of (k,p) on the estimation error in different settings: (a) $(d,r)=(20,2)$, noise level $c=0.05$ and sampling ratio 15%, (b) $(d,r)=(30,3)$, noise level $c=0.05$ and sampling ratio 12.5%.

B. EFFECTIVENESS OF THE PROPOSED ESTIMATOR

The motivation of this work is to exploit the flexibility of spectral (k,p)-support norm in fitting the decaying pattern of spectrum in a squared reshaping manner for tensor completion. To show the superiority of the proposed model over the nuclear norm based ones, i.e., the square nuclear norm model [23] ('SquareNN' for short) and the tensor nuclear norm based model [2] ('SNN' for short), experiments for noisy tensor completion on synthetic datasets and color videos are carried out.

Since the SquareNN model is a special case of our balanced spectral (k,p)-support norm based model with $k=1$, we simply use the result of the proposed model by setting $k=1$. Since the observed entries are noisy, we consider the the SNN model formulated as follows:

$$\min_{\mathcal{L}\in\mathbb{R}^{d_1\times\cdots\times d_K}}\ \sum_{k=1}^{K}\omega_k\|\boldsymbol{L}_{(k)}\|_*,$$
$$\text{s.t. }\|\boldsymbol{y}-\mathfrak{X}(\mathcal{L})\|_2\leq B_{cn}. \tag{34}$$

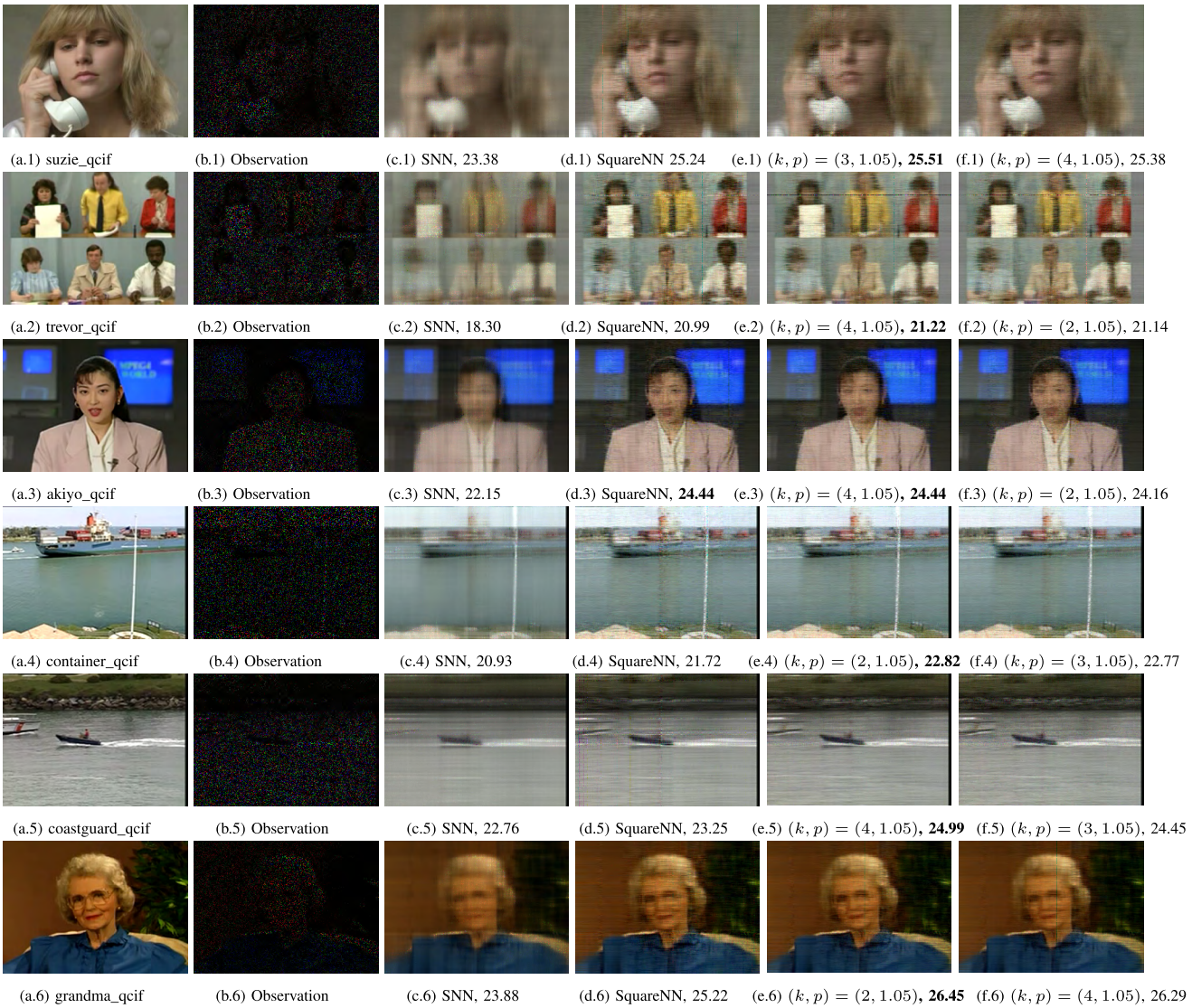

FIGURE 4. Qualitative comparison in noisy video inpainting. (a) is the ground truth frame, (b) is the observed frame with sampling ratio $N/D = 8\%$ and noise level $c = 0.05$. (c)-(f) show the inpainting results of SNN, SquareNN and two examples of the proposed balanced spectral (k, p)-support norm based model with different (k, p) choices. The PSNR values are put behind each method.

Problem (34) is solved through the framework of ADMM [39] and coded in Matlab, for more details, see [13].

1) SYNTHETIC SIMULATION

In this subsection, random tensors of different sizes and Tucker ranks are generated with different noise levels to compare performances of the proposed balanced spectral (k, p)-support norm based model with the SquareNN and SNN model. The ground truth tensor are generated through Eq. (31) with additional Gaussian noise satisfying Eq. (32). For the SNN model, the elements of the weight vector $\boldsymbol{\omega}$ are set to 0.25. For the SquareNN model and the proposed model, the permutations (l_1, l_2, l_3, l_4) are all $[1, 2, 3, 4]$ with balancing mode $j = 2$ in Eq. (6). The parameter B_{cn} is simply set to $2\varsigma\sqrt{N}$ for the three models . The remaining parameters are manually tuned for better performances in most cases. For an estimation $\hat{\mathcal{L}}$ of the underlying tensor $\mathcal{L}^*$, the relative root squared error (RSE), i.e.,

$$\text{RSE} = \frac{\|\hat{\mathcal{L}} - \mathcal{L}^*\|_{\text{F}}}{\|\mathcal{L}^*\|_{\text{F}}},$$

is employed to evaluate the quality of $\hat{\mathcal{L}}$.

In Table 1, the results of noisy tensor completion are reported. According to Table 1, we can see that the proposed method outperforms SquareNN and SNN thanks to the flexibility of fitting the spectrum of spectral (k, p)-support norm. To illustrate the impact of different choices of (k, p) on the estimation error, the RSE values obtained by the proposed method with different (k, p) pairs are shown in Fig. 3. It can be seen from Fig. 3 that the proposed balanced spectral

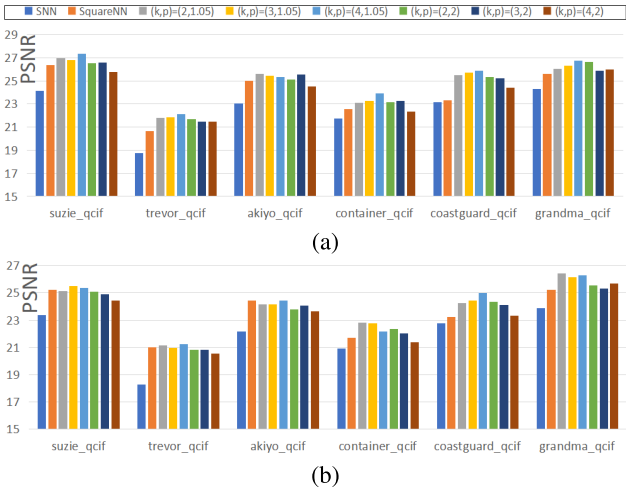

FIGURE 5. The quantitative comparison in PSNR values for noisy video in-painting with sampling ratio 8% and noise level $c \in \{0, 0.05\}$. The proposed balanced spectral (k, p)-support norm based method (with different choices of (k, p) pairs) is compared with the nuclear norm based methods SNN and SquareNN. (a) $c = 0$. (b) $c = 0.05$.

(k, p)-support norm based model performs better when p is around 2.

2) VIDEO IN-PAINTING

In video in-painting, the goal is to reconstruct a video from incomplete and noisy observations. Since color videos can be represented as a 4-order tensor (height $\times$ width $\times$ channels $\times$ frames), we fulfill the video in-painting task using tensor completion. In this subsection, the proposed algorithm is compared with the SNN and the SquareNN on video data. The color videos include suzie_qcif, trevor_qcif, akiyo_qcif, container_qcif, coastguard_qcif and grandma_qcif[1]. For the limitation of computation, we use the first 32 frames of each video and get 4-th order tensors of size height $\times$ width $\times 3 \times$ 32. The sampling ratio N/D is set to 8% and the noise setting follows Eq. (32) with noise level parameter $c \in \{0, 0.05\}$. For the SNN model, the elements of the weight vector $\boldsymbol{\omega}$ are set to 0.25. For the SquareNN and the proposed method, the permutations (l_1, l_2, l_3, l_4) are all $[1, 4, 2, 3]$ with balancing mode $j = 2$ in Eq. (6). The parameter B_{cn} is simply set to $2\varsigma\sqrt{N}$. The remaining parameters are manually tuned for better performances. For an estimation $\hat{\mathcal{L}}$ of the underlying tensor $\mathcal{L}^*$, the peak signal-to-noise ratio (PSNR), i.e.,

$$\text{PSNR} = 10\log_{10}\left(\frac{n_1 n_2 n_3 \|\mathcal{L}^*\|_\infty^2}{\|\hat{\mathcal{L}} - \mathcal{L}^*\|_{\text{F}}^2}\right),$$

is employed to evaluate the quality of $\hat{\mathcal{L}}$.

The qualitative comparison results with observation ratio 8% of six videos are shown in Fig. 4. For quantitative comparison, we report the PSNR values on the six test videos with noise level $c \in \{0, 0.05\}$ in Fig. 5. It can be seen that the SquareNN performs better than the SNN. Besides, the recovery results by the proposed method outperform those by the SNN and the SquareNN in terms of PSNR values and visual quality[2].

[1]The videos are available from https://sites.google.com/site/subudhibadri/fewhelpfuldownloads.

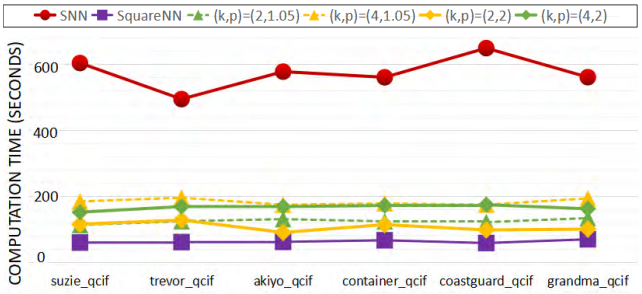

FIGURE 6. Comparison of computation time in seconds for video in-painting with sampling ratio 8% and noise level $c = 0$. The proposed method (with different choices of (k, p) pairs) is compared with SNN. Note that SquareNN is a special case of our balanced spectral (k, p)-support norm based method with $k = 1$.

Comparison of computation time is also presented in Fig. 6 for the case where $c = 0$. As shown in Fig. 6, the SquareNN is the most efficient, models based on the spectral (k, p)-support norm are the second efficient, and the SNN is the most time-consuming. Note that the SquareNN is a special case of the proposed model and it is solved through the proposed Algorithm 1. The efficiency of SquareNN and other spectral (k, p)-support norm based models benefits from the APG solver in Algorithm 1 which only computes first k leading singular values and singular vectors in each iteration. Since the SNN is implemented in the ADMM framework which involves full SVD on the unfolding matrices along all modes in each iteration, it is computationally expensive.

VII. CONCLUSION

To estimate a K-th order tensor from its partial noisy observations, we defined an estimator based on a newly defined 'balanced spectral (k, p)-support norm'. A scalable algorithm was proposed to efficiently compute the estimator. Instead of directly solving the primal problem which involves full SVD in each iteration, the proposed algorithm benefits from the Lagrangian dual through minimizing the dual norm of the (k, p)-support norm which only computes the first-k leading singular values and singular vectors in each iteration. For statistical performance of the proposed estimator, upper bounds on the sample complexity and the estimation error were established and verified through simulation study. The proposed method is shown to outperform the nuclear norm based methods through experimental results on synthetic and real datasets. It is interesting to consider fast algorithms like [45] and [46], other atomic norm regularization [47], [48] or other tensor decomposition approaches [49] for low-rank tensor recovery in future research.

[2]As pointed out by one kind reviewer of this manuscript, in the $c = 0.05$ case, the improvement of the proposed model on the SquareNN is not significant on suzie_qcif, trevor_qcif and akiyo_qcif. The reason is still under study.

ACKNOWLEDGMENT

The authors would be thankful to the reviewers for their comments aimed at improving the manuscript. The authors also thank Mr. Jian Lou, one author of [37], for his generosity in sharing their code. The authors also thank Miss. Fangfang Yang, Prof. Zhong Jin, Mr. Huan Wang and Mr. Jing Lou for their long-time support.

REFERENCES

[1] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[2] 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. 2013.

[3] M. Tang, G. Ding, Q. Wu, Z. Xue, and T. A. Tsiftsis, "A joint tensor completion and prediction scheme for multi-dimensional spectrum map construction," *IEEE Access*, vol. 4, pp. 8044–8052, 2016.

[4] A. Cichocki *et al.*, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Mar. 2015.

[5] A. Wang and Z. Jin, "Near-optimal noisy low-tubal-rank tensor completion via singular tube thresholding," in *Proc. IEEE Int. Conf. Data Mining Workshop (ICDMW)*, 2017, pp. 553–560.

[6] Y. Liu and F. Shang, "An efficient matrix factorization method for tensor completion," *IEEE Signal Process. Lett.*, vol. 20, no. 4, pp. 307–310, Apr. 2013.

[7] G. Zhong, S. Xiang, P. Zhou, and L. Yu, "Spatially adaptive tensor total variation-Tikhonov model for depth image super resolution," *IEEE Access*, vol. 5, pp. 13857–13867, 2017.

[8] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "A survey of multilinear subspace learning for tensor data," *Pattern Recognit.*, vol. 44, no. 7, pp. 1540–1551, 2011.

[9] Z. Lai, Y. Xu, J. Yang, J. Tang, and D. Zhang, "Sparse tensor discriminant analysis," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3904–3915, Oct. 2013.

[10] A. Cichocki, N. Lee, I. V. Oseledets, A.-H. Phan, Q. Zhao, and D. Mandic. (2016). "Low-rank tensor networks for dimensionality reduction and large-scale optimization problems: Perspectives and challenges PART 1." [Online]. Available: https://arxiv.org/abs/1609.00893

[11] E. E. Papalexakis, C. Faloutsos, and N. D. Sidiropoulos, "Tensors for data mining and data fusion: Models, applications, and scalable algorithms," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 2, 2016, Art. no. 16.

[12] 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.

[13] R. Tomioka, K. Hayashi, and H. Kashima. (2010). "Estimation of low-rank tensors via convex optimization." [Online]. Available: https://arxiv.org/abs/1010.0789

[14] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multimodal factor analysis," in *Proc. UCLA Working Papers Phonetics*, 1970, pp. 1–84.

[15] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.

[16] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, "Novel methods for multilinear data completion and de-noising based on tensor-SVD," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3842–3849.

[17] L. Wang, K. Xie, T. Semong, and H. Zhou, "Missing data recovery based on tensor-CUR decomposition," *IEEE Access*, vol. 6, pp. 532–544, 2017.

[18] C. J. Hillar and L.-H. Lim, "Most tensor problems are NP-hard," *J. ACM*, vol. 60, no. 6, p. 45, 2009.

[19] R. Tomioka, T. Suzuki, K. Hayashi, and H. Kashima, "Statistical performance of convex tensor decomposition," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 972–980.

[20] Q. Song, H. Ge, J. Caverlee, and X. Hu. (2017). "Tensor completion algorithms in big data analytics." [Online]. Available: https://arxiv.org/abs/1711.10105

[21] R. Tomioka and T. Suzuki, "Convex tensor decomposition via structured schatten norm regularization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 1331–1339.

[22] B. Romera-Paredes and M. Pontil, "A new convex relaxation for tensor completion," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 2967–2975.

[23] C. Mu, B. Huang, J. Wright, and D. Goldfarb, "Square deal: Lower bounds and improved relaxations for tensor recovery," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 73–81.

[24] Z. Zhang and S. Aeron, "Exact tensor completion using t-SVD," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1511–1526, Mar. 2017.

[25] J. A. Bengua, H. N. Phiem, 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.

[26] Y. Xu, R. Hao, W. Yin, and Z. Su, "Parallel matrix factorization for low-rank tensor completion," *Inverse Problems Imag.*, vol. 9, no. 2, pp. 601–624, 2015.

[27] Q. Zhao, L. Zhang, and A. Cichocki, "Bayesian CP factorization of incomplete tensors with automatic rank determination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1751–1763, Sep. 2015.

[28] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.

[29] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ., Stanford, CA, USA, 2002.

[30] Y. Chen, S. Bhojanapalli, S. Sanghavi, and R. Ward, "Completing any low-rank matrix, provably," *J. Mach. Learn. Res.*, vol. 16, pp. 2999–3034, Dec. 2015.

[31] M. A. Davenport and J. Romberg, "An overview of low-rank matrix recovery from incomplete observations," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 4, pp. 608–622, Jun. 2016.

[32] D. Goldfarb and Z. Qin, "Robust low-rank tensor recovery: Models and algorithms," *SIAM J. Matrix Anal. Appl.*, vol. 35, no. 1, pp. 225–253, 2014.

[33] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos. (2016). "Tensor decomposition for signal processing and machine learning." [Online]. Available: https://arxiv.org/abs/1607.01668

[34] M. Signoretto, Q. T. Dinh, L. De Lathauwer, and J. A. K. Suykens, "Learning with tensors: A framework based on convex optimization and spectral regularization," *Mach. Learn.*, vol. 94, no. 3, pp. 303–351, 2014.

[35] A. M. McDonald, M. Pontil, and D. Stamos, "Fitting spectral decay with the k-support norm," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2016, pp. 1–11.

[36] R. T. Rockerfeller, *Convex Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1970, p. 12.

[37] Y.-M. Cheung and J. Lou, "Scalable spectral k-support norm regularization for robust low rank subspace learning," in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage.*, 2016, pp. 1151–1160.

[38] O. Klopp, "Noisy low-rank matrix completion with general sampling distribution," *Bernoulli*, vol. 20, no. 1, pp. 282–303, 2014.

[39] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[40] C. Mu, Y. Zhang, J. Wright, and D. Goldfarb, "Scalable robust matrix recovery: Frank–Wolfe meets proximal methods," *SIAM J. Sci. Comput.*, vol. 38, no. 5, pp. A3291–A3317, 2016.

[41] A. Yurtsever, Q. T. Dinh, and V. Cevher, "A universal primal-dual convex optimization framework," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 3150–3158.

[42] S. Gunasekar, A. Banerjee, and J. Ghosh, "Unified view of matrix completion under general structural constraints," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 1180–1188.

[43] G. A. Watson, "Characterization of the subdifferential of some matrix norms," *Linear Algebra Appl.*, vol. 170, no. 6, pp. 33–45, Jun. 1992.

[44] B. Huang, C. Mu, D. Goldfarb, and J. Wright, "Provable models for robust low-rank tensor completion," *Pacific J. Optim.*, vol. 11, no. 2, pp. 339–364, 2015.

[45] X.-Y. Liu, S. Aeron, V. Aggarwal, and X. Wang, "Low-tubal-rank tensor completion using alternating minimization," *Proc. SPIE*, vol. 9848, p. 984809, May 2016.

[46] P. Zhou, C. Lu, Z. Lin, and C. Zhang, "Tensor factorization for low-rank tensor completion," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1152–1163, Mar. 2018.

[47] E. Richard, G. R. Obozinski, and J.-P. Vert, "Tight convex relaxations for sparse matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3284–3292.

[48] E. Belilovsky, A. Argyriou, G. Varoquaux, and M. Blaschko, "Convex relaxations of penalties for sparse correlated variables with bounded total variation," *Mach. Learn.*, vol. 100, nos. 2–3, pp. 533–553, 2015.

[49] Q. Zhao, G. Zhou, S. Xie, L. Zhang, and A. Cichocki. (2016). "Tensor ring decomposition." [Online]. Available: https://arxiv.org/abs/1606.05535

DONGXU WEI received the master's degree from the College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2007. He is currently pursuing the Ph.D. degree with the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing. He is currently with the School of Physics and Electronic Electrical Engineering, Huaiyin Normal University, and Jiangsu Province Key Construction Laboratory of Modern Measurement Technology and Intelligent System. His current research interests include tensor decomposition, radar fuze, and robot for intelligent welding.

ANDONG WANG received the B.Sc. degree in software engineering from the Nanjing University of Science and Technology, Nanjing, China, in 2012, where he is currently pursuing the Ph.D. degree in pattern recognition theory and application with the School of Computer Science and Engineering. His research interests include low-rank tensor analysis and autonomous driving.

BO WANG received the master's degree from the School of Materials Science and Engineering, Nanjing University of Science and Technology, Nanjing, China, in 2004. He is currently a Senior Manager with the Jiangsu Shuoshi Welding Technology Company, Ltd. His current research interests include digital manufacturing shop and robot for intelligent welding.

XIAOQIN FENG received the master's degree from the Nanjing University of Aeronautics and Astronautics, China, in 2011. She is currently with the School of Physics and Electronic Electrical Engineering, Huaiyin Normal University. Her current research interests include structural and physical properties of nanomaterials.