Noname manuscript No.
(will be inserted by the editor)

Reshaped Tensor Nuclear Norms for Higher Order Tensor Completion

Kishan Wimalawarne · Hiroshi Mamitsuka

Received: date / Accepted: date

Abstract We investigate optimal conditions for inducing low-rankness of higher order tensors by using convex tensor norms with reshaped tensors. We propose the reshaped tensor nuclear norm as a generalized approach to reshape tensors to be regularized by using the tensor nuclear norm. Furthermore, we propose the reshaped latent tensor nuclear norm to combine multiple reshaped tensors using the tensor nuclear norm. We analyze the generalization bounds for tensor completion models regularized by the proposed norms and show that the novel reshaping norms lead to lower Rademacher complexities. Through simulation and real-data experiments, we show that our proposed methods are favorably compared to existing tensor norms consolidating our theoretical claims.

Keywords Tensor nuclear norm, Reshaping, CP rank, Generalization bounds

Kishan Wimalawarne
Department of Mathematical Informatics
The University of Tokyo
Hongo 7-3-1, Bunkyo-ku,
Tokyo 113-8656, Japan
E-mail: kishanwn@gmail.com

Hiroshi Mamitsuka
Bioinformatics Center,
Institute for Chemical Research,
Kyoto University,
Gokasho, Uji, Japan.
Department of Computer Science,
Aalto University,
Espoo, Finland.
E-mail: mami@kuicr.kyoto-u.ac.jp

1 Introduction

Tensor formatted data is becoming abundant in machine learning applications. Among the many tensor related machine learning problems, tensor completion has gained an increased popularity in recent years. Tensor completion performs imputation of unknown elements of a partially observed tensor by exploiting its low-rank structure. Some of the popular real-world applications of tensor completion are found in recommendation systems [11,26], computer vision [15] and multi-relational link prediction [18]. Though there exist many methods to perform tensor completion [21], global optimal solutions are obtained mainly by convex low-rank tensor norms, making them an active area of research.

Over the years, many researchers have proposed different low-rank inducing norms to minimize the rank of tensors, however, none of these norms are universally better compared to others. The main challenge in designing norms for tensors is that they have multiple dimensions and different definitions of ranks (Tucker rank, CP rank, TT-rank), making it difficult for a single norm to induce low-rankness with respect to all the properties of tensors. Most tensor norms have been designed with a focus to a specific rank; overlapped trace norm [22] and latent trace norms [23] to constrain the multilinear ranks, tensor nuclear norm [25,24,14] to constrain the CP rank, and the Schatten TT rank [10] to constrain the TT-rank. However, targeting a specific rank to constrain may not always be practical, since we may not know the most suitable rank for a tensor in advance.

Most tensor norms reshape tensors by rearranging its elements as matrices to induce low-rankness with respect to a mode or a set of modes. However, this reshaping method is specific to obtaining relevant ranks that a norm constrains. An alternative view was presented by [16] with the square norm, where the tensor is reshaped as a balanced matrix without considering the structure of its ranks. The square norm has shown to have better sample complexities for higher order tensors (tensor with more than three modes) than some of the existing norms such as the overlapped trace norm [25]. However, this norm only considers the special case of reshaping a tensor as a matrix such that its dimensions are close to each other. Other possibilities of how reshaping tensors beyond matrices affect the inducement of low-rankness have not been investigated.

In this paper, we propose generalized reshaping strategies to reshape tensors and develop low-rank inducing tensor norms. We demonstrate that reshaping higher order tensors as another tensor and applying the tensor nuclear norm leads to better inducement of low-rankness compared to applying existing low-rank norms on the original tensor or its matrix unfoldings. Furthermore, we propose the latent reshaped tensor nuclear norm that combines multiple reshaped tensors to obtain a better performance among possible reshaping tensors. Using the generalization bounds, we show that the proposed norms are able to give lower Rademacher complexities compared to exiting norms. Using simulations and real-world data experiments we justify our theoretical analysis and show that our proposed methods are able to give better performance for tensor completion compared other convex norms.

Throughout this paper we use the following notations. We represent a K-mode (K-way) tensor as $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$. The mode-$k$ unfolding [12] of a tensor $\mathcal{T}$ is given by $T_{(k)} \in \mathbb{R}^{n_k \times \prod_{j \neq k} n_j}$, which is obtained by concatenating all slices along the mode-k. We indicate the tensor product [6] between vectors $u_i \in \mathbb{R}^{n_i},\ i = 1, \ldots, K$ using the notation $\otimes$ as $(u_1 \otimes \cdots \otimes u_K)_{i_1,\ldots,i_K} = \prod_{l=1}^{K} u_{l,i_l}$. The k-mode product of a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_k \cdots \times n_K}$ and a vector $v \in \mathbb{R}^{n_k}$ is defined as $\mathcal{T} \times_k v = \sum_{i_k=1}^{n_k} \mathcal{T}_{i_1,i_2,\ldots,i_k,\ldots,i_K} v_{i_k}$. The largest singular value of $\mathcal{T}$ is given by $\gamma_1(\mathcal{T})$. The rank of a matrix $A \in \mathbb{R}^{n \times m}$ is given by $\mathrm{Rank}(A)$.

2 Review of Low-Rank Tensor Norms

Designing of convex low-rank inducing norms for tensors is a challenging task. Over the years, several tensor norms have been proposed with each norm having certain advantages over the others. The main challenge with defining tensor norms is the multi-dimensionality of tensors and the existence of different ranks (e.g. CP rank, multilinear (Tucker) rank). A common criterion for designing low-rank tensor norms is to induce low-rankness by minimizing a particular rank. A commonly used rank is the multilinear rank, which represents the rank with respect to each mode of a tensor. Given a tensor $\mathcal{W} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$, we obtain the rank of each unfolding $r_k = \mathrm{Rank}(W_{(k)}),\ k = 1, \ldots, K$, and define the multilinear rank as $(r_1, \ldots, r_K)$. To minimize the multilinear rank the overlapped trace norm has been defined [15,22], which for a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$, as

$$\|\mathcal{T}\|_{\mathrm{overlap}} = \sum_{k=1}^{K} \|T_{(k)}\|_{\mathrm{tr}},$$

where $\|\cdot\|_{\mathrm{tr}}$ is the matrix nuclear norm (a.k.a. trace norm) [4], which is the sum of the non-zero singular values of a matrix. A limitation with this norm is that for tensors with high variations in the multilinear rank this norm stays at poor performances [22,23].

The latent trace norm [22] has been proposed to overcome limitations of the overlapped trace norm, which allows freedom to learn ranks with respect to each mode unfolding by considering a latent decomposition of the tensor. More specifically, the latent tensor norm learns latent tensors $\mathcal{T}^{(k)},\ k = 1, \ldots, K$ as

$$\|\mathcal{T}\|_{\mathrm{latent}} = \inf_{\mathcal{T}^{(1)} + \ldots + \mathcal{T}^{(K)} = \mathcal{T}} \sum_{k=1}^{K} \|T_{(k)}^{(k)}\|_{\mathrm{tr}}.$$

This norm was shown to be more robust for tensors with high variations in the multilinear rank compared to the overlapped trace norm [22]. The latent trace norm has been further extended to develop the scaled latent trace norm

[23] by considering the relative rank of each latent tensor by scaling using the inverse squared mode dimension.

Another popular rank for tensors is the CANDECOMP/PARAFAC (CP) rank [2,7,9,12], which can be considered as the higher order extension of the matrix rank. Recently, minimization of the CP rank has gained attention of many researchers, who have shown that it leads to a better sample complexity than multilinear rank based norms [25]. The tensor nuclear norm [25,24,14] has been defined as an approximation to minimize the CP rank of a tensor. For a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ with rank R, $\mathrm{Rank}(\mathcal{T}) = R$, the tensor nuclear norm is defined as

$$\|\mathcal{T}\|_* = \inf \Bigg\{ \sum_{j=1}^{R} \gamma_j | \mathcal{T} = \sum_{j=1}^{R} \gamma_j u_{1j} \otimes u_{2j} \cdots \otimes u_{Kj}, \\ \|u_{kj}\|_2^2 = 1, \gamma_j \geq \gamma_{j+1} > 0 \Bigg\}, \quad (1)$$

where $u_{kj} \in \mathbb{R}^{n_k}$ for $k = 1, \ldots, K$ and $j = 1, \ldots, R$.

The latest addition to convex low-rank tensor norms is the Schatten TT norm [10], which minimizes the tensor train rank [17] of tensors. The Schatten TT norm is defined as

$$\|\mathcal{T}\|_{s,T} = \frac{1}{K-1} \sum_{k=1}^{K-1} \|Q_k(\mathcal{T})\|_{\mathrm{tr}},$$

where $Q_k : \mathcal{T} \rightarrow \mathbb{R}^{n_{\geq k} \times n_{k<}}$ is an operator that reshapes the tensor $\mathcal{T}$ to a matrix by combining the first k modes as rows and the rest of the $K-k$ modes as columns. This norm has been shown to be suitable for high-order tensors.

It has also been shown that low-rank tensor norms can be designed without restricting to a specific rank. The square norm [16] reshapes a tensor as a matrix and apply the matrix nuclear norm as

$$\|\mathcal{T}\|_{\square} = \left\| \mathrm{reshape}\left(\mathcal{T}_{(1)}, \prod_{i=1}^{j}, \prod_{i=j+1}^{K} \right) \right\|_{\mathrm{tr}},$$

where the function reshape() reshapes $\mathcal{T}$ to a matrix with approximately equal dimensions for some $j > 1$. This norm has shown to have a better sample complexity for tensor completion compared to the overlapped trace norm.

We point out that all the existing tensor norms except the tensor nuclear norm reshape tensors as matrices to induce the low-rankness with respect to two sets of mode arrangements. As a result these norms are focused on constraining the multilinear rank of a tensor. However, tensor nuclear norm has shown to lead to a better sample complexity compared to multilinear rank based tensors norms for tensor completion [25]. Hence, lack of tensor nuclear norm regularization for reshaped tensors among existing norms may results in sub-optimal solutions.

3 Proposed Method: Tensor Reshaping and Tensor Nuclear Norm

In this paper, we investigate on extending the tensor nuclear norm for high order tensors. We explore methods to combine tensor reshaping with the tensor nuclear norm.

3.1 Generalized Tensor Reshaping

First, we introduce the following notation to compute the product of tensor dimensions. For a given vector $(n_1, \ldots, n_p)$, we present its element-wise product by $\mathsf{prod}(n_1, \ldots, n_p) = n_1 n_2 \cdots n_p$. Next, we define generalized reshaping for tensors.

Definition 1 (Tensor Reshaping) Let us consider a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ and its mode dimensions as $D = \{n_1, n_2, \ldots, n_K\}$. Given M sets $D_i \subset D,\ i = 1, \ldots, M$, that are disjoint, $D_i \cap D_j = \emptyset$ for $i \neq j$, the reshaping operator is defined as

$$\Pi_{(D_1,\ldots,D_M)} : \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K} \to \mathbb{R}^{\mathsf{prod}(D_1) \times \cdots \times \mathsf{prod}(D_M)},$$

and the inverse operator is represented by $\Pi^{\top}_{(D_1,\ldots,D_M)}$. Further, we present the reshaping of $\mathcal{X}$ by the set $(D_1, \ldots, D_M)$ as $\mathcal{X}_{(D_1,\ldots,D_M)}$.

We point out that when $|D_1| = \cdots = |D_M| = 1$, there is no reshaping of the tensor, $\mathcal{X}_{(D_1,\ldots,D_M)} = \mathcal{X}$. Unfolding of a tensor along the mode k [12] is equivalent to defining two sets with $D_1 = n_k$ and $D_2 = (n_1, \ldots, n_{k-1}, n_{k+1}, \ldots, n_K)$. Further, we can obtain reshaping of a tensor as a matrix for the square norm [16] by specifying two sets D_1 and D_2 with $\mathsf{prod}(D_1) \approx \mathsf{prod}(D_2)$.

3.2 Reshaped Tensor Nuclear Norm

We propose a class of tensor norms by combining generalized tensor reshaping and the tensor nuclear norm. We name the proposed norms *Reshaped Tensor Nuclear Norms*. In order to define the proposed norms, we consider a K-mode tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ and a set of $D_i,\ i = 1, \ldots, M$, adhering to Definition 1. We define the reshaped tensor nuclear norm as

$$\|\mathcal{X}_{(D_1,\ldots,D_M)}\|_*,$$

where $\|\cdot\|_*$ is the tensor nuclear norm as defined in (1). It is understood that this norm is a convex norm, since the tensor nuclear norm (1) is convex.

3.3 Reshaped Latent Nuclear Norm

A practical limitation in applying reshaping the proposed tensor norm is the difficulty to select the most suitable reshaping set out of all possible reshaping combinations. This is critical since we would not know the ranks of the tensor prior to training a learning model. To overcome this difficulty we propose the *Reshaped Latent Tensor Nuclear Norm* by extending the latent trace norm [22] for reshaping tensors.

Let us consider a collection of G reshaping sets $D_{\mathrm{L}} = (D^{(1)}, \ldots, D^{(G)})$ where each $D^{(s)} = (D_1^{(s)}, \ldots, D_{m_s}^{(s)}),\ s = 1, \ldots, G$ consists a reshaping set for a m_s-mode reshaped tensor. Further, we consider the $\mathcal{W}$ as a summation of G latent tensors $\mathcal{W}^{(g)},\ g = 1, \ldots, G$ as $\mathcal{W} = \sum_{k=1}^{G} \mathcal{W}^{(k)}$. We define the reshaped latent tensor nuclear norm as

$$\|\mathcal{W}\|_{\mathrm{r\_latent}(D_{\mathrm{L}})} = \inf_{\mathcal{W}^{(1)}+\cdots+\mathcal{W}^{(G)}=\mathcal{W}} \sum_{k=1}^{G} \|\mathcal{W}^{(k)}_{(D_1^{(k)},\ldots,D_{m_k}^{(k)})}\|_*. \tag{2}$$

We point out that the above norm differs from the latent trace norm [22] since it considers reshaping sets defined by the user where the latent trace norm considers all the mode-wise tensor unfolding. Furthermore, the above norm uses the tensor nuclear norm while the latent trace norm is build using the matrix nuclear norm.

3.4 Completion Models

Now, we propose tensor completion models for the proposed norms. Let us consider a partially observed tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$. Given that $\mathcal{X}$ has m observed elements, we define the mapping of the observed elements from $\mathcal{X}$ by $\Omega : \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K} \to \mathbb{R}^m$. Given a reshaping set $(D_1, \ldots, D_M)$, the completion model that is regularized by the reshaped norm is given as

$$\begin{aligned} &\min_{\mathcal{W}} \frac{1}{2}\|\Omega(\mathcal{X}) - \Omega(\mathcal{W})\|_{\mathrm{F}}^2 \\ &\text{s.t.}\ \ \|\mathcal{W}_{(D_1,\ldots,D_M)}\|_* \le \lambda, \end{aligned} \tag{3}$$

where λ is a regularization parameter. For a selected set of reshaping sets $D_{\mathrm{L}} = (D^{(1)}, \ldots, D^{(G)})$, a completion model regularized by the reshaped latent tensor nuclear norm is given as

$$\begin{aligned} \min_{\mathcal{W}^{(1)}+\cdots+\mathcal{W}^{(G)}=\mathcal{W}} &\frac{1}{2}\|\Omega(\mathcal{X}) - \Omega(\mathcal{W}^{(1)} + \cdots + \mathcal{W}^{(G)})\|_{\mathrm{F}}^2 \\ &\text{s.t.}\ \ \|\mathcal{W}\|_{\mathrm{r\_latent}(D_{\mathrm{L}})} \le \lambda, \end{aligned} \tag{4}$$

where λ is a regularization parameter.

4 Theory

We investigate theoretical properties of our proposed methods to identify the optimal conditions for reshaping of tensors. For our analysis, we use generalization bounds based on transductive Rademacher complexity analysis [3,20].

We consider the learning problem in (3) and we denote the indexes of the observed elements of $\mathcal{X}$ by S, where each index $(i_1, \cdots, i_K)$ of observed elements of $\mathcal{X}$ is assigned as an element $\alpha_j \in \mathrm{S}$ for some $1 \leq j \leq |\mathrm{S}|$. We consider observed elements as the training set denoted by $\mathrm{S}_{\mathrm{Train}}$ and the rest belonging to the test set denoted by $\mathrm{S}_{\mathrm{Test}}$. For the convenience of deriving the Rademacher complexity, we consider the special case of $|\mathrm{S}_{\mathrm{Ttrain}}| = |\mathrm{S}_{\mathrm{Test}}| = |\mathrm{S}|/2$ as in [20].

Given a reshaping set $(D_1, \ldots, D_M)$, we consider the hypothesis class $\mathsf{W} = \{\mathcal{W} | \|\mathcal{W}_{(D_1,\ldots,D_M)}\|_* \leq t\}$ for a given t. Given a loss function $l(\cdot,\cdot)$ and a set S, we define the empirical loss as

$$L_{\mathrm{S}}(l \circ \mathcal{W}) := \frac{1}{|\mathrm{S}|}\left[\sum_{(i_1,\ldots,i_K)\in \mathrm{S}} l(\mathcal{X}_{i_1,\ldots,i_K}, \mathcal{W}_{i_1,\ldots,i_K})\right].$$

Given that $\max_{i_1,\ldots,i_K \; \mathcal{W}\in\mathsf{W}} |l(\mathcal{X}_{i_1,\ldots,i_K}, \mathcal{W}_{i_1,\ldots,i_K})| \leq b_l$, it is straight forward to extend generalizing bounds for matrices from [20] to tensors, which holds with probability $1-\delta$ as

$$L_{\mathrm{S}_{\mathrm{Test}}}(l \circ \mathcal{W}) - L_{\mathrm{S}_{\mathrm{Train}}}(l \circ \mathcal{W}) \leq 4R_{\mathrm{S}}(l \circ \mathcal{W}) + b_l\left(\frac{11 + 4\sqrt{\log\frac{1}{\delta}}}{\sqrt{|\mathrm{S}_{\mathrm{Train}}|}}\right),$$

where $R_{\mathrm{S}}(l \circ \mathcal{W})$ is transductive Rademacher complexity theory [3,20] defined as

$$R_{\mathrm{S}}(l \circ \mathcal{W}) = \frac{1}{|\mathrm{S}|}\mathbb{E}_{\sigma}\left[\sup_{\mathcal{W}\in\mathsf{W}} \sum_{j=1}^{|\mathrm{S}|} \sigma_j l(\mathcal{X}_{\alpha_j}, \mathcal{W}_{\alpha_j})\right], \tag{5}$$

where $\sigma_j \in \{-1, 1\},\ j = 1, \ldots, |\mathrm{S}|$ with probability of 0.5 are Rademacher variables.

The following theorem gives the Rademacher complexity for completion model regularized by a reshaped tensor nuclear norm.

Theorem 1 *Consider a K-mode tensor $\mathcal{W} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_K}$. Let us consider any M reshaping sets $(D_1, \ldots, D_M)$ with a hypothesis class of $\mathsf{W} = \{\mathcal{W} | \|\mathcal{W}_{(D_1,\ldots,D_M)}\|_* \leq t\}$. Suppose that for all $(i_1, \ldots, i_K)$, $l(\mathcal{X}_{i_1,\ldots,i_K}, \cdot)$ is Λ-Lipschitz continuous. Then,*
(a) given that $\mathcal{W}$ has a multilinear rank of $(r_1, \ldots, r_K)$, we obtain

$$R_{\mathrm{S}}(l \circ \mathcal{W}) \leq \frac{c\Lambda}{|\mathrm{S}|}\left(\frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\ldots,M}\prod_{i\in D_j} r_i}\right)\gamma_1(\mathcal{W}_{(D_1,\ldots,D_M)})$$
$$\log(4M)\sum_{j=1}^{M}\sqrt{\prod_{p\in D_j} n_p},$$

Table 1: Rademacher complexities for convex norm regularized completion models for a K-mode tensor $\mathcal{T} \in \mathbb{R}^{n\times\cdots\times n}$ with a multilinear rank $(r_1,\ldots,r_K)$. $\gamma_1(\mathcal{X})$ is the largest singular value of $\mathcal{X}$, G reshaping sets of $D^{(s)} = (D_1^{(s)},\ldots,D_{m_s}^{(s)})$, $g = 1,\ldots,G$, and c, Λ, and $B_{\mathcal{T}}$ are constants.

Norm	**Rademacher complexity** $R_{\mathrm{S}}(l\circ\mathcal{W})$
Overlapped norm	$\frac{c\Lambda}{\lvert\mathrm{S}\rvert}\sum_{j=1}^{K}\sqrt{r_j}B_{\mathcal{T}}\big(\sqrt{n^{K-1}}+\sqrt{n}\big)$
Latent trace norm	$\frac{c\Lambda}{\lvert\mathrm{S}\rvert}\min_{j=1,\ldots,K}\sqrt{r_j}B_{\mathcal{T}}\big(\sqrt{n^{K-1}}+\sqrt{n}\big)$
Scaled latent trace norm	$\frac{c\Lambda}{\lvert\mathrm{S}\rvert}\min_{j=1,\ldots,K}\sqrt{\frac{r_j}{n}}B_{\mathcal{T}}\big(\sqrt{n^{K}}+n\big)$
Schatten TT norm	$\frac{c\Lambda}{\lvert\mathrm{S}\rvert(K-1)}\sum_{k=1}^{K-1}\min\big(\prod_{i=1}^{k}\sqrt{r_i},\prod_{j=k+1}^{K}\sqrt{r_j}\big) B_{\mathcal{T}}\sqrt{n^{\lceil K/2\rceil}}$
Square norm	$\frac{c\Lambda}{\lvert\mathrm{S}\rvert}\Big(\frac{\prod_{k=1}^{K}r_k}{\max_{j=1,2}\prod_{i\in D_j}r_i}\Big)\gamma_1(\mathcal{W}_{(D_1,D_2)}) \log(8)\big(\sqrt{n^{\lvert D_1\rvert}}+\sqrt{n^{\lvert D_2\rvert}}\big)$
Tensor nuclear norm	$\frac{c\Lambda}{\lvert\mathrm{S}\rvert}\Big(\frac{\prod_{k=1}^{K}r_k}{\max_{j=1,\ldots,K}r_j}\Big)\gamma_1(\mathcal{W})\log(4K)\sqrt{Kn}$
Reshaped tensor nuclear norm	$\frac{c\Lambda}{\lvert\mathrm{S}\rvert}\Big(\frac{\prod_{k=1}^{K}r_k}{\max_{j=1,\ldots,M}\prod_{i\in D_j}r_i}\Big)\gamma_1(\mathcal{W}_{(D_1,\ldots,D_M)}) \log(4M)\sum_{j=1}^{M}\sqrt{n^{\lvert D_j\rvert}}$
Reshaped latent tensor nuclear norm	$\frac{c\Lambda}{\lvert\mathrm{S}\rvert}\min_{g\in G}\Big(\frac{\prod_{k=1}^{K}r_k}{\max_{j=1,\ldots,M}\prod_{i\in D_j^{(g)}}r_i}\Big) \gamma_1(\mathcal{W}^{(g)}_{(D_1^{(g)},\ldots,D_{M_g}^{(g)})})\max_{g\in G}\log(4M_g) \sum_{j=1}^{M_g}\sqrt{n^{\lvert M_g\rvert}}$

(b) given that $\mathcal{W}$ has a CP rank of r_{cp}, we obtain

$$R_{\mathrm{S}}(l\circ\mathcal{W}) \leq \frac{c\Lambda}{\lvert\mathrm{S}\rvert}r_{cp}\gamma_1(\mathcal{W}_{(D_1,\ldots,D_M)})\log(4M)\sum_{j=1}^{M}\sqrt{\prod_{p\in D_j}n_p},$$

where c is a constant.

Using the Theorem 1, we can obtain the Rademacher complexities for tensor nuclear norm by considering $|D_1| = |D_2| = \cdots = |D_K| = 1$ and the square norm by two reshaping sets of $|D_1|$ and $|D_2|$ such that $\prod_{p\in D_1}n_p \approx \prod_{q\in D_2}n_q$. We summarize Rademacher complexities of convex low-rank tensor norms in Table 1 for a tensor with equal mode dimensions ($n_1 = n_2 = \ldots = n_K = n$).

From Table 1 and Theorem 1, we see that norms constructed using the tensor nuclear norm lead to better bounds compared to the overlapped trace norm, latent trace norm, and the scaled latent trace norm. Further, we see that the mode based components of the Rademacher complexity would have the smallest value with the tensor nuclear norm ($\log(4K)\sqrt{Kn}$). It is also clear that for any reshaping set, we find that $\log(4K)\sqrt{Kn} \leq \log(4M)\sum_{j=1}^{M}\sqrt{n^{|D_j|}}$. This observation might lead us to conclude that the tensor nuclear norm is better than all the other norms. However, considering the multilinear rank such that $1 < r_1 \leq r_2 \leq \cdots \leq r_K$, we can always find $M < K$ reshaping sets $D_1, D_2, \ldots, D_M$ such that $\frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\ldots,K} r_j} \geq \frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\ldots,M} \prod_{i \in D_j} r_i}$. In other words, we can reshape the tensor such that the Rademacher complexity for the reshaped tensor nuclear norm is bounded with a smaller rank based component compared to the tensor nuclear norm.

It is not known how reshaping a tensor changes the CP rank of original tensor into the rank of the reshaped tensor except that $\mathrm{Rank}(\mathcal{X}_{(D_1,\ldots,D_M)}) \leq r_{cp}$ ([16] and Lemma 4 in appendix). However, Theorem 1 shows that reshaping results in a mode based component of $\log(4M)\sum_{j=1}^{M}\sqrt{n^{|D_j|}}$ for the Rademacher complexity, which indicates that selecting a reshaping set that gives a lower mode based components can lead to a lower generalization bound compared the square norm or the tensor nuclear norm. Furthermore, it is clear that the reshaping a tensor and regularization using the tensor nuclear norm lead to a lower generalization bound compared with multilinear rank based norms such as the overlapped trace norm, latent trace norm, and scaled latent trace norms and tensor train rank based Schatten TT norm.

The next theorem provides the Rademacher complexity for completion models regularized by the reshaped latent tensor nuclear norm.

Theorem 2 *Let us consider a K-mode tensor $\mathcal{W} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$. Let us consider a collection of G collection of reshaping sets $D_{\mathrm{L}} = (D^{(1)}, \ldots, D^{(G)})$ where each $D^{(s)} = (D_1^{(s)}, \ldots, D_{M_s}^{(s)})$, $s = 1, \ldots, G$ consists a reshaping set for a M_s-mode reshaped tensor. Consider the hypothesis class $\mathsf{W}_{\mathsf{rl}} = \{\mathcal{W} | \mathcal{W}^{(1)} + \cdots + \mathcal{W}^{(G)} = \mathcal{W}, \|\mathcal{W}\|_{\mathrm{r\_latent}(D_{\mathrm{L}})} \leq t\}$ for a given set of reshaping set $(D_1, \ldots, D_M)$. Suppose that for all $\mathcal{X}_{i_1,\ldots,i_K}$, $l(\mathcal{X}_{i_1,\ldots,i_K}, \cdot)$ is Λ-Lipschitz continuous. Then, (a) when $\mathcal{W}$ has a multilinear rank of $(r_1, \ldots, r_K)$, we obtain*

$$R_{\mathrm{S}}(l \circ \mathcal{W}) \leq \frac{c\Lambda}{|\mathrm{S}|} \min_{g \in G} \left(\frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\ldots,M} \prod_{i \in D_j^{(g)}} r_i} \right) \gamma_1(\mathcal{W}^{(g)}_{(D_1^{(g)},\ldots,D_{M_g}^{(g)})}) \max_{g \in G} \log(4M_g) \sum_{j=1}^{M_g} \sqrt{\prod_{p \in D_j^{(g)}} n_p}.$$

(b) when $\mathcal{W}$ has a CP rank of r_{cp}, we obtain

$$R_{\mathrm{S}}(l \circ \mathcal{W}) \leq \frac{c\Lambda}{|\mathrm{S}|} r_{cp} \min_{g} \gamma_1(\mathcal{W}^{(g)}_{(D_1^{(g)},\ldots,D_{M_g}^{(g)})}) \max_{g \in G} \log(4M_g) \sum_{j=1}^{M_g} \sqrt{\prod_{p \in D_j^{(g)}} n_p}.$$

where c is a constant.

Theorem 2 shows that latent reshaped tensor nuclear norm bounds the Rademacher complexity by the largest mode based component that results from all the reshaping sets. Further, with the multilinear rank of the tensor the Rademacher complexity is bounded by the smallest rank based component that results from all the reshaping sets. This observation indicates that properly selecting a set of reshaping sets to use with the latent reshaped tensor nuclear norm can lead to a lower generalization bound.

We want to point out that the largest singular values ($\gamma_1(\cdot)$) that appear in both Theorems 1 and 2 can be upper bounded by taking the largest singular value with respect to all possible reshaping sets for a tensor. However, we do not use such a bounding to keep the Rademacher complexities small.

4.1 Optimal Reshaping Strategies

Given that we have an understanding of the ranks of the tensor, Theorem 1 can be used to select a reshaping set such that reshaped tensor has a smaller rank and relatively smaller mode dimensions. However, since we do not know the rank in advance, selecting a reshaping set such that the reshaped tensor does not have large mode dimensions would lead to a better performance.

To avoid the difficulty in choosing a single reshaping set, we can use the reshaped latent tensor nuclear norm by choosing several reshaping sets that agree with our observation in Theorem 1. However, since the Rademacher complexity if bounded by the largest mode based components as shown in Theorem 2, it is important not to select reshaping sets that result in a tensor with large dimensions. A general strategy to create the reshaping sets by selecting the original tensor and other reshaping sets that do not result in large mode dimensions compared to the original tensor.

5 Optimization Procedures

It has been shown that learning by constraining the tensor nuclear norm is the NP-Complete problem [8], which makes solving the problems (3) and (4) computationally difficult. In [24] an approximation method have been proposed to compute the spectral norm by computing largest singular vectors on each mode, which is combined with Frank-Wolfe optimization method to solve (3). We adopt the approximation methods [24] to solve our proposed completion models with reshaped tensors (3) and (4). We found that solutions using the approximation methods provide agreements with our theoretical results related to excess risk bounds as we show in the Section 7. However, there is no theoretical analysis available to understand how well the approximation method results in a solution compared to a exact solution.

The optimization method proposed in [24] uses an approximation method for the spectral norm using a recursive algorithm based on singular value decomposition with respect to each mode. However, we adopt a more simpler approach as given in Algorithm 1, which we believe is more easier to implement. Using the approximation method, we provide an optimization procedure to solve the completion model that is regularized by a single reshaped norm in the Algorithm 2. The optimization procedure in Algorithm 2 is also similar to the

Frank-Wolfe based optimization procedure proposed in [24]. The additions in Algorithm 2 to [24] are the computation of the spectral norm of the reshaped tensor in step 7 and the conversion of the reshaped tensor to the original dimensions in step 12. Here, we want to recall Definition 1 to refer to the reshaping operator $\Pi_{(D_1,\ldots,D_M)}()$ and its inverse operator $\Pi^{\top}_{(D_1,\ldots,D_M)}()$ for any given reshaping set $(D_1, D_2, \ldots, D_M)$.

Input: $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_K}$
Output: $w_1, \ldots, w_K, sv$
$\mathcal{Y} = \mathcal{A}$
for $k = 1, \ldots, K-1$ **do**
    $M = \text{reshape}(\mathcal{Y}, [n_k, n_{k+1} \cdots n_K])$
    $(w_k, s_k, v_k) = \text{svd}(M, 1)$
    $\mathcal{Y} = \mathcal{Y} \times_1 w_k$
    $\mathcal{Y} = \text{reshape}(\mathcal{Y}, [n_{k+1}, \ldots, n_K])$
end for
$w_K = \frac{\mathcal{Y} \times_1 w_{K-1}}{\|\mathcal{Y} \times_1 w_{K-1}\|_2}$
$sv = \mathcal{A} \times_1 w_1 \times_2 w_2 \cdots \times_K w_K$

Algorithm 1: ApproxSpectralNorm($\mathcal{A}$)

1: **Input:** $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_K}$ with observed indexes Ω, Regularization parameter λ. Initial $\mathcal{W}^0$, Maximum iterations T, Reshaping dimensions $(D_1, D_2, \ldots, D_M)$
2: **Output:** $\mathcal{W}^T$
3: $t = 0$
4: **repeat**
5:     $t = t + 1$
6:     $f_{\mathcal{W}}(\mathcal{W}^t) = \frac{1}{2}\|\Omega(\mathcal{W}^t) - \Omega(\mathcal{X})\|_{\mathrm{F}}^2$
7:     $w_1, \ldots, w_K, sv = \text{ApproxSpectralNorm}(\Pi_{(D_1,\ldots,D_M)}(\nabla_{\mathcal{W}} f_{\mathcal{W}}(\mathcal{W}^t)))$
8:     $\mathcal{W}^t_{descent} = -\lambda w_1 \otimes w_2, \cdots \otimes w_K$
9:     **if** linesearch == True **then**
10:         Using an appropriate line search method (e.g.Yuan et al (2016))
11:     **else**
12:         $\mathcal{W}^{t+1} = \mathcal{W}^t + \frac{2}{t+2} \Pi^{\top}_{(D_1,\ldots,D_M)}(\mathcal{W}^t_{descent})$
13:     **end if**
14: **until** $t = T$

Algorithm 2: A Frank-Wolfe optimization method for a regularization with a reshaped tensor norm

Next, we give an algorithm to solve the completion model regularized by the reshaped latent tensor nuclear norm. The Frank-Wolfe optimization method has also been applied to efficiently solve learning models regularized by the latent trace norms [5]. We follow their approach to design Frank-Wolfe method for the reshaped latent tensor nuclear norm and Algorithm 3 shows the steps for optimization. From Lemma 1, we know that we need to find the reshaping with the largest spectral norm each t step to update the Frank-Wolfe procedure. This is shown in the lines 7-11 in the Algorithm 3.

Input: $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_K}$ with observed indexes Ω, Regularization parameter λ. Initial $\mathcal{W}^0$, Maximum iterations T, G Reshaping sets $(D^{(1)}, \ldots, D^{(G)})$ with $D^{(g)} = (D_1^{(g)}, \ldots, D_{M_g}^{(g)}),\ g = 1, \ldots, G$
Output: $\mathcal{W}^T$
$t = 0$
repeat
  $t = t + 1$
  $f_{\mathcal{W}}(\mathcal{W}^t) = \frac{1}{2}\|\Omega(\mathcal{W}^t) - \Omega(\mathcal{X})\|_{\mathrm{F}}^2$
  for $g = 1, \ldots, G$ **do**
    $w_1, \ldots, w_{M_g}, sv = \mathrm{ApproxSpectralNorm}(\Pi_{(D_1^{(g)}, \ldots, D_{M_g}^{(g)})}(\nabla_{\mathcal{W}} f_{\mathcal{W}}(\mathcal{W}^t)))$
    $sv\_array(g) = sv$
  end for
  $i = \arg\max_i(sv\_array)$
  $w_1, \ldots, w_{M_i}, sv = \mathrm{ApproxSpectralNorm}(\Pi_{(D_1^{(i)}, \ldots, D_{M_i}^{(i)})}(\nabla_{\mathcal{W}} f_{\mathcal{W}}(\mathcal{W}^t)))$
  $\mathcal{W}_{descent}^t = -\lambda \Pi^{\top}_{(D_1^{(i)}, \ldots, D_{M_i}^{(i)})}(w_1 \otimes w_2, \cdots \otimes w_{M_i})$
  if linesearch == True **then**
    Using an appropriate line search method (e.g.Yuan et al (2016))
  else
    $\mathcal{W}^{t+1} = \mathcal{W}^t + \frac{2}{t+2}(\mathcal{W}_{descent}^t)$
  end if
until $t = T$

Algorithm 3: A Frank-Wolfe optimization method for a regularization with the reshaped latent tensor nuclear norm

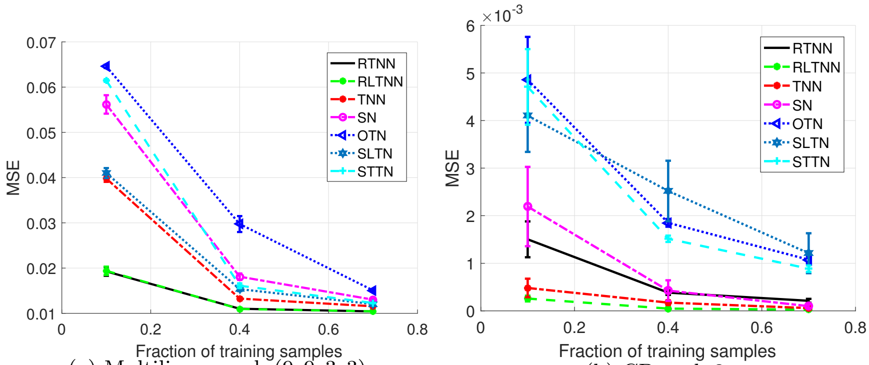
6 Experiments

In this section, we give simulation and real-data experiments.

6.1 Simulation Experiments

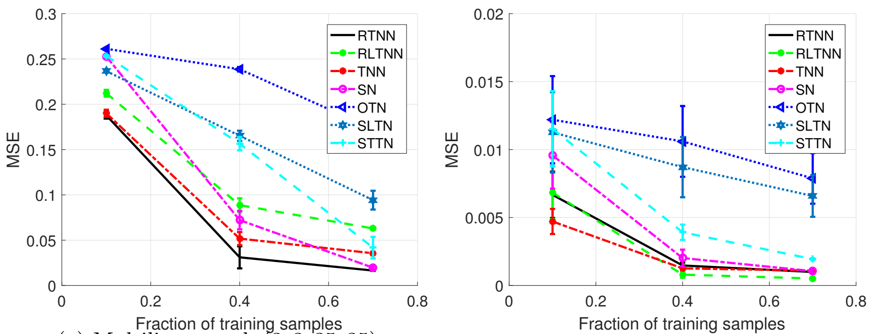
We created simulation experiments for tensor completion using tensors with some fixed multilinear rank and CP rank. We create a K-mode tensor with the multilinear rank of $(r_1, \cdots, r_K)$ by generating a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ using the Tucker decomposition [12] as $\mathcal{T} = \mathcal{C} \times_1 U_1 \times_2 U_2 \times_3 \cdots \times_K U_K$, where $\mathcal{C} \in \mathbb{R}^{r_1 \times \cdots \times r_K}$ is a core tensor whose elements are sampled from a normal distribution specifying the multilinear rank $(r_1, \cdots, r_K)$ and $U_k \in \mathbb{R}^{r_k \times n_k},\ k = 1, \ldots, K$ are orthogonal component matrices. We create a tensor with the CP rank of r using the CP decomposition [12] as $\mathcal{T} = \sum_{i=1}^{r} c_i u_{1i} \otimes u_{2i} \otimes \cdots \otimes u_{Ki}$ where $u_{ki} \in \mathbb{R}^{n_k},\ k = 1, \ldots, K,\ i = 1, \ldots, r$ are sampled from a normal distribution and normalized such that $\|u_{ki}\|_2^2 = 1$ and $c_i \in \mathbb{R}^{+}$. From the total number of elements in the generated tensors, we randomly selected 10, 40, and 70 percentages as training sets, and from the remaining we selected 10 percent of elements as validation set, and the rest were taken as test data. We conducted 3 simulations for each randomly generated tensor.

For all simulation experiments, we tested completion using our proposed completion models (3) with the reshaped tensor nuclear norm (abbreviated as



(a) Multilinear rank $(9, 9, 3, 3)$ (b) CP rank 3

Fig. 1: Performances of completion of the tensors (a) Tensor $\mathcal{T} \in \mathbb{R}^{10\times10\times40\times40}$ with a multilinear rank $(9, 9, 3, 3)$ and (b) $\mathcal{T} \in \mathbb{R}^{10\times10\times10\times10\times10}$ with a CP rank 3



(a) Multilinear rank $(3, 3, 35, 35)$ (b) CP rank 243

Fig. 2: Performances of completion of the tensors (a) Tensor $\mathcal{T} \in \mathbb{R}^{10\times10\times40\times40}$ with a multilinear rank $(3, 3, 35, 35)$ and (b) $\mathcal{T} \in \mathbb{R}^{10\times10\times10\times10\times10}$ with a CP rank 243

RTNN) and (4) with the reshaped latent tensor nuclear norm (abbreviated as RLTNN). Additionally, we performed completion using the tensor nuclear norm (abbreviated as TNN) without reshaping and the square norm (abbreviated as SN). As further baseline methods, we used tensor completion with regularization using the overlapped trace norm (abbreviated as OTN), scaled latent trace norm (abbreviated as SLTN), and the Schatten TT norm [10] (abbreviated as STTN). As the performance measure of completion, we calculated the mean squared error (MSE) on the validation data and test data. For all completion models, we performed cross-validation of regularization parameters in power of 2^x, with x ranging from -5 to 15 with intervals of 0.25.

For our first simulation experiment, we created a 4-way tensors $\mathcal{T}_1 \in \mathbb{R}^{n_1\times n_2\times n_3\times n_4}$ with $n_1 = n_2 = 10, n_3 = n_4 = 40$ with a multilinear rank of $(r_1, r_2, r_3, r_4) = (9, 9, 3, 3)$. From [16] we can reshape $\mathcal{T}$ by using a reshaping set of $(D_1, D_2) = ((n_1, n_3), (n_2, n_4))$ such that it creates a square matrix for the square norm. From Theorem 1, we see that the rank components in the Rademacher complexity for the nuclear norm and the square norm are $\prod_{k=1}^{K} r_k/(\max_{j=1,\ldots,4} r_j) = 243$ and $\prod_{k=1}^{K} r_k/(\max_{j=1,2} \prod_{i \in D_j} r_i) = 27$, respectively. Further, Theorem 1 shows that the mode based components for the nuclear norm and the square norm are $\log(4 \cdot 4)(\sum_{k=1}^{4} \sqrt{n_k}) \approx 53$ and $\log(4 \cdot 2)(\sqrt{n_1 n_3} + \sqrt{n_2 n_4}) \approx 83$, respectively. This leads to a lower generalization bound for the nuclear norm compared to the square norm justifying its better performance as shown in the Figure 1 (a). However, Theorem 1 indicates that the lowest generalization bound is obtained by using the reshaping set $(D'_1, D'_2, D'_3) = ((n_1, n_2), n_3, n_4)$, which combines the high ranked modes (mode 1 and mode 2) together resulting in a rank based component of $\prod_{k=1}^{K} r_k/(\max_{j=1,2,3} \prod_{i \in D'_j} r_i) = 9$ and a mode based component of $\log(4 \cdot 3)(\sqrt{n_1 n_2} + \sqrt{n_3} + \sqrt{n_4}) \approx 56$. Figure 1 (a) agrees with our theoretical analysis showing that our proposed reshaped tensor nuclear norm obtains the best performance compared to other norms. For the reshaped latent tensor nuclear norm, we combined the two reshaping sets $((n_1, n_2, n_3, n_4), ((n_1, n_2), n_3, n_4))$. Applying Theorem 2, we see that this reshaping set combination leads to a lower Rademacher complexity. However, this combination only gave a comparable performance to the reshaped tensor nuclear norm.

As our second simulation, we created a 5-mode tensor $\mathcal{T}_2 \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_5}$, where $n_1 = n_2 = \ldots = n_5 = 10$ with a CP rank of 3. From Theorem 1, we know that we can only consider the mode based component of the Rademacher complexity to obtain a lower generalization bound. For the square norm we can use a reshaping set such as $(D_1, D_2) = ((n_1, n_2), (n_3, n_4, n_5))$, which results in the mode based component as $\log(4 \cdot 2)(\sqrt{n_1 n_2} + \sqrt{n_3 n_4 n_5}) \approx 86$. The tensor nuclear norm leads to a mode based component of $\log(4 \cdot 5)(\sum_{k=1}^{5} \sqrt{n_k}) \approx 47$. As an alternative reshaping method, we propose to combine any two modes together to create a reshaping set such as $(D'_1, D'_2, D'_3) = (n_1, n_2, n_3, (n_4, n_5))$ for the reshaped tensor nuclear norm, which lead to a mode based component of $\log(4 \cdot 4)(\sqrt{n_1} + \sqrt{n_2} + \sqrt{n_3} + \sqrt{n_4 n_5}) \approx 54$. Comparing the Rademacher complexities using the mode based components we see that the lowest generalization bound is given by the tensor nuclear norm. Figure 1 (b) shows that our theoretical observation is accurate since the tensor nuclear norm gives the best performance compared to other two reshaped norms. For the reshaped latent tensor nuclear norm we used all the 10 combinations of two modes combinations, which resulted in reshaping sets of $D = (((n_1, n_2), n_3, n_4, n_5), (n_1, (n_2, n_3), n_4, n_5), \ldots, (n_1, n_2, n_3, (n_4, n_5)))$. Figure 1 (b) shows that the reshaped latent tensor nuclear norm has outperformed the tensor nuclear norm.

The next simulation again focuses on a different multilinear rank for the 4-way tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times n_4}$ with $n_1 = n_2 = 10, n_3 = n_4 = 40$. Figure 2 (a) shows the simulation experiment with multilinear rank of $(3, 3, 35, 35)$. Again from [16] we can reshape $\mathcal{T}$ by using a reshaping set of $(D_1, D_2) = ((n_1, n_3), (n_2, n_4))$ or $(D_1, D_2) = ((n_1, n_4), (n_2, n_3))$ to create a square matrix to use with the square norm. From Theorem 1, the square norm will result in a rank based component of $\prod_{k=1}^{K} r_k/(\max_{j=1,2}\prod_{i\in D_j} r_i) = 105$ and a mode based component of $\log(4\cdot 2)(\sqrt{n_1 n_3}+\sqrt{n_2 n_4}) \approx 63$. However, if we combine the high ranked modes 3 and 4 together to create a reshaping set $(D'_1, D'_2, D'_3) = (n_1, n_2, (n_3, n_4))$ for the reshaped tensor nuclear norm, then the rank based component will decrease to $\prod_{k=1}^{K} r_k/(\max_{j=1,2,3}\prod_{i\in D'_j} r_i) = 9$ and mode based component will decrease to $\log(4\cdot 2)(\sqrt{n_1}\sqrt{n_3}+\sqrt{n_2 n_4}) \approx 55$. Furthermore, the tensor nuclear norm leads to a rank based component of $\prod_{k=1}^{K} r_k/(\max_{j=1,\ldots,4} r_j) = 315$ and mode based component of $\log(4\cdot 4)(\sum_{k=1}^{4}\sqrt{n_k}) \approx 53$ resulting in a larger generalization bound compared to the proposed reshaped set $(D'_1, D'_2, D'_3) = (n_1, n_2, (n_3, n_4))$. This analysis is also confirmed with the experimental results as shown in Figure 2 (a) where the reshaped tensor nuclear norm gives the best performance. Using the Theorem 2, we find that if we use reshaping sets $((n_1, n_2, n_3, n_4), ((n_1, n_2), n_3, n_4))$ for the reshaped latent tensor nuclear norm, the Rademacher complexity will be bounded by the smaller rank based component from the reshaping set (n_1, n_2, n_3, n_4) and the mode based component from $((n_1, n_2), n_3, n_4)$. However, the reshaped latent tensor nuclear norm was not able to perform better than the tensor nuclear norm or the proposed reshaped norm with $(D'_1, D'_2, D'_3) = (n_1, n_2, (n_3, n_4))$.

The final simulation result shown in Figure 2 (b) is for a tensor $\mathcal{T} \in \mathbb{R}^{10\times 10\times 10\times 10\times 10}$ with CP rank of 243. For this experiment, we used the same reshaping strategies as in the previous experiment with CP rank in Figure 1 (b). We see that when the fraction of training samples is less than 40 percent the tensor nuclear norm has given the best performance. When the fraction of the training samples increases the reshaped latent tensor nuclear norm has outperformed the tensor nuclear norm.

6.2 Multi-View Video Completion

We performed completion on multi-view video data using the EPFL data set: Multi-camera Pedestrian Videos data [1]. Videos in this data set capture sequentially entering a room and walking around of four people from 4 views using 4 synchronized cameras. We down-sampled each video frame to a height of 96 and width of 120 to obtain a frame as a RGB-color image with dimensions of $96 \times 120 \times 3$. We sequentially selected 391 frames from each video. Combining all the video frames from all views resulted in a tensor of dimensions of $96 \times 120 \times 3 \times 391 \times 4$ (height $\times$ width $\times$ color $\times$ frames $\times$ views).

To evaluate completion, we randomly removed entries from the multi-view tensor and performed completion using the remaining elements. We randomly selected percentages of 2, 4, 8, 16, 32, and 64 of the total number of elements in the tensor as training elements. As our validation set we selected 10 percent of the total number of elements. The rest of the remaining elements were taken as the test set. For the square norm, we considered the reshaping set $((\text{height}, \text{width}), (\text{color}, \text{frames}, \text{views}))$. For the reshaped tensor nuclear norm, we experimentally found that the reshaping set $((\text{height}, \text{views}), (\text{width}, \text{color}), (\text{frames}))$ gives the best performance.

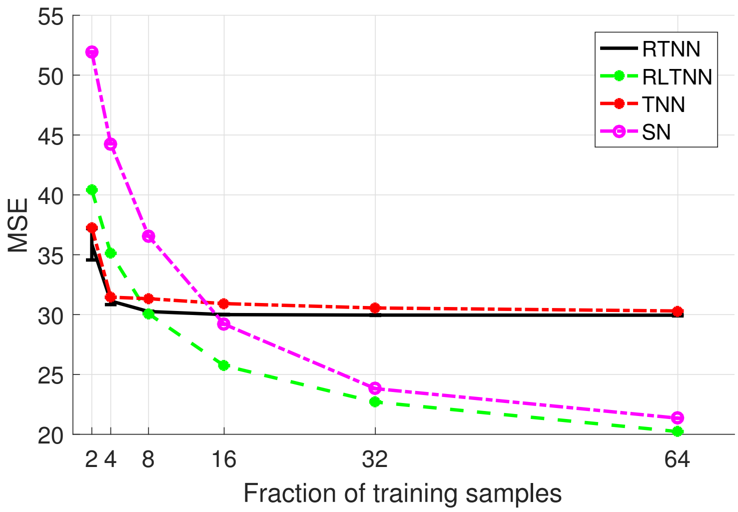


Fig. 3: Tensor completion of the multi-video tensor.

To create the reshaping set for the reshaped tensor nuclear norm we combined the reshaping sets for the square norm and the reshaped tensor nuclear norm with the unreshaped original tensor. The resulting set is $D = ((\text{height}, \text{width}, \text{color}, \text{frames}, \text{views}),\ ((\text{height}, \text{width}),(\text{color}, \text{frames}, \text{views})),$ $((\text{height}, \text{views}),(\text{width}, \text{color}), (\text{frames})))$. We cross-validated all the completion models with regularization parameters out of $10^{-1}, 10^{-0.75}, 10^{-0.5}, \ldots, 10^{7}$.

Figure 3 shows that when the training set is small (or the reshaped tensor nuclear norm is sparse) the reshaped tensor nuclear norm and the tensor nuclear norm have given good performance compared to the square norm. When the percentage of observed elements increases more than 16 percent, the square norm outperforms the other norms. However, the reshaped latent tensor nuclear norm has shown to be adaptive to all fractions of training samples and has given the overall best performance.

7 Conclusions

In this paper, we generalize tensor reshaping for low-rank tensor regularization and introduce the reshaped tensor nuclear norm and the reshaped latent tensor nuclear norm. We propose tensor completion models that are regularized by the proposed norm. Using generalization bound analysis of the proposed completion models we show that the proposed norms lead to smaller Rademacher complexity bounds compared to exiting norms. Further, using our theoretical analysis we discuss optimal conditions to create reshaped tensor nuclear norms. Simulation and real-data experiments confirm our theoretical analysis.

Our research opens up several future research directions. The most important research should be focused on developing theoretical guaranteed methods for optimization of completion models regularized by the the proposed tensor nuclear norms. Though the approximation methods we have adopted for computing the tensor spectral norm to be used with the Frank-Wolfe from [24] provide performances that agrees with our generalization bounds we do not know its approximation error. We believe that future theoretical investigations are needed to understand qualitative properties of the proposed optimization procedures using the approximation method. Furthermore, optimization methods for nuclear norms that can scale for large-scale higher order tensors would be another important future research direction. Another important research direction is to further explore the theoretical foundation of tensor completion using the reshaped tensor nuclear norm. In this regard, recovery bounds [25] would provide us with stronger bounds on sample complexities for our proposed method.

Acknowledgements H.M. has been supported in part by JST ACCEL [grant number JPMJAC1503], MEXT Kakenhi [grant numbers 19H04169] and AIPSE by Academy of Finland.

References

1. Berclaz, J., Fleuret, F., Turetken, E., and Fua, P. Multiple Object Tracking using K-Shortest Paths Optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011.
2. Carroll, J. D. and Chang, J.-J. Analysis of individual differences in multidimensional scaling via an n-way generalization of “eckart-young” decomposition. *Psychometrika*, 35 (3):283–319, Sep 1970.
3. El-Yaniv, R. and Pechyony, D. Transductive rademacher complexity and its applications. In *Learning Theory*, volume 4539, pp. 157–171. 2007.
4. Fazel, M., Hindi, H., and Boyd, S. P. A rank minimization heuristic with application to minimum order system approximation. In *Proceedings of the 2001 American Control Conference. (Cat. No.01CH37148)*, volume 6, pp. 4734–4739, June 2001.
5. Guo, X., Yao, Q., and Kwok, J. T. Efficient sparse low-rank tensor completion using the frank-wolfe algorithm. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, pp. 1948–1954, 2017.
6. Hackbusch, W. *Tensor Spaces and Numerical Tensor Calculus*. Springer Series in Computational Mathematics. Springer Berlin Heidelberg, 2012. ISBN 9783642280276.
7. Harshman, R. A. Foundations of the PARAFAC procedure: models and conditions for an explanatory multimodal factor analysis. *UCLA Working Papers in Phonetics*, 16:1–84, 1970.
8. Hillar, C. J. and Lim, L.-H. Most tensor problems are np-hard. *J. ACM*, 60(6), November 2013. ISSN 0004-5411.
9. Hitchcock, F. L. The expression of a tensor or a polyadic as a sum of products. *J. Math. Phys*, 6(1):164–189, 1927.
10. Imaizumi, M., Maehara, T., and Hayashi, K. On tensor train rank minimization : Statistical efficiency and scalable algorithm. In *NIPS*, pp. 3933–3942, 2017.
11. Karatzoglou, A., Amatriain, X., Baltrunas, L., and Oliver, N. Multiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering. In *RecSys*, pp. 79–86. ACM, 2010.
12. Kolda, T. G. and Bader, B. W. Tensor Decompositions and Applications. *SIAM Review*, 51(3):455–500, 2009.

13. Latała, R. Some estimates of norms of random matrices. *Proc. Amer. Math. Soc.*, 133 (5):1273–1282, 2005.
14. Lim, L. and Comon, P. Blind multilinear identification. *IEEE Trans. Information Theory*, 60(2):1260–1280, 2014.
15. Liu, J., Musialski, P., Wonka, P., and Ye, J. Tensor completion for estimating missing values in visual data. In *ICCV*, pp. 2114–2121, 2009.
16. Mu, C., Huang, B., Wright, J., and Goldfarb, D. Square deal: Lower bounds and improved relaxations for tensor recovery. In *ICML*, pp. 73–81, 2014.
17. Oseledets, I. V. Tensor-train decomposition. *SIAM J. Sci. Comput.*, 33(5):2295–2317, September 2011. ISSN 1064-8275.
18. Rai, P., Hu, C., Harding, M., and Carin, L. Scalable probabilistic tensor factorization for binary and count data. IJCAI'15, pp. 3770–3776. AAAI Press, 2015. ISBN 978-1-57735-738-4.
19. Raskutti, G., Chen, H., and Yuan, M. Convex regularization for high-dimensional multi-response tensor regression. *CoRR*, abs/1512.01215v2, 2015.
20. Shamir, O. and Shalev-Shwartz, S. Matrix completion with the trace norm: Learning, bounding, and transducing. *Journal of Machine Learning Research*, 15:3401–3423, 2014.
21. Song, Q., Ge, H., Caverlee, J., and Hu, X. Tensor completion algorithms in big data analytics. *CoRR*, abs/1711.10105, 2017.
22. Tomioka, R. and Suzuki, T. Convex tensor decomposition via structured schatten norm regularization. In *NIPS*, 2013.
23. Wimalawarne, K., Sugiyama, M., and Tomioka, R. Multitask learning meets tensor factorization: task imputation via convex optimization. In *NIPS*. 2014.
24. Yang, Y., Feng, Y., and Suykens, J. A. K. A rank-one tensor updating algorithm for tensor completion. *IEEE Signal Processing Letters*, 22(10):1633–1637, Oct 2015. ISSN 1070-9908.
25. Yuan, M. and Zhang, C.-H. On tensor completion via nuclear norm minimization. *Foundations of Computational Mathematics*, 16(4):1031–1068, Aug 2016.
26. Zheng, V. W., Cao, B., Zheng, Y., Xie, X., and Yang, Q. Collaborative filtering meets mobile recommendation: A user-centered approach. In *AAAI*, AAAI'10, pp. 236–241. AAAI Press, 2010.

A Dual Norms of Reshaped Tensor Nuclear Norms

In this section, we discuss the dual norm of the proposed reshaped tensor nuclear norm. The dual norm is useful in developing optimization procedures and proving theoretical bounds.

The dual norm the tensor nuclear norm [24,25] of a K-mode tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ is given by

$$\|\mathcal{T}\|_{\mathrm{op}} = \max_{\|y_i\|_2 = 1, 1 \leq i \leq K} \langle \mathcal{T}, y_1 \otimes y_2 \otimes \cdots \otimes y_K \rangle. \tag{6}$$

This definition applies to all the tenor nuclear norms including reshaped norms.

The next Lemma provides the dual norm for the reshaped latent tensor nuclear norm.

Lemma 1 *The dual norm of the reshaped latent tensor nuclear norm for a tensor $\mathcal{W} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ for a collection of G reshaping sets $D_{\mathrm{L}} = (D^{(1)}, \ldots, D^{(G)})$ is*

$$\|\mathcal{W}\|_{\mathrm{r\_latent}(D_{\mathrm{L}})^*} = \max_g \|\mathcal{W}_{(D^{(g)})}\|_{\mathrm{op}}.$$

Proof Using the standard formulation of the dual norm, we write the dual norm for $\|\mathcal{W}\|_{\mathrm{r\_latent}(D_{\mathrm{L}})^*}$ as

$$\|\mathcal{W}\|_{\mathrm{r\_latent}(D_{\mathrm{L}})^*} = \sup \left\langle \sum_{k=1}^{G} \mathcal{X}^{(k)}, \mathcal{W} \right\rangle \quad \text{s.t.} \inf_{\mathcal{X}^{(1)} + \cdots + \mathcal{X}^{(G)} = \mathcal{X}} \sum_{k=1}^{G} \|\mathcal{X}^{(k)}_{(D^{(k)})}\|_\star \leq 1. \tag{7}$$

The solution to (7) resides on the simplex of $\inf_{\mathcal{X}^{(1)}+\cdots+\mathcal{X}^{(G)}=\mathcal{X}} \sum_{k=1}^{G} \|\mathcal{X}^{(k)}_{(D^{(k)})}\|_\star \leq 1$ and one of the edges of the simplex is a solution. Then, we can take any $g \in 1,\ldots,G$ such that $\mathcal{X}^{(g)} = \mathcal{X}$ and all $\mathcal{X}^{(k\neq g)} = 0$, such that we arrange (7) as

$$\|\mathcal{W}\|_{\text{r\_latent}(D_{\text{L}})^*} = \sup_{g\in 1,\ldots,G} \left\langle \mathcal{X}_{(D^{(g)})}, \mathcal{W}_{(D^{(g)})} \right\rangle \quad \text{s.t. } \|\mathcal{X}_{(D^{(g)})}\|_\star \leq 1,$$

which results in the following

$$\|\mathcal{W}\|_{\text{r\_latent}(D_{\text{L}})^*} = \max_{g\in 1,\ldots,G} \|\mathcal{W}_{(D^{(g)})}\|_{\text{op}}.$$

□

B Proofs of theoretical analysis

In this section, we provide proofs of the theoretical analysis in Section 4.

First, we prove the following useful lemmas. These lemmas bound the tensor nuclear norm and the reshaped tensor nuclear norms with respect to the multilinear rank of a tensor.

Lemma 2 *Let $\mathcal{X} \in \mathbb{R}^{n_1\times\cdots\times n_K}$ be a random K-mode tensor with a multilinear rank of $(r_1,\ldots,r_K)$. Let r_{cp} be the CP rank of $\mathcal{X}$, then*

$$\|\mathcal{X}\|_\star = \left\{ \sum_{j=1}^{r_{cp}} \gamma_j | \mathcal{X} = \sum_{j=1}^{r_{cp}} \gamma_j u_{1j} \otimes u_{2j} \cdots \otimes u_{Kj}, \|u_{kj}\|_2^2 = 1, \gamma_j \geq \gamma_{j+1} > 0 \right\}$$
$$\leq \frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\ldots,K} r_i} \gamma_1,$$

where γ_i is the ith singular value of $\mathcal{X}$.

Proof Let us consider the Tucker decomposition of $\mathcal{X}$ as

$$\mathcal{X} = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \cdots \sum_{j_K=1}^{r_K} \mathcal{C}_{j_1,\ldots,j_K} u^{(1)}_{j_1} \otimes u^{(2)}_{j_2} \otimes \cdots \otimes u^{(K)}_{j_K},$$

where $\mathcal{C} \in \mathbb{R}^{r_1\times\cdots\times r_K}$ is the core tensor and $u^j_{(i)} \in \mathbb{R}^{n_j}$, $\|u^j_{(i)}\|_2 = 1$, $i = 1,\ldots,r_i$, $j = 1,\ldots,K$ are component vectors.

Following Chapter 8 of [6], we can express the above Tucker decomposition as

$$\mathcal{X} = \sum_{j_2=1}^{r_2} \cdots \sum_{j_K=1}^{r_K} \underbrace{\left(\sum_{j_1=1}^{r_1} \mathcal{C}_{j_1,\ldots,j_K} u^{(1)}_{j_1} \right)}_{\hat{u}^{(1)}[j_2,\ldots,j_K] \in \mathbb{R}^{n_1}} \otimes u^{(2)}_{j_2} \otimes \cdots \otimes u^{(K)}_{j_K}, \tag{8}$$

where we have taken summation over the multiplications of core tensor elements with component vectors of the mode 1. It is easy to see that we can consider the summation over component vectors of any other mode.

By considering $\hat{u}^{(1)}[j_2,\ldots,j_K] = \gamma[j_2,\ldots,j_K] \frac{\hat{u}^{(1)}[j_2,\ldots,j_K]}{\|\hat{u}^{(1)}[j_2,\ldots,j_K]\|_2}$ where $\gamma[j_2,\ldots,j_K] = \|\hat{u}^{(1)}[j_2,\ldots,j_K]\|_2$, it leads to a CP decomposition with rank of $r_{cp} = \frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\ldots,K} r_i}$.

By arranging $\gamma[j_2,\ldots,j_K]$ in descending order along with component vectors $\hat{u}^{(1)}[j_2,\ldots,j_K]$ and renaming them as $\gamma_1 \geq \gamma_2 \geq \ldots$ and u_{1j}, respectively, we obtain

$$\|\mathcal{X}\|_\star = \left\{ \sum_{j=1}^{r_{cp}} \gamma_j | \mathcal{X} = \sum_{j=1}^{r_{cp}} \gamma_j u_{1j} \otimes u_{2j} \cdots \otimes u_{Kj}, \|u_{kj}\|_2^2 = 1, \gamma_j \geq \gamma_{j+1} > 0 \right\},$$

where $u_{kj} \in [u_1^{(k)}, \dots u_{r_k}^{(k)}]$ are component vectors from (8) for each $k = 2, \dots, K$.

Then the final bound is trivial

$$\|\mathcal{X}\|_\star = \left\{ \sum_{j=1}^{r_{cp}} \gamma_j | \mathcal{X} = \sum_{j=1}^{r_{cp}} \gamma_j u_{1j} \otimes u_{2j} \cdots \otimes u_{Kj}, \|u_{kj}\|_2^2 = 1, \gamma_j \geq \gamma_{j+1} > 0 \right\} \leq \frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\dots,K} r_i} \gamma_1.$$

□

Lemma 3 *Let $\mathcal{X} \in \mathbb{R}^{n \times \dots \times n}$ be a random K-mode tensor with multilinear rank of $(r_1, \dots, r_K)$. We consider a set of M reshaping modes D_i, $i = 1, \dots, M$. Let r_{cp} be the CP rank of $\mathcal{X}$, then*

$$\|\mathcal{X}_{(D_1,\dots,D_M)}\|_\star = \left\{ \sum_{j=1}^{r_{cp}} \gamma_j | \mathcal{X}_{(D_1,\dots,D_M)} = \sum_{j=1}^{r_{cp}} \gamma_j u_{1j} \otimes u_{2j} \cdots \otimes u_{Mj}, \|u_{kj}\|_2^2 = 1, \gamma_j \geq \gamma_{j+1} > 0 \right\} \leq \frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\dots,M} \prod_{i \in D_j} r_i} \gamma_1,$$

where γ_i is the ith singular value of $\mathcal{X}_{(D_1,\dots,D_M)}$.

Proof Let us consider the Tucker decomposition of $\mathcal{X}$ as

$$\mathcal{X} = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \cdots \sum_{j_K=1}^{r_K} \mathcal{C}_{j_1,\dots,j_K} u_{j_1}^{(1)} \otimes u_{j_2}^{(2)} \otimes \cdots \otimes u_{j_K}^{(K)},$$

where $\mathcal{C} \in \mathbb{R}^{r_1 \times \cdots \times r_K}$ is the core tensor, $(r_1, \dots, r_K)$ is the multilinear rank and $u_i^j \in \mathbb{R}^{n_j}$, $\|u_i^j\|_2 = 1$, $i = 1, \dots, r_i$, $j = 1, \dots, K$ are component vectors. We rearrange the Tucker decomposition for the reshaped tensor $\mathcal{X}_{(D_1,\dots,D_M)}$ as

$$\mathcal{X}_{(D_1,\dots,D_M)} = \sum_{j'_{a'}, j'_{b'}, \dots \in D_2} \cdots \left(\sum_{\sum_{j''_{a''}, j''_{b''}, \dots \in D_M}} \underbrace{\left(\sum_{j_a, j_b, \dots \in D_1} \mathcal{C}_{j_1,\dots,j_K} \mathit{\Pi}_{D_1}(u_{j_a}^{(a)} \otimes u_{j_b}^{(b)} \cdots) \right)}_{\hat{u}_1[D'_2,\dots,D'_M] \in \mathbb{R}^{\mathrm{prod}(D_1)}} \otimes \mathit{\Pi}_{D_M}(u_{j'_a}^{(a')} \otimes u_{j'_b}^{(b')} \cdots) \right) \otimes \cdots .$$

Taking $\hat{u}_1[D'_2, \dots, D'_M] = \gamma[D'_2, \dots, D'_M] \frac{\hat{u}_1[D'_2,\dots,D'_M]}{\|\hat{u}_1[D'_2,\dots,D'_M]\|_2}$ where $\gamma[D'_2, \dots, D'_M] = \|\hat{u}_1[D'_2, \dots, D'_M]\|_2$. We can consider the about summation over any reshaping set and it is easy to see that the arrangement takes a CP decomposition with a CP rank of $r_{cp} = \frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\dots,M} \prod_{i \in D_j} r_i}$.

By arranging $\gamma[D_2, \dots, D_M]$ in descending order order along with component vectors $\hat{u}^{(1)}[D_2, \dots, D_M]$ and renaming them as $\gamma_1 \geq \gamma_2 \geq \dots$ and u_{1j}, respectively, we obtain

$$\|\mathcal{X}_{(D_1,\dots,D_M)}\|_\star = \left\{ \sum_{j=1}^{r_{cp}} \gamma_j | \mathcal{X}_{(D_1,\dots,D_M)} = \sum_{j=1}^{r_{cp}} \gamma_j u_{1j} \otimes u_{2j} \cdots \otimes u_{Mj}, \|u_{kj}\|_2^2 = 1, \gamma_j \geq \gamma_{j+1} > 0 \right\},$$

where $u_{kj} \in [\mathit{\Pi}_{D_k}(u_1^{(a')} \otimes u_1^{(b')} \cdots), \dots, \mathit{\Pi}_{D_k}(u_{r'_a}^{(a')} \otimes u_{r'_b}^{(b')} \cdots)]$ are components for each $k = 2, \dots, M$ and $a', b', \dots \in D_k$.

The following inequality is trivial,

$$\|\mathcal{X}_{(D_1,\ldots,D_M)}\|_\star = \Bigg\{ \sum_{j=1}^{r_{cp}} \gamma_j | \mathcal{X}_{(D_1,\ldots,D_M)} = \sum_{j=1}^{r_{cp}} \gamma_j u_{1j} \otimes u_{2j} \cdots \otimes u_{Mj},$$
$$\|u_{kj}\|_2^2 = 1, \gamma_j \geq \gamma_{j+1} > 0 \Bigg\} \leq \frac{\prod_{k=1}^K r_k}{\max_{j=1,\ldots,M} \prod_{i \in D_j} r_i} \gamma_1,$$

□

Lemma 4 *Let $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ be a random K-mode tensor with CP rank of r_{cp}. We consider a set of M reshaping sets D_i, $i = 1, \ldots, M$. Then*

$$\|\mathcal{X}_{(D_1,\ldots,D_M)}\|_\star \leq r_{cp}\gamma_1,$$

where γ_i is the ith singular value of $\mathcal{X}_{(D_1,\ldots,D_M)}$.

Proof Let us consider $\mathcal{X}$ as

$$\mathcal{X} = \sum_{j=1}^{r_{cp}} \gamma_j u_{1j} \otimes u_{2j} \cdots \otimes u_{Kj},$$

with $\|u_{kj}\|_2^2 = 1, \gamma_j \geq \gamma_{j+1} > 0$. For the reshaping set $(D_1, \ldots, D_M)$, we rearrange the $\mathcal{X}$ as

$$\mathcal{X}_{(D_1,\ldots,D_M)} = \sum_{j=1}^{r_{cp}} \gamma_j (\circ_{i_1 \in D_1} u_{i_1 j}) \otimes (\circ_{i_2 \in D_2} u_{i_2 j}) \cdots \otimes (\circ_{i_M \in D_M} u_{i_M j}),$$

where $a \circ b = [a_1 b, a_2 b, \ldots, a_n b]^\top$ is the Khatri-Rao product [12]. It is easy to verify that $\text{vec}((a \circ b) \otimes (c \circ d)) = \text{vec}(a \otimes b \otimes c \otimes d)$, which indicates that $\text{vec}(\mathcal{X}) = \text{vec}(\mathcal{X}_{(D_1,\ldots,D_M)})$.

Using the fact that $\text{Rank}(a \otimes b) \leq \text{Rank}(a)\text{Rank}(b)$ from [12], we have

$$\text{Rank}(\mathcal{X}_{(D_1,\ldots,D_M)}) \leq \text{Rank}(\mathcal{X}) = r_{cp}.$$

This lead to the final observation

$$\|\mathcal{X}_{(D_1,\ldots,D_M)}\|_\star \leq r_{cp}\gamma_1.$$

□

In order to prove Rademacher complexities in Theorem 1 and 2, we use the following Lemma form [19].

Lemma 5 (Raskutti, Chen and Yuan,2015) *Consider a K-mode tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ with random samples from an i.i.d. Gaussian tensor ensemble. Then*

$$\mathbb{E}\|\mathcal{X}\|_{\text{op}} \leq 4\log(4K) \sum_{k=1}^K \sqrt{n_k}.$$

Given a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ with Gaussian entries, we can write

$$\mathbb{E}\mathcal{X} = \mathbb{E} \sum_{i_1,i_2,\ldots,i_K} \mathcal{X}_{i_1,i_2,\ldots,i_K} e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_K},$$

where e_{i_k} is the vector with 1 at the kth element and rest of the elements are zero. Due to each $\mathcal{X}_{i_1,i_2,\ldots,i_K}$ being a Gaussian entry, we have

$$\mathbb{E}\mathcal{X} = \mathbb{E}_g \mathbb{E}_\epsilon \sum_{i_1,i_2,\ldots,i_K} \epsilon_{i_1,i_2,\ldots,i_K} |\mathcal{X}_{i_1,i_2,\ldots,i_K}| e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_K},$$

where $\epsilon_{i_1,i_2,\ldots,i_K} \in \{-1,1\}$. Using the Jensen's inequality, we have

$$\begin{aligned}
&\mathbb{E}_g\mathbb{E}_\epsilon \sum_{i_1,i_2,\ldots,i_K} \epsilon_{i_1,i_2,\ldots,i_K}|\mathcal{X}_{i_1,i_2,\ldots,i_K}| e_{i_1}\otimes e_{i_2}\otimes\cdots e_{i_K}\\
&\geq \mathbb{E}_\epsilon \sum_{i_1,i_2,\ldots,i_K} \epsilon_{i_1,i_2,\ldots,i_K}\mathbb{E}_g|\mathcal{X}_{i_1,i_2,\ldots,i_K}| e_{i_1}\otimes e_{i_2}\otimes\cdots e_{i_K}\\
&\geq \sqrt{2\pi}\mathbb{E}_\epsilon \sum_{i_1,i_2,\ldots,i_K} \epsilon_{i_1,i_2,\ldots,i_K} e_{i_1}\otimes e_{i_2}\otimes\cdots e_{i_K}.
\end{aligned}$$

This shows that we can use the Lemma 3 to bound tensors with Bernoulli random variables.

Next we give the detailed proof of Theorem 1.

Proof of Theorem 1: We expand the Rademacher complexity in (5) as

$$R_{\mathrm{S}}(l\circ\mathcal{W}) = \frac{1}{|\mathrm{S}|}\mathbb{E}_\sigma\left[\sup_{\mathcal{W}\in\mathsf{W}}\sum_{i_1,..,i_K}\Sigma_{i_1,\ldots,i_K} l(\mathcal{X}_{i_1,\ldots,i_K},\mathcal{W}_{i_1,\ldots,i_K})\right],$$

where $\Sigma_{i_1,\ldots,i_K} = \sigma_j$ when $(i_1,\ldots,i_K) = \sigma_j \in \mathrm{S}$ and $\Sigma_{i_1,\ldots,i_K} = 0$, otherwise.

We analyze the Rademacher complexity

$$\begin{aligned}
R_{\mathrm{S}}(l\circ\mathcal{W}) &= \frac{1}{|\mathrm{S}|}\mathbb{E}_\sigma\left[\sup_{\mathcal{W}\in\mathsf{W}}\sum_{i_1,..,i_K}\Sigma_{i_1,\ldots,i_K} l(\mathcal{X}_{i_1,\ldots,i_K},\mathcal{W}_{i_1,\ldots,i_K})\right],\\
&\leq \frac{\Lambda}{|\mathrm{S}|}\mathbb{E}_\sigma\left[\sup_{\mathcal{W}\in\mathsf{W}}\sum_{i_1,..,i_K}\Sigma_{i_1,\ldots,i_K}\mathcal{W}_{i_1,\ldots,i_K}\right], \qquad \text{(Rademacher contraction)} \quad (9)\\
&\leq \frac{\Lambda}{|\mathrm{S}|}\mathbb{E}_\sigma\sup_{\mathcal{W}\in\mathsf{W}}\|\mathcal{W}_{(D_1,\ldots,D_M)}\|_\star\|\Sigma_{(D_1,\ldots,D_M)}\|_{\star^*}, \quad \text{(Duality relationship)}
\end{aligned}$$

(a) Given that tensor has a multilinear rank of $(r_1,\ldots,r_K)$, using the Lemma 3, we know that

$$\|\mathcal{W}_{(D_1,\ldots,D_M)}\|_\star \leq \left(\frac{\prod_{k=1}^K r_k}{\max_{j=1,\ldots,M}\prod_{i\in D_j} r_i}\right)\gamma_1(\mathcal{W}_{(D_1,\ldots,D_M)}). \tag{10}$$

Using Lemma 5, we can bound $\mathbb{E}_\sigma\|\Sigma_{(D_1,\ldots,D_M)}\|_{\star^*}$ as

$$\mathbb{E}_\sigma\|\Sigma_{(D_1,\ldots,D_M)}\|_{\star^*} \leq 4\log(4M)\sum_{j=1}^M\sqrt{\prod_{p\in D_j} n_p}. \tag{11}$$

By substituting (10) and (11) to (9), we obtain the following bound

$$R_{\mathrm{S}}(l\circ\mathcal{W}) \leq \frac{c\Lambda}{|\mathrm{S}|}\left(\frac{\prod_{k=1}^K r_k}{\max_{j=1,\ldots,M}\prod_{i\in D_j} r_i}\right)\gamma_1(\mathcal{W}_{(D_1,\ldots,D_M)})\log(4M)\sum_{j=1}^M\sqrt{\prod_{p\in D_j} n_p}. \tag{12}$$

(b) Given that tensor has a CP rank of r_{cp}, using the Lemma 4, we have

$$\|\mathcal{W}_{(D_1,\ldots,D_M)}\|_\star \leq r_{cp}\gamma_1(\mathcal{W}_{(D_1,\ldots,D_M)}). \tag{13}$$

From Lemma 5, we have

$$\mathbb{E}_\sigma\|\Sigma_{(D_1,\ldots,D_M)}\|_{\star^*} \leq 4\log(4M)\sum_{j=1}^M\sqrt{\prod_{p\in D_j} n_p}. \tag{14}$$

By substituting (13) and (14) to (9), we obtain the desired bound

$$R_{\mathrm{S}}(l \circ \mathcal{W}) \leq \frac{c\Lambda}{|\mathrm{S}|} r_{cp} \gamma_1(\mathcal{W}_{(D_1,\ldots,D_M)}) \log(4M) \sum_{j=1}^{M} \sqrt{\prod_{p \in D_j} n^{|D_j|}}. \tag{15}$$

□

Next, we give the proof for theorem 2.

Proof of theorem 2: We expand the Rademacher complexity in (5) using latent tensors $\mathcal{W}^{(1)},\ldots,\mathcal{W}^{(G)}$ for the reshaped latent tensor nuclear norm as

$$R_{\mathrm{S}}(l \circ (\mathcal{W}^{(1)} + \cdots + \mathcal{W}^{(G)})) = \frac{1}{|\mathrm{S}|}\mathbb{E}_\sigma \Bigg[\sup_{\mathcal{W}^{(1)}+\cdots+\mathcal{W}^{(G)}=\mathcal{W}\in \mathsf{W}_{\mathrm{rl}}} \sum_{i_1,\ldots,i_K} \Sigma_{i_1,\ldots,i_K} l(\mathcal{X}_{i_1,\ldots,i_K}, \mathcal{W}_{i_1,\ldots,i_K}) \Bigg],$$

where $\Sigma_{i_1,\ldots,i_K} = \sigma_j$ when $(i_1,\ldots,i_K) = \sigma_j \in \mathrm{S}$ and $\Sigma_{i_1,\ldots,i_K} = 0$, otherwise.

We can analyze the Rademacher complexity as

$$\begin{aligned} R_{\mathrm{S}}(l \circ (\mathcal{W}^{(1)} + \cdots + \mathcal{W}^{(G)})) &= \frac{1}{|\mathrm{S}|}\mathbb{E}_\sigma \Bigg[\sup_{\mathcal{W}^{(1)}+\cdots+\mathcal{W}^{(G)}=\mathcal{W}\in \mathsf{W}_{\mathrm{rl}}} \sum_{i_1,\ldots,i_K} \Sigma_{i_1,\ldots,i_K} l(\mathcal{X}_{i_1,\ldots,i_K}, \mathcal{W}_{i_1,\ldots,i_K}) \Bigg], \\ &\leq \frac{\Lambda}{|\mathrm{S}|}\mathbb{E}_\sigma \Bigg[\sup_{\mathcal{W}^{(1)}+\cdots+\mathcal{W}^{(G)}=\mathcal{W}\in \mathsf{W}_{\mathrm{rl}}} \sum_{i_1,\ldots,i_K} \Sigma_{i_1,\ldots,i_K} \mathcal{W}_{i_1,\ldots,i_K} \Bigg], \\ &\qquad \text{(Rademacher contraction)} \\ &\leq \frac{\Lambda}{|\mathrm{S}|}\mathbb{E}_\sigma \sup_{\mathcal{W}^{(1)}+\cdots+\mathcal{W}^{(G)}=\mathcal{W}\in \mathsf{W}_{\mathrm{rl}}} \|\mathcal{W}\|_{\mathrm{r\_latent}} \|\Sigma\|_{\mathrm{r\_latent}^*} \\ &\qquad \text{(Duality relationship).} \end{aligned} \tag{16}$$

(a) For tensor with multilinear rank, using Lemma 4, we obtain

$$\begin{aligned} \|\mathcal{W}\|_{\mathrm{r\_latent}} &= \inf_{\mathcal{W}^{(1)}+\cdots+\mathcal{W}^{(G)}=\mathcal{W}} \sum_{g=1}^{G} \|\mathcal{W}^{(k)}_{(D_1^{(g)},\ldots,D_{m_g}^{(g)})}\|_\star \\ &\leq \min_{g\in G} \left(\frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\ldots,M} \prod_{i\in D_j^{(g)}} r_i} \right) \gamma_1(\mathcal{W}^{(g)}_{(D_1^{(g)},\ldots,D_{M_g}^{(g)})}). \end{aligned} \tag{17}$$

Using Lemma 1 we can bound $\mathbb{E}_\sigma \|\Sigma\|_{\mathrm{r\_latent}^*}$ as

$$\mathbb{E}_\sigma \|\Sigma\|_{\mathrm{r\_latent}^*} = \max_{g\in G} \left\| \mathcal{W}^{(g)}_{(D_1^{(g)},\ldots,D_{M_g}^{(g)})} \right\|_\star \leq 4 \max_{g\in G} \log(4M_g) \sum_{j=1}^{M_g} \sqrt{\prod_{p\in D_j^{(g)}} n_p}. \tag{18}$$

By substituting (17) and (18) to (16), we obtain the following bound

$$\begin{aligned} R_{\mathrm{S}}(l \circ \mathcal{W}) \leq \frac{c\Lambda}{|\mathrm{S}|} \min_{g\in G} &\left(\frac{\prod_{k=1}^{K} r_k}{\max_{j=1,\ldots,M} \prod_{i\in D_j^{(g)}} r_i} \right) \gamma_1(\mathcal{W}^{(g)}_{(D_1^{(g)},\ldots,D_{M_g}^{(g)})}) \\ &\max_{g\in G} \log(4M_g) \sum_{j=1}^{M_g} \sqrt{\prod_{p\in D_j^{(g)}} n_p}. \end{aligned}$$

(b) For tensor with CP rank, using Lemma 4, we obtain

$$\|\mathcal{W}\|_{\mathrm{r\_latent}} = \inf_{\mathcal{W}^{(1)}+\cdots+\mathcal{W}^{(G)}=\mathcal{W}} \sum_{g=1}^{G} \|\mathcal{W}^{(k)}_{(D_1^{(g)},\ldots,D_{m_g}^{(g)})}\|_\star \leq \min_{g\in G} r_{cp}\gamma_1(\mathcal{W}^{(g)}_{(D_1^{(g)},\ldots,D_{M_g}^{(g)})}). \tag{19}$$

By substituting (19) and (18) to (16), we obtain the following bound

$$R_{\mathrm{S}}(l\circ\mathcal{W}) \leq \frac{c\Lambda}{|\mathrm{S}|} \min_{g\in G} r_{cp}\gamma_1(\mathcal{W}^{(g)}_{(D_1^{(g)},\ldots,D_{M_g}^{(g)})}) \max_{g\in G}\log(4M_g) \sum_{j=1}^{M_g} \sqrt{\prod_{p\in D_j^{(g)}} n_p}.$$

□

Finally, we derive the Rademacher complexity for the tensor completion model regularized by the Schatten TT norm.

Theorem 3 *Consider a K-mode tensor $\mathcal{W}\in\mathbb{R}^{n_1\times\cdots\times n_K}$ with a multilinear rank of $(r_1,\ldots,r_K)$. Let us consider the hypothesis class $\mathsf{W}_{\mathrm{TT}} = \{\mathcal{W} | \|\mathcal{T}\|_{s,T}\leq t\}$. Then Rademacher complexity is bounded as*

$$R_{\mathrm{S}}(l\circ\mathcal{W}) \leq \frac{c'\Lambda}{|\mathrm{S}|}\sum_{k=1}^{K-1}\min\left(\prod_{i=1}^{k}\sqrt{r_i}, \prod_{j=k+1}^{K}\sqrt{r_j}\right) B_{\mathcal{T}} \min_{k=1,\ldots,K-1}\left(\sqrt{\prod_{i<k} n_i} + \sqrt{\prod_{j\geq k}^{K} n_j}\right), \tag{20}$$

where $\|\mathcal{W}\|_{\mathrm{F}}\leq B_{\mathcal{T}}$ and c' is a constant.

Proof For this case we consider the hypothesis class $\mathcal{W}_{\mathrm{TT}}$ for the Rademacher complexity follows as

$$R_{\mathrm{S}}(l\circ\mathcal{W}) = \frac{1}{|\mathrm{S}|}\mathbb{E}_\sigma\left[\sup_{\mathcal{W}_{\mathrm{TT}}\in\mathsf{W}} \sum_{i_1,\ldots,i_K} \Sigma_{i_1,\ldots,i_K} l(\mathcal{X}_{i_1,\ldots,i_K}, \mathcal{W}_{i_1,\ldots,i_K})\right],$$

where $\Sigma_{i_1,\ldots,i_K} = \sigma_j$ when $(i_1,\ldots,i_K) = \sigma_j \in \mathrm{S}$ and $\Sigma_{i_1,\ldots,i_K} = 0$, otherwise.

Now we analyze the Rademacher complexity for the hypothesis class $\mathcal{W}_{\mathrm{TT}}$. We have

$$\begin{aligned} R_{\mathrm{S}}(l\circ\mathcal{W}) &= \frac{1}{|\mathrm{S}|}\mathbb{E}_\sigma\left[\sup_{\mathcal{W}\in\mathsf{W}_{\mathrm{TT}}} \sum_{i_1,\ldots,i_K} \Sigma_{i_1,\ldots,i_K} l(\mathcal{X}_{i_1,\ldots,i_K}, \mathcal{W}_{i_1,\ldots,i_K})\right], \\ &\leq \frac{\Lambda}{|\mathrm{S}|}\mathbb{E}_\sigma\left[\sup_{\mathcal{W}\in\mathsf{W}_{\mathrm{TT}}} \sum_{i_1,\ldots,i_K} \Sigma_{i_1,\ldots,i_K} \mathcal{W}_{i_1,\ldots,i_K}\right], \qquad \text{(Rademacher contraction)} \\ &\leq \frac{\Lambda}{|\mathrm{S}|}\mathbb{E}_\sigma \sup_{\mathcal{W}\in\mathsf{W}_{\mathrm{TT}}} \|\mathcal{W}\|_{s,T}\|\Sigma\|_{s,T^*}, \qquad \text{(Duality relationship)} \end{aligned} \tag{21}$$

where $\|\cdot\|_{s,T^*}$ is the dual norm of $\|\cdot\|_{s,T}$. The last step can be obtained by applying the Holder's inequality to the sum of trace norms in the Schatten TT norm.

Considering $\|\mathcal{W}\|_{s,T}$, we can expand it as

$$\|\mathcal{W}\|_{s,T} = \frac{1}{K-1}\sum_{k=1}^{K-1}\|Q_k(\mathcal{T})\|_{\mathrm{tr}} = \frac{1}{K-1}\sum_{k=1}^{K-1}\sum_{i_k=1}^{\hat{r}_k}\gamma_{i_k}(Q_k(\mathcal{T})),$$

where $Q_k : \mathcal{T}\to\mathbb{R}^{n_{\geq k}\times n_{k<}}$ is a reshaping operator, and $\gamma_{i_k}()$ and $\hat{r}_k$ are the i_kth singular value and the rank of the reshaped tensor by Q_k, respectively. Using the Cauchy-Schwarz inequality, we have

$$\|\mathcal{W}\|_{s,T} \leq \frac{1}{K-1}\sum_{k=1}^{K-1}\sqrt{\hat{r}_k}\sqrt{\sum_{i_k=1}^{\hat{r}_k}\gamma_{i_k}^2(Q_k(\mathcal{T}))} = \frac{1}{K-1}\sum_{k=1}^{K-1}\sqrt{\hat{r}_k}B_{\mathcal{T}},$$

where $\|\mathcal{T}\|_{\mathrm{F}} = B_{\mathcal{T}}$. Using Lemmas 1 and 2, we can infer that

$$\|\mathcal{W}\|_{s,T} \leq \frac{1}{K-1} \sum_{k=1}^{K-1} \min\left(\prod_{i=1}^{k} \sqrt{r_i}, \prod_{j=k+1}^{K} \sqrt{r_j} \right) B_{\mathcal{T}}, \tag{22}$$

Similar to the overlapped trace norm [22], the Schatten TT norm also sums nuclear norms of the the same tensors reshaped into different matrices. Hence, we can extend the dual norm of the overlapped trace norm in [22] to the Schatten TT norm. Using [22], it is easy to the dual norm of Schatten TT norm as

$$\|\Sigma\|_{s,T^*} = \inf_{\Sigma^{(1)}+\ldots+\Sigma^{(K)}=\Sigma} \sum_{k=1}^{K-1} \|Q_k(\Sigma^{(k)})\|_{\mathrm{op}}.$$

We want to bound

$$\mathbb{E}\|\Sigma\|_{s,T^*} = \mathbb{E} \inf_{\Sigma^{(1)}+\ldots+\Sigma^{(K)}=\Sigma} \sum_{k=1}^{K-1} \|Q_k(\Sigma^{(k)})\|_{\mathrm{op}},$$

and since we can take any of $\Sigma^{(k)},\ k = 1, \ldots, K$ to be equal to Σ, we have

$$\mathbb{E}\|\Sigma\|_{s,T^*} \leq \min_{k=1,\ldots,K-1} \|Q_k(\Sigma)\|_{\mathrm{op}}.$$

We apply Latała's Theorem [13, 20] for the reshaping by the Q_k operator and bound $\mathbb{E}\|Q_k(\varSigma)\|_{\mathrm{op}}$ as

$$\mathbb{E}\|Q_k(\varSigma)\|_{\mathrm{op}} \leq C_1 \left(\sqrt{\prod_{i<k} n_i} + \sqrt{\prod_{j \geq k}^{K} n_j} + \sqrt[4]{|Q_k(\varSigma)|} \right),$$

and since $\sqrt[4]{|Q_k(\varSigma)|} \leq \sqrt[4]{\prod_{i=1}^{K} n_i} \leq \frac{1}{2}\left(\sqrt{\prod_{i<k} n_i} + \sqrt{\prod_{j \geq k}^{K} n_j} \right)$, we have,

$$\mathbb{E}\|Q_k(\varSigma)\|_{\mathrm{op}} \leq \frac{3C_1}{2} \left(\sqrt{\prod_{i<k} n_i} + \sqrt{\prod_{j \geq k}^{K} n_j} \right).$$

This gives us the bounds for $\mathbb{E}\|\Sigma\|_{s,T^*}$ as

$$\mathbb{E}\|\Sigma\|_{s,T^*} \leq \min_{k=1,\ldots,K-1} \frac{3C_1}{2} \left(\sqrt{\prod_{i<k} n_i} + \sqrt{\prod_{j \geq k}^{K} n_j} \right). \tag{23}$$

By combining (22) and (23) with (21), we obtain

$$R_{\mathrm{S}}(l \circ \mathcal{W}) \leq \frac{c'\varLambda}{|\mathrm{S}|(K-1)} \sum_{k=1}^{K-1} \min\left(\prod_{i=1}^{k} \sqrt{r_i}, \prod_{j=k+1}^{K} \sqrt{r_j} \right) B_{\mathcal{T}} \min_k \left(\sqrt{\prod_{i<k} n_i} + \sqrt{\prod_{j \geq k}^{K} n_j} \right). \tag{24}$$

□