Dear Author:

Attached you will find a pdf of the proofs of your article scheduled to appear in a forthcoming issue of *Neural Computation*.

Please print out the pdf of your proof on standard size paper (8 ½ x 11) single-sided, and check for accuracy and consistency, making especially sure to check the accuracy of material that only you can verify (such as numerical data or spelling of proper names). Please also check all figures on the proofs for orientation, legibility, and overall quality and initial your approval next to all figures. Throughout, please limit your changes to those necessary to correct errors or inconsistencies. Major changes to the manuscript or replacement of figures may not be made. Note that changes resulting in repagination of the issue will not be accepted. Please do not send a revision of your article and do not alter the pdf that was sent to you. If you are unable to get the corrections back to us within the requested time frame, we can not guarantee that your changes will be incorporated.

Please mark corrections in red pen directly on the proofs. Please also include a typed list detailing changes. If each figure is approved please indicate this on the list. If you are not approving the figures and new figures have been sent, please indicate why. Please send only one set of corrections on one master copy and one list of changes, even if there is more than one author making changes.

**Within three days of receipt** of the email notification, please return your corrections by e-mail (preferred) or fax to the contact below.

Please be sure to provide the following:

- A list containing all changes, including figure approval.
- A copy of every page and figure where corrections are marked.
- If, because of error in placement or reproduction, adjustments to figures are required or new figures are sent, include on the list of changes an explanation of why this is necessary.

You may choose to return your corrections in hardcopy form via an overnight delivery service but this is not required.

**Contact Information:**
Eric Witz
MIT Press Journals
One Rogers Street
Cambridge, MA 02142-1209
Email: ewitz@mit.edu
Phone: (617) 258-0586
Fax: (617) 812-6363

Thank you for your cooperation.

Yours sincerely,

Eric Witz
Senior Production Editor
MIT Press Journals
Email: ewitz@mit.edu

*LETTER* Communicated by Evrim Acar

# Convex Coupled Matrix and Tensor Completion

**Kishan Wimalawarne**
*kishanwn@gmail.com*
[1]*Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji 611-004, Japan*

**Makoto Yamada**
*makoto.yamada@riken.jp*
[2]*RIKEN, Center for Advanced Intelligence Project, Chuo-ku, Tokyo 103-0027, Japan, Institute of Statistical Mathematics, Tachikawa, Tokyo 190-8562, Japan, and PRESTO, Japan Science and Technological Agency, Kawaguchi-shi, Saitama 332-0012, Japan*

**Hiroshi Mamitsuka**
*mami@kuicr.kyoto-u.ac.jp*
[3]*Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji 611-0011, Japan, and Department of Computer Science, Aalto University, Espoo 02150 Finland*



**We propose a set of convex low-rank inducing norms for coupled matrices and tensors (hereafter referred to as coupled tensors), in which information is shared between the matrices and tensors through common modes. More specifically, we first propose a mixture of the overlapped trace norm and the latent norms with the matrix trace norm, and then, propose a completion model regularized using these norms to impute coupled tensors. A key advantage of the proposed norms is that they are convex and can be used to find a globally optimal solution, whereas existing methods for coupled learning are nonconvex. We also analyze the excess risk bounds of the completion model regularized using our proposed norms and show that they can exploit the low-rankness of coupled tensors, leading to better bounds compared to those obtained using uncoupled norms. Through synthetic and real-data experiments, we show that the proposed completion model compares favorably with existing ones.**

## 1 Introduction

Learning from a matrix or a tensor has long been an important problem in machine learning. In particular, matrix and tensor factorization using low-rank inducing norms has been studied extensively, and many applications have been considered, such as missing value imputation (Signoretto,

*Neural Computation* **30, 1–33** (2018)
doi:10.1162/neco_a_01123

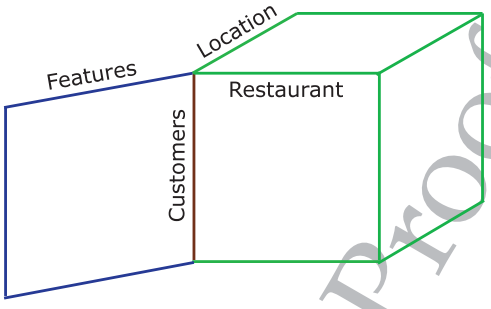



Figure 1: Illustration of information sharing between matrix and tensor in coupled tensor, through customers mode.



Dinh, De Lathauwer, & Suykens, 2013; Liu, Musialski, Wonka, & Ye, 2009), multitask learning (Argyriou, Evgeniou, & Pontil, 2006; Romera-Paredes, Aung, Bianchi-Berthouze, & Pontil, 2013; Wimalawarne, Sugiyama, & Tomioka, 2014), subspace clustering (Liu, Lin, & Yu, 2010), and inductive learning (Signoretto et al., 2013; Wimalawarne, Tomioka, & Sugiyama, 2016). Though useful in many applications, factorization based on an individual matrix or tensor tends to perform poorly under the cold start setup condition (Singh & Gordon, 2008), when, for example, it is not possible to observe click information for new users in collaborative filtering. It therefore cannot be used to recommend possible items for new users. Potential ways to address this issue are matrix or tensor factorization with side information (Narita, Hayashi, Tomioka, & Kashima, 2011). Both have been applied to recommendation systems (Singh & Gordon, 2008; Gunasekar, Yamada, Yin, & Chang, 2015) and personalized medicine (Khan & Kaski, 2014).

Both matrix and tensor factorization with side information can be regarded as the joint factorization of coupled matrices and tensors (hereafter referred to as coupled tensors; see Figure 1). Acar, Kolda, and Dunlavy (2011) introduced a coupled factorization method based on CANDECOMP/PARAFAC (CP) decomposition that simultaneously factorizes matrices and tensors by sharing the low-rank structures in the matrices and tensors. The coupled factorization approach has been applied to joint analysis of fluorescence and proton nuclear magnetic resonance (NMR) measurements (Acar, Nilsson, & Saunders, 2014) and joint NMR and liquid chromatography-mass spectrometry (LCMS; Acar, Bro, and Smilde, 2015). More recently, a Bayesian approach proposed by Ermis, Acar, and Cemgil (2015) was applied to link prediction problems. However, existing coupled factorization methods are nonconvex and can obtain only a poor local optimum. Moreover, the ranks of the coupled tensors need to be determined beforehand. In practice, it is difficult to specify the true ranks of the tensor and the matrix without prior knowledge. Furthermore, existing algorithms are not theoretically guaranteed.

We propose in this letter convex norms for coupled tensors that overcome the nonconvexity problem. The norms are a mixture of tensor norms: the overlapped trace norm (Tomioka, Suzuki, Hayashi, & Kashima, 2011), the latent trace norm (Tomioka & Suzuki, 2013), the scaled latent norm (Wimalawarne et al., 2014), and the matrix trace norm (Argyriou et al., 2006). A key advantage of the proposed norms is that they are convex and thus can be used to find a globally optimal solution, whereas existing coupled factorization approaches are nonconvex. Furthermore, we analyze the excess risk bounds of the completion model regularized using our proposed norms. Through synthetic and real-data experiments, we show that it compares favorably with existing ones.

In this letter, we:

- Propose a set of convex coupled norms for matrices and tensors that extend low-rank tensor and matrix norms.
- Propose mixed norms that combine features from both the overlapped norm and latent norms.
- Propose a convex completion model regularized using the proposed coupled norms.
- Analyze the excess risk bounds for the proposed completion model with respect to the proposed norms and show that it leads to lower excess risk.
- Show through synthetic and real-data experiments that our norms lead to performance comparable to that of existing nonconvex methods.
- Show that our norms are applicable to coupled tensors based on both the CP rank and the multilinear rank without prior assumptions about their low-rankness.
- Show that the convexity of the proposed norms leads to global solutions, eliminating the need to deal with local optimal solutions as is necessary with nonconvex methods.

The remainder of the letter is organized as follows. In section 2, we discuss related work on coupled tensor completion. In section 3, we present our proposed method, first introducing a coupled completion model and then proposing a set of norms called coupled norms. In section 4, we give optimization methods for solving the coupled completion model. In section 5, we theoretically analyze it using excess risk bounds for the proposed coupled norms. In section 6, we present the results of our evaluation using synthetic and real-world data experiments. Finally, in section 7, we summarize the key points and suggest future work.

## 2 Related Work

Most of the models proposed for learning with multiple matrices or tensors use joint factorization of matrices and tensors. The regularization-based

model proposed by Acar et al. (2011) for completion of coupled tensors, which was further studied (Acar, Nilsson et al., 2014; Acar, Papalexakis et al., 2014; Acar et al., 2015) uses CP decomposition (Carroll & Chang, 1970; Harshman, 1970; Hitchcock, 1927; Kolda & Bader, 2009) to factorize the tensor and operates under the assumption that the factorized components of its coupled mode are in common with the factorized components of the matrix on the same mode. Bayesian models have also been proposed for imputing missing values with applications in link prediction (Ermis et al., 2015) and nonnegative factorization (Takeuchi, Tomioka, Ishiguro, Kimura, & Sawada, 2013), which use similar factorization models. Applications that have used collective factorization of tensors are multiview factorization (Khan & Kaski, 2014) and multiway clustering (Banerjee, Basu, & Merugu, 2007). Due to their use of factorization-based learning, all of these models are nonconvex.

The use of common adjacency graphs has more recently been proposed for incorporating similarities among heterogeneous tensor data (Li, Zhao, Li, Cichocki, & Guo, 2015). Though this method does not require assumptions about rank for explicit factorization of tensors, it depends on the modeling of the common adjacency graph and does not incorporate the low-rankness created by the coupling of tensors.

## 3 Proposed Method

We investigate a method for coupling a matrix and a tensor that forms when they share a common mode (Acar et al., 2015; Acar, Nilsson et al., 2014; Acar, Papalexakis, 2014). An example of the most basic coupling is shown in Figure 1, where a three-way (third-order) tensor is attached to a matrix on a specific mode. As depicted, we may have a problem predicting recommendations for customers on the basis of their preferences of restaurants in different locations, and we may also have side information about the characteristics for each customer. We can utilize this side information by coupling the customer-characteristic matrix with the sparse customer-restaurant-location tensor of the customer mode and then impute the missing values in the tensor.

Let us consider a partially observed matrix $\hat{M} \in \mathbb{R}^{n_1 \times m}$ and a partially observed three-way tensor $\hat{\mathcal{T}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with mappings to observed elements indexed by $\Omega_M$ and $\Omega_{\mathcal{T}}$, respectively, and let us assume that they are coupled on the first mode. Our ultimate goal of this letter is to introduce convex coupled norms $\|\mathcal{T}, M\|_{\mathrm{cn}}$ for use in solving

$$\min_{\mathcal{T},M} \frac{1}{2}\|\Omega_M(M - \hat{M})\|_{\mathrm{F}}^2 + \frac{1}{2}\|\Omega_{\mathcal{T}}(\mathcal{T} - \hat{\mathcal{T}})\|_{\mathrm{F}}^2 + \lambda\|\mathcal{T}, M\|_{\mathrm{cn}}, \tag{3.1}$$

where $\lambda \geq 0$ is the regularization parameter. We also investigate the theoretical properties of problem 3.1.

The mode-$k$ unfolding of tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ is represented as $T_{(k)} \in \mathbb{R}^{n_k \times \prod_{j \neq k}^{K} n_j}$, which is obtained by concatenating all the $\prod_{j \neq k}^{K} n_j$ vectors with dimension $n_k$ obtained by fixing all except the $k$th index on mode-$k$ along its columns. We use $vec()$ to indicate the conversion of a matrix or a tensor into a vector and $unvec()$ to represent the reverse operation. The spectral norm (operator norm) of a matrix $X$ is the $\|X\|_{\text{op}}$ that is the largest singular value of $X$. The Frobenius norm of a tensor $\mathcal{T}$ is defined as $\|\mathcal{T}\|_{\text{F}} = \sqrt{\langle \mathcal{T}, \mathcal{T} \rangle} = \sqrt{vec(\mathcal{T})^\top vec(\mathcal{T})}$. We use $[M; N]$ as the concatenation of matrices $M \in \mathbb{R}^{m_1 \times m_2}$ and $N \in \mathbb{R}^{m_1 \times m_3}$ along their mode 1.

**3.1 Existing Matrix and Tensor Norms.** Before we introduce our new norms, we first briefly review the existing low-rank inducing matrix and tensor norms. Among matrices, the matrix trace norm (Argyriou et al., 2006) is a commonly used convex relaxation for the minimization of the rank of a matrix. For a given matrix $M \in \mathbb{R}^{n_1 \times m}$ with rank $J$, we can define its trace norm as

$$\|M\|_{\text{tr}} = \sum_{j=1}^{J} \sigma_j,$$

where $\sigma_j$ is the $j$th nonzero singular value of the matrix.

Low-rank inducing norms for tensors have received revived attention in recent years. One of the earliest low-rank inducing tensor norm is the tensor nuclear norm (Liu et al., 2009), also known as the overlapped trace norm (Tomioka & Suzuki, 2013) which can be expressed for a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ as

$$\|\mathcal{T}\|_{\text{overlap}} = \sum_{k=1}^{K} \|T_{(k)}\|_{\text{tr}}. \tag{3.2}$$

Tomioka and Suzuki (2013) proposed the latent trace norm:

$$\|\mathcal{T}\|_{\text{latent}} = \inf_{\mathcal{T}^{(1)} + \ldots + \mathcal{T}^{(K)} = \mathcal{T}} \sum_{k=1}^{K} \|T^{(k)}_{(k)}\|_{\text{tr}}. \tag{3.3}$$

The scaled latent trace norm was proposed as an extension of the latent trace norm (Wimalawarne et al., 2014):

$$\|\mathcal{T}\|_{\text{scaled}} = \inf_{\mathcal{T}^{(1)} + \ldots + \mathcal{T}^{(K)} = \mathcal{T}} \sum_{k=1}^{K} \frac{1}{\sqrt{n_k}} \|T^{(k)}_{(k)}\|_{\text{tr}}. \tag{3.4}$$

The behaviors of these two tensor norms have been studied on the basis of multitask learning (Wimalawarne et al., 2014) and inductive learning (Wimalawarne et al., 2016). The results show that for a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ with multilinear rank $(r_1, \ldots, r_K)$, the excess risk is bounded above with respect to regularization with the overlapped trace norm by $\mathcal{O}(\sum_{k=1}^{K} \sqrt{r_k})$, the latent trace norm by $\mathcal{O}(\min_k \sqrt{r_k})$, and the scaled latent trace norm by $\mathcal{O}\Big(\min_k \sqrt{\frac{r_k}{n_k}}\Big)$.

**3.2 Coupled Tensor Norms.** As with individual matrices and tensors, having convex and low-rank inducing norms for coupled tensors would be useful in achieving global solutions for coupled tensor completion with theoretical guarantees. To achieve this, we propose a set of norms for coupled tensors that are coupled on specific modes using existing matrix and tensor trace norms. We first define a new coupled norm with the format $\|.\|^a_{(b,c,d)}$, where the superscript $a$ specifies the mode in which the tensor and matrix are coupled and the subscripts $b, c, d \in \{\mathrm{O}, \mathrm{L}, \mathrm{S}, -\}$ indicate how the modes are regularized. The notations for $b, c, d$ are defined as follows:

O: The mode is regularized with the trace norm. The same tensor is regularized on other modes similar to the overlapped trace norm.
L: The mode is considered to be a latent tensor that is regularized using the trace norm only with respect to that mode.
S: The mode is regularized as a latent tensor, but it is scaled similar to the scaled latent trace norm.
$-$: The mode is not regularized.

Given a matrix $M \in \mathbb{R}^{n_1 \times m}$ and a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we introduce three norms that are coupled extensions of the overlapped trace norm, the latent trace norm, and the scaled latent trace norm, respectively.

*Coupled overlapped trace norm:*

$$\|\mathcal{T}, M\|^1_{(\mathrm{O},\mathrm{O},\mathrm{O})} := \|[T_{(1)}; M]\|_{\mathrm{tr}} + \sum_{k=2}^{3} \|T_{(k)}\|_{\mathrm{tr}}. \tag{3.5}$$

*Coupled latent trace norm:*

$$\|\mathcal{T}, M\|^1_{(\mathrm{L},\mathrm{L},\mathrm{L})} = \inf_{\mathcal{T}^{(1)}+\mathcal{T}^{(2)}+\mathcal{T}^{(3)}=\mathcal{T}} \left( \|[T^{(1)}_{(1)}; M]\|_{\mathrm{tr}} + \sum_{k=2}^{3} \|T^{(k)}_{(k)}\|_{\mathrm{tr}} \right). \tag{3.6}$$

*Coupled scaled latent trace norm:*

$$\|\mathcal{T}, M\|^1_{(\mathrm{S,S,S})} = \inf_{\mathcal{T}^{(1)}+\mathcal{T}^{(2)}+\mathcal{T}^{(3)}=\mathcal{T}} \left( \frac{1}{\sqrt{n_1}} \|[T^{(1)}_{(1)}; M]\|_{\mathrm{tr}} + \sum_{k=2}^{3} \frac{1}{\sqrt{n_k}} \|T^{(k)}_{(k)}\|_{\mathrm{tr}} \right). \tag{3.7}$$

In addition to these norms, we can also create norms as mixtures of overlapped and latent or scaled latent norms. For example, if we want to create a norm that is regularized using the scaled latent trace norm on the second mode while the other modes are regularized using the overlapped trace norm, we can define it as

$$\|\mathcal{T}, M\|^1_{(\mathrm{O,S,O})} = \inf_{\mathcal{T}^{(1)}+\mathcal{T}^{(2)}=\mathcal{T}} \left( \|[T^{(1)}_{(1)}; M]\|_{\mathrm{tr}} + \frac{1}{\sqrt{n_2}} \|T^{(2)}_{(2)}\|_{\mathrm{tr}} + \|T^{(1)}_{(3)}\|_{\mathrm{tr}} \right). \tag{3.8}$$

This norm has two latent tensors, $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$. Tensor $\mathcal{T}^{(1)}$ is regularized using the overlapped method for modes 1 and 3, while the tensor $\mathcal{T}^{(2)}$ is regularized as a scaled latent tensor on mode 2. Given this use of a mixture of regularization methods, we call the resulting norm a *mixed norm*.

In a similar manner, we can create other mixed norms distinguished by their subscripts: (L, O, O), (O, L, O), (O, O, L), (S, O, O), (O, S, O), and (O, O, S). The main advantage gained by using these mixed norms is the additional freedom to regularize low-rank constraints among coupled tensors. Other combinations of norms in which two modes are latent tensors, such as (L, L, O), will make the third mode also a latent tensor since overlapped regularization requires that more than one mode be regularized of the same tensor. Though we have considered using the latent trace norm, in practice it has been shown to be weaker in performance than the scaled latent trace norm (Wimalawarne et al., 2014, 2016). Therefore, in our experiments, we considered only mixed norms based on the scaled latent trace norm.

*3.2.1 Extensions for Multiple Matrices and Tensors.* Our newly defined norms can be extended to multiple matrices coupled to a tensor on different modes. For instance, we can couple two matrices $M_1 \in \mathbb{R}^{n_1 \times m_1}$ and $M_2 \in \mathbb{R}^{n_3 \times m_2}$ to a three-way tensor $\mathcal{T}$ on its first and third modes. If we regularize the coupled tensor with the overlapped trace norm on modes 1 and 3 and the scaled latent trace norm on mode 2, we obtain a mixed norm:

$$\|\mathcal{T}, M_1, M_2\|^{1,3}_{(\mathrm{O,S,O})}$$
$$= \inf_{\mathcal{T}^{(1)}+\mathcal{T}^{(2)}=\mathcal{T}} \left( \|[T^{(1)}_{(1)}; M_1]\|_{\mathrm{tr}} + \frac{1}{\sqrt{n_2}} \|T^{(2)}_{(2)}\|_{\mathrm{tr}} + \|[T^{(1)}_{(3)}; M_2]\|_{\mathrm{tr}} \right).$$

Coupled norms for multiple three-mode or higher-dimensional tensors could also be designed using our proposed method. However, such extension may require extending coupled norms further. Extensions to coupled norms for multiple tensors are a promising area for future research.

**3.3 Dual Norms.** We now briefly look at dual norms for the coupled norms. Dual norms are useful in deriving excess risk bounds, as discussed in section 4. Due to space limitations, we derive dual norms for only two coupled norms to better understand their nature. To derive them, we first need to know the Schatten norm (Tomioka & Suzuki, 2013) for the coupled tensor norms. We first define the Schatten-$(p, q)$ norm for the coupled norm $\|\mathcal{T}, M\|^1_{(\mathrm{O,O,O})}$ with an additional subscript notation $\underline{S_p/q}$:

$$\|\mathcal{T}, M\|^1_{(\mathrm{O,O,O}),\underline{S_p/q}} := \left(\left(\sum_i^{r_1} \sigma_i\left([T_{(1)}; M]\right)^p\right)^{\frac{q}{p}} + \left(\sum_j^{r_2} \sigma_j\left(T_{(2)}\right)^p\right)^{\frac{q}{p}} + \left(\sum_k^{r_3} \sigma_k\left(T_{(3)}\right)^p\right)^{\frac{q}{p}}\right)^{\frac{1}{q}}, \tag{3.9}$$

where $p$ and $q$ are constants; $r_1$, $r_2$, and $r_3$ are the ranks; and $\sigma_i$, $\sigma_j$, and $\sigma_k$ are the singular values for each unfolding.

The following theorem presents the dual norm of $\|\mathcal{T}, M\|^1_{(\mathrm{O,O,O}),\underline{S_p/q}}$ (see appendix A for proof).

**Theorem 1.** *Let a matrix $M \in \mathbb{R}^{n_1 \times m}$ and a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be coupled on their first modes. The dual norm of $\|\mathcal{T}, M\|^1_{(O,O,O),\underline{S_p/q}}$ with $1/p + 1/p^* = 1$ and $1/q + 1/q^* = 1$ is*

$$\|\mathcal{T}, M\|^1_{(O,O,O),\overline{S_{p^*}/q^*}} = \inf_{\mathcal{T}^{(1)}+\mathcal{T}^{(2)}+\mathcal{T}^{(3)}=\mathcal{T}} \left(\left(\sum_i^{r_1} \sigma_i\left([T^{(1)}_{(1)}; M]\right)^{p^*}\right)^{\frac{q^*}{p^*}} + \left(\sum_j^{r_2} \sigma_j\left(T^{(2)}_{(2)}\right)^{p^*}\right)^{\frac{q^*}{p^*}} + \left(\sum_k^{r_3} \sigma_k\left(T^{(3)}_{(3)}\right)^{p^*}\right)^{\frac{q^*}{p^*}}\right)^{\frac{1}{q^*}},$$

*where $r_1$, $r_2$, and $r_3$ are the ranks for each mode and $\sigma_i$, $\sigma_j$, and $\sigma_k$ are the singular values for each unfolding of the coupled tensor.*

In the special case of $p = 1$ and $q = 1$, we see that $\|\mathcal{T}, M\|^1_{(\mathrm{O,O,O}),\underline{S_1/1}} = \|\mathcal{T}, M\|^1_{(\mathrm{O,O,O})}$. Its dual norm is the spectral norm, as shown in the following corollary:

**Corollary 1.** *Let a matrix $M \in \mathbb{R}^{n_1 \times m}$ and a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be coupled on their first mode. The dual norm of $\|\mathcal{T}, M\|^1_{(O,O,O),\underline{S_1/1}}$ is*

$$\|\mathcal{T}, M\|^1_{(O,O,O),\overline{S_\infty/\infty}} = \inf_{\mathcal{T}^{(1)}+\mathcal{T}^{(2)}+\mathcal{T}^{(3)}=\mathcal{T}} max\Big(\|[T^{(1)}_{(1)}; M]\|_{\text{op}}, \|T^{(2)}_{(2)}\|_{op}, \|T^{(3)}_{(3)}\|_{\text{op}}\Big).$$

The Schatten-$(p, q)$ norm for the mixed norm $\|\cdot\|^1_{(L,O,O)}$ is defined as

$$\|\mathcal{T}, M\|^1_{(L,O,O),\underline{S_p/q}} = \inf_{\mathcal{T}^{(1)}+\mathcal{T}^{(2)}=\mathcal{T}} \Bigg(\Big(\sum_i^{r_1} \sigma_i\big([T^{(1)}_{(1)}; M]\big)^p\Big)^{\frac{q}{p}} + \Big(\sum_j^{r_2} \sigma_j\big(T^{(2)}_{(2)}\big)^p\Big)^{\frac{q}{p}} + \Big(\sum_k^{r_3} \sigma_k\big(T^{(2)}_{(3)}\big)^p\Big)^{\frac{q}{p}}\Bigg)^{\frac{1}{q}}.$$

Its dual norm is defined by the following theorem (see appendix A for the proof):

**Theorem 2.** *Let a matrix $M \in \mathbb{R}^{n_1 \times m}$ and a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be coupled on their first mode. The dual norm of the mixed coupled norm $\|\mathcal{T}, M\|^1_{(L,O,O),\underline{S_p/q}}$ with $1/p + 1/p^* = 1$ and $1/q + 1/q^* = 1$ is*

$$\|\mathcal{T}, M\|^1_{(L,O,O),\overline{S_{p^*}/q^*}} = \Bigg(\Big(\sum_i^{r_1} \sigma_i\big([T_{(1)}; M]\big)^{p^*}\Big)^{\frac{q^*}{p^*}} + \inf_{\hat{\mathcal{T}}^{(1)}+\hat{\mathcal{T}}^{(2)}=\mathcal{T}} \Bigg(\Big(\sum_j^{r_2} \sigma_j\big(\hat{T}^{(1)}_{(2)}\big)^{p^*}\Big)^{\frac{q^*}{p^*}} + \Big(\sum_k^{r_3} \sigma_k\big(\hat{T}^{(2)}_{(3)}\big)^{p^*}\Big)^{\frac{q^*}{p^*}}\Bigg)\Bigg)^{\frac{1}{q^*}},$$

*where $r_1$, $r_2$, and $r_3$ are the ranks of $T_{(1)}$, $\hat{T}^{(1)}_{(2)}$, and $\hat{T}^{(2)}_{(3)}$, respectively, and $\sigma_i$, $\sigma_j$, and $\sigma_k$ are their singular values.*

The dual norms of other mixed norms can be similarly derived.

## 4 Optimization

In this section, we discuss optimization of the proposed completion model, 3.1. The model can be easily solved for each coupled norm using a state-of-the-art optimization method such as the alternating direction method of multipliers (ADMM) method (Boyd, Parikh, Chu, Peleato, & Eckstein, 2011). The optimization steps for the coupled norm $\|\mathcal{T}, M\|^1_{(\text{S,O,O})}$ are derived using the ADMM method. The optimization steps for the other norms are similarly derived.

We express equation 3.1 using the $\|\mathcal{T}, M\|^1_{(\mathrm{S,O,O})}$ norm:

$$\min_{\mathcal{T}^{(1)},\mathcal{T}^{(2)},M} \frac{1}{2}\|\Omega_M(M-\hat{M})\|_{\mathrm{F}}^2 + \frac{1}{2}\|\Omega_{\mathcal{T}}(\mathcal{T}^{(1)}+\mathcal{T}^{(2)}-\hat{\mathcal{T}})\|_{\mathrm{F}}^2 + \lambda\left(\frac{1}{\sqrt{n_1}}\|[T^{(1)}_{(1)};M]\|_{\mathrm{tr}} + \|T^{(2)}_{(2)}\|_{\mathrm{tr}} + \|T^{(2)}_{(3)}\|_{\mathrm{tr}}\right). \tag{4.1}$$

By introducing auxiliary variables $X \in \mathbb{R}^{n_1\times m}$ and $\mathcal{Y} \in \mathbb{R}^{n_1\times n_2\times n_3}$, we can formulate the objective function of ADMM for equation 4.1:

$$\begin{aligned}\min_{\mathcal{T}^{(1)},\mathcal{T}^{(2)},M} & \frac{1}{2}\|\Omega_M(M-\hat{M})\|_{\mathrm{F}}^2 + \frac{1}{2}\|\Omega_{\mathcal{T}}(\mathcal{T}^{(1)}+\mathcal{T}^{(2)}-\hat{\mathcal{T}})\|_{\mathrm{F}}^2 \\ & + \lambda\left(\frac{1}{\sqrt{n_1}}\|[Y^{(1)}_{(1)};X]\|_{\mathrm{tr}} + \|Y^{(2)}_{(2)}\|_{\mathrm{tr}} + \|Y^{(2)}_{(3)}\|_{\mathrm{tr}}\right) \\ & \text{s.t.} \quad X = M, \quad \mathcal{Y}^{(1)} = \mathcal{T}^{(1)}, \quad \mathcal{Y}^{(k)} = \mathcal{T}^{(2)} \quad k = 2,3.\end{aligned} \tag{4.2}$$

We introduce Lagrangian multipliers $W^M \in \mathbb{R}^{n_1\times m}$ and $\mathcal{W}^{\mathcal{T}^{(k)}} \in \mathbb{R}^{n_1\times n_2\times n_3}$, $(k = 1, 2, 3)$ and formulate the Lagrangian as

$$\begin{aligned}\min_{\mathcal{T}^{(1)},\mathcal{T}^{(2)},M} & \frac{1}{2}\|\Omega_M(M-\hat{M})\|_{\mathrm{F}}^2 + \frac{1}{2}\|\Omega_{\mathcal{T}}(\mathcal{T}^{(1)}+\mathcal{T}^{(2)}-\hat{\mathcal{T}})\|_{\mathrm{F}}^2 \\ & + \lambda\left(\frac{1}{\sqrt{n_1}}\|[Y^{(1)}_{(1)};X]\|_{\mathrm{tr}} + \|Y^{(2)}_{(2)}\|_{\mathrm{tr}} + \|Y^{(2)}_{(3)}\|_{\mathrm{tr}}\right) + \langle W^M, M-X\rangle \\ & + \langle \mathcal{W}^{\mathcal{T}^{(1)}}, \mathcal{T}^{(1)}-\mathcal{Y}^{(1)}\rangle + \sum_{k=2}^{3}\langle \mathcal{W}^{\mathcal{T}^{(k)}}, \mathcal{T}^{(2)}-\mathcal{Y}^{(k)}\rangle + \frac{\beta}{2}\|M-X\|_F^2 \\ & + \frac{\beta}{2}\|\mathcal{T}^{(1)}-\mathcal{Y}^{(1)}\|_F^2 + \frac{\beta}{2}\sum_{k=2}^{3}\|\mathcal{T}^{(2)}-\mathcal{Y}^{(k)}\|_F^2,\end{aligned} \tag{4.3}$$

where $\beta$ is a proximity parameter. Using this Lagrangian formulation, we can obtain solutions for unknown variables $M$, $\mathcal{T}^{(1)}$, $\mathcal{T}^{(2)}$, $W^M$, $\mathcal{W}^{\mathcal{T}^{(k)}}$ $(k = 1, 2, 3)$, $X$, and $\mathcal{Y}^{(k)}$ $(k = 1, 2, 3)$ iteratively. We use superscripts $[t]$ and $[t-1]$ to represent the variables at iteration steps $t$ and $t-1$, respectively.

The solutions for $M$ at each iteration can be obtained by solving the following subproblem:

$$M^{[t]} = unvec\big((\Omega_M^\top\Omega_M + \beta I_M)^{-1} vec\big(\Omega_M(\hat{M}) - W^{M[t-1]} + \beta X^{[t-1]}\big)\big).$$

Solutions for $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ at iteration step $t$ can be obtained from the following subproblem:

$$\begin{aligned}&\begin{bmatrix} \Omega_{\mathcal{T}}^{\top}\Omega_{\mathcal{T}} + 2\beta I_{\mathcal{T}} & I_{\mathcal{T}} \\ I_{\mathcal{T}} & \Omega_{\mathcal{T}}^{\top}\Omega_{\mathcal{T}} + 2\beta I_{\mathcal{T}} \end{bmatrix} \begin{bmatrix} vec(\mathcal{T}^{(1)[t]}) \\ vec(\mathcal{T}^{(2)[t]}) \end{bmatrix} \\ &= \begin{bmatrix} vec\Big(\Omega_{\hat{\mathcal{T}}}(\hat{\mathcal{T}}) - \sum_{k=2}^{3} \mathcal{W}^{\mathcal{T}^{(k)}[t-1]} + \beta \sum_{k=2}^{3} \mathcal{Y}^{(k)[t-1]}\Big) \\ vec\Big(\Omega_{\hat{\mathcal{T}}}(\hat{\mathcal{T}}) - \sum_{k=2}^{3} \mathcal{W}^{\mathcal{T}^{(k)}[t-1]} + \beta \sum_{k=2}^{3} \mathcal{Y}^{(k)[t-1]}\Big) \end{bmatrix},\end{aligned} \tag{4.4}$$

where $I_M$ and $I_{\mathcal{T}}$ are unit diagonal matrices with dimensions $n_1 m \times n_1 m$ and $n_1 n_2 n_3 \times n_1 n_2 n_3$, respectively.

The updates for $X$ and $\mathcal{Y}^{(k)}$, $(k = 1, 2, 3)$ at iteration step $t$ are given as

$$[Y_{(1)}^{(1)[t-1]}; X^{[t-1]}] = \text{prox}_{\lambda/(\sqrt{n_1}\beta)}\left(\left[\frac{W_{(1)}^{\mathcal{T}^{(1)}[t-1]}}{\beta}; \frac{W^{M[t-1]}}{\beta}\right] + [T_{(1)}^{(1)[t]}; M^{[t]}]\right) \tag{4.5}$$

and

$$Y_{(k)}^{(k)[t-1]} = \text{prox}_{\lambda/\beta}\left(\frac{W_{(k)}^{\mathcal{T}^{(t)}[t-1]}}{\beta} + T_{(k)}^{(2)[t]}\right), \quad k = 2, 3, \tag{4.6}$$

where $\text{prox}_{\lambda}(\mathrm{X}) = \mathrm{U}(\mathrm{S} - \lambda)_{+}\mathrm{V}^{\top}$ for $X = USV^{\top}$.

The update rules for the dual variables are

$$\begin{aligned} W^{M[t]} &= W^{M[t-1]} + \beta(M^{[t]} - X^{[t]}), \\ \mathcal{W}^{\mathcal{T}^{(1)}[t-1]} &= \mathcal{W}^{\mathcal{T}^{(1)}[t]} + \beta(\mathcal{T}^{(1)[t]} - \mathcal{Y}^{(1)[t]}), \\ \mathcal{W}^{\mathcal{T}^{(k)}[t-1]} &= \mathcal{W}^{\mathcal{T}^{(k)}[t]} + \beta(\mathcal{T}^{(k)[t]} - \mathcal{Y}^{(k)[t]}), \quad k = 2, 3. \end{aligned}$$

We can modify the above optimization procedures by replacing the variables in equation 4.1 in accordance with the norm that is used to regularize the tensor and by adjusting operations in equations 4.2 and 4.4 to 4.6. For example, for the norm $\|\cdot\|_{(\mathrm{O},\mathrm{O},\mathrm{O})}^{1}$, there is only a single $\mathcal{T}$, so the subproblem for equation 4.4 becomes

$$(\Omega_{\mathcal{T}}^{\top}\Omega_{\mathcal{T}} + 3\beta I_{\mathcal{T}})vec(\mathcal{T}^{[t]}) = vec\Big(\Omega_{\hat{\mathcal{T}}}(\hat{\mathcal{T}}) - \sum_{k=1}^{3} \mathcal{W}^{\mathcal{T}^{(k)}[t-1]} + \beta \sum_{k=1}^{3} \mathcal{Y}^{[t-1]}\Big),$$

and that for equation 4.5 becomes

$$[Y_{(1)}^{(1)[t]}; X^{[t]}] = \text{prox}_{\lambda/\beta}\left(\left[\frac{W_{(1)}^{\mathcal{T}^{(k)}[t-1]}}{\beta}; \frac{W^{M[t-1]}}{\beta}\right] + [T_{(1)}^{[t]}; M^{[t]}]\right)$$

and

$$Y_{(k)}^{(k)[t-1]} = \text{prox}_{\lambda/\beta}\left(\frac{W_{(k)}^{\mathcal{T}^{(k)}[t-1]}}{\beta} + T_{(k)}^{[t]}\right), \quad k = 1, 2, 3.$$

Additionally, the dual update rule with $\mathcal{T}$ becomes

$$\mathcal{W}^{\mathcal{T}^{(k)}[t-1]} = \mathcal{W}^{\mathcal{T}^{(k)}[t]} + \beta(\mathcal{T}^{[t]} - \mathcal{Y}^{(k)[t]}), \quad k = 1, 2, 3.$$

The optimization procedures for the other norms can be similarly derived.

## 5 Theoretical Analysis

In this section, we analyze the excess risk bounds of the completion model introduced in equation 3.1 for the coupled norms defined in section 3 using transductive Rademacher complexity (El-Yaniv & Pechyony, 2007; Shamir & Shalev-Shwartz, 2014). Let us again consider matrix $M$ and tensor $\mathcal{T}$ and use them as a single structure $\boldsymbol{X} = \mathcal{T} \cup M$ with a training sample index set $\text{S}_{\text{Train}}$ and a testing sample index set $\text{S}_{\text{Test}}$ with the total set of observed samples $\text{S} = \text{S}_{\text{Train}} \cup \text{S}_{\text{Test}}$. We rewrite equation 3.1 with our new notations as an equivalent model:

$$\min_{\boldsymbol{W}} \frac{1}{|\text{S}_{\text{Train}}|} \sum_{(i_1,i_2,i_3)\in \text{S}_{\text{Train}}} l(\boldsymbol{X}_{i_1,i_2,i_3}, \boldsymbol{W}_{i_1,i_2,i_3}) \ \text{ s.t. } \ \|\boldsymbol{W}\|_{\text{cn}} \leq B, \tag{5.1}$$

where $l(a, b) = (a - b)^2$, $\boldsymbol{W} = \mathcal{W} \cup W_M$ is the learned coupled structure consisting of components $\mathcal{W}$ and $W_M$ of the tensor and matrix, respectively; $B$ is a constant; and $\|\cdot\|_{\text{cn}}$ is any norm defined in section 3.2.

Given that $l(\cdot, \cdot)$ is a $\Lambda$-Lipschitz loss function bounded by $\sup_{i_1,i_2,i_3} |l(\boldsymbol{X}_{i_1,i_2,i_3}, \boldsymbol{W}_{i_1,i_2,i_3})| \leq b_l$ and assuming that $|\text{S}_{\text{Train}}| = |\text{S}_{\text{Test}}| = |S|/2$, we can obtain the following excess risk bound based on transductive Rademacher complexity theory (El-Yaniv & Pechyony, 2007; Shamir & Shalev-Shwartz, 2014) with probability $1 - \delta$,

$$\frac{1}{|\text{S}_{\text{Test}}|} \sum_{(i_1,i_2,i_3)\in \text{S}_{\text{Test}}} l(\boldsymbol{X}_{i_1,i_2,i_3}, \boldsymbol{W}_{i_1,i_2,i_3})$$

$$-\frac{1}{|\text{S}_{\text{Train}}|} \sum_{(i_1,i_2,i_3)\in \text{S}_{\text{Train}}} l(\boldsymbol{X}_{i_1,i_2,i_3}, \boldsymbol{W}_{i_1,i_2,i_3})$$

$$\leq 4R(\mathcal{W}) + b_l\left(\frac{11 + 4\sqrt{\log\frac{1}{\delta}}}{\sqrt{|\mathrm{S}_{\mathrm{Train}}|}}\right), \tag{5.2}$$

where $R(\mathcal{W})$ is the transductive Rademacher complexity defined as

$$R(\mathcal{W}) = \frac{1}{|\mathrm{S}|}\mathbb{E}_\sigma\left[\sup_{\|\mathcal{W}\|_{\mathrm{cn}}\leq B}\sum_{(i_1,i_2,i_3)\in \mathrm{S}}\sigma_{i_1,i_2,i_3}l(\mathcal{W}_{i_1,i_2,i_3}, \mathcal{X}_{i_1,i_2,i_3})\right], \tag{5.3}$$

where $\sigma_{i_1,i_2,i_3} \in \{-1, 1\}$ with probability 0.5 if $(i_1, i_2, i_3) \in \mathrm{S}$, or 0 otherwise (see appendix B for derivation).

Next we give the bounds for equation 5.3 with respect to different coupled norms. We assume that $|\mathrm{S}_{\mathrm{Train}}| = |\mathrm{S}_{\mathrm{Test}}|$, as in Shamir and Shalev-Shwartz (2014) but our theorem can be extended to more general cases. Detailed proofs of the theorems in this section are given in appendix B.

The following two theorems give the Rademacher complexities for coupled completion regularized using the coupled norms $\|\cdot\|^1_{(\mathrm{O},\mathrm{O},\mathrm{O})}$ and $\|\cdot\|^1_{(\mathrm{S},\mathrm{S},\mathrm{S})}$.

**Theorem 3.** *Let $\|\cdot\|_{cn} = \|\cdot\|^1_{(O,O,O)}$; then, with probability $1-\delta$,*

$$R(\mathcal{W}) \leq \frac{3\Lambda}{2|S|}\left[\sqrt{r_{(1)}}(B_{\mathcal{T}} + B_M) + \sum_{k=2}^{3}\sqrt{r_k}B_{\mathcal{T}}\right]$$
$$max\left\{C_2\left(\sqrt{n_1} + \sqrt{\textstyle\prod_{j=2}^{3} n_j + m}\right), \min_{k\in 2,3} C_1\left(\sqrt{n_k} + \sqrt{\textstyle\prod_{j\neq k}^{3} n_j}\right)\right\},$$

*where $(r_1, r_2, r_3)$ is the multilinear rank of $\mathcal{W}$, $r_{(1)}$ is the rank of the coupled unfolding on mode 1, and $B_M$, $B_T$, $C_1$, and $C_2$ are constants.*

**Theorem 4.** *Let $\|\cdot\|_{cn} = \|\cdot\|^1_{(S,S,S)}$. Then, with probability $1-\delta$,*

$$R(\mathcal{W}) \leq \frac{3\Lambda}{2|S|}\left[\sqrt{\frac{r_{(1)}}{n_1}}(B_M + B_{\mathcal{T}}) + \min_{k\in 2,3}\sqrt{\frac{r_k}{n_k}}B_{\mathcal{T}}\right]$$
$$max\left\{C_2\left(n_1 + \sqrt{\textstyle\prod_{i=1}^{3} n_i + n_1 m}\right), C_1 \max_{k=2,3}\left(n_k + \sqrt{\textstyle\prod_{i=1}^{3} n_i}\right)\right\},$$

*where $(r_1, r_2, r_3)$ is the multilinear rank of $\mathcal{W}$, $r_{(1)}$ is the rank of the coupled unfolding on mode 1, and $B_M$, $B_T$, $C_1$, and $C_2$ are constants.*

We can see that in both of these theorems, the Rademacher complexity of the coupled tensor is divided by the total number of observed samples of

both the matrix and the tensor. If the tensor or the matrix is completed separately, then the Rademacher complexity is divided only by their individual samples (see theorems 7 to 9 in appendix B and a discussion in Shamir & Shalev-Shwartz, 2014). This means that coupled tensor learning can lead to better performance than separate matrix or tensor learning. We can also see that due to coupling, the excess risks are bounded by the ranks of both the tensors and the concatenated matrix of the unfolded tensors on the coupled mode. Additionally, the maximum term on the right takes the combinations of both the tensor and the concatenated matrix of the unfolded tensors on the coupled mode.

Finally, we consider the Rademacher complexity of the mixed norm $\|\cdot\|_{\mathrm{cn}} = \|\cdot\|^1_{(\mathrm{S,O,O})}$:

**Theorem 5.** *Let* $\|\cdot\|_{cn} = \|\cdot\|^1_{(S,O,O)}$. *Then, with probability* $1-\delta$,

$$R(\boldsymbol{W}) \leq \frac{3\Lambda}{2|S|}\left[\sqrt{\frac{r_{(1)}}{n_1}}(B_M + B_{\mathcal{T}}) + \sum_{i=2,3}\sqrt{r_i}B_{\mathcal{T}}\right]$$
$$max\left\{C_2\left(n_1 + \sqrt{\prod_{i=1}^{3} n_i + n_1 m}\right), \min_{k=2,3} C_1\left(\sqrt{n_k} + \sqrt{\prod_{i\neq k}^{3} n_i}\right)\right\},$$

*where* $(r_1, r_2, r_3)$ *is the multilinear rank of* $\mathcal{W}$, $r_{(1)}$ *is the rank of the coupled unfolding on mode 1, and* $B_M$, $B_{\mathcal{T}}$, $C_1$, *and* $C_2$ *are constants.*

We see that for the mixed norm $\|\cdot\|_{\mathrm{cn}} = \|\cdot\|^1_{(\mathrm{S,O,O})}$, the excess risk is bounded by the scaled rank of the coupled unfolding along the first mode. For this norm, we can see that the terms related to ranks are smaller in theorem 3 and that the maximum term could be smaller than in theorem 4. This means that this norm can perform better than $\|\cdot\|^1_{(\mathrm{O,O,O})}$ and $\|\cdot\|^1_{(\mathrm{S,S,S})}$ depending on the ranks and mode dimensions of the coupled tensor. The bounds of the other two mixed norms can also be derived and explained in a manner similar to theorem 5.

## 6 Evaluation

We evaluated our proposed method experimentally using synthetic and real-world data.

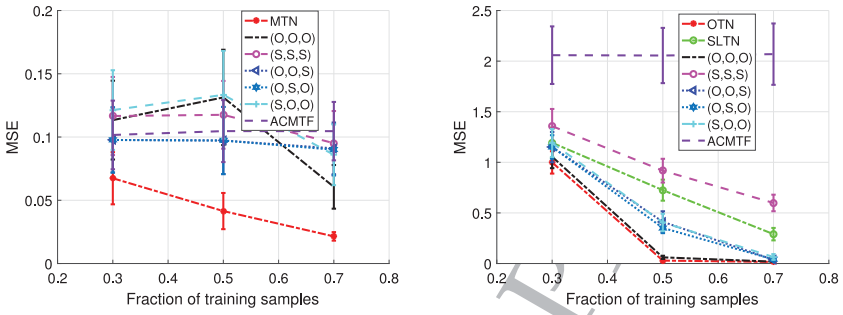
**6.1 Synthetic Data.** Our main objectives were to evaluate how the proposed norms perform depending on the ranks and dimensions of the coupled tensors. We used simulation data based on CP rank and Tucker rank in these experiments.

*6.1.1 Experiments Using CP Rank.* To create coupled tensors with the CP rank, we first generated a three-mode tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with CP rank $r$ using CP decomposition (Kolda & Bader, 2009) as $\mathcal{T} = \sum_{i=1}^{r} c_i u_i \circ v_i \circ w_i$ where $u_i \in \mathbb{R}^{n_1}$, $v_i \in \mathbb{R}^{n_2}$ and $w_i \in \mathbb{R}^{n_3}$ and $c_i \in \mathbb{R}^{+}$. For our experiments, we used two approaches to create CP-rank-based tensors in which all the component vectors $u_i$, $v_i$, and $w_i$ were nonorthogonal vectors or orthogonal vectors. We coupled matrix $X \in \mathbb{R}^{n_1 \times m}$ with rank $r$ to $\mathcal{T}$ on mode 1 by generating $X = USV^\top$ with $U(1:r,:) = [u_1, \ldots, u_r]$, $S \in \mathbb{R}^{r \times r}$, and $V \in \mathbb{R}^{m \times r}$ is an orthogonal matrix. We also added noise sampled from a gaussian distribution with mean zero and variance of 0.01 to the elements of the matrix and the tensor.

In our experiments using synthetic data, we considered coupled structures of tensors with dimension $20 \times 20 \times 20$ and matrices with dimension $20 \times 30$ coupled on their first modes. To simulate completion, we randomly selected observed samples with percentages of 30, 50, and 70 of the total number of elements in both the matrix and the tensor; selected a validation set with a percentage of 10; and took the remainder as test samples. We performed coupled completion using the proposed coupled norms of $\|\cdot\|^1_{(O,O,O)}$, $\|\cdot\|^1_{(S,S,S)}$, $\|\cdot\|^1_{(S,O,O)}$, $\|\cdot\|^1_{(O,S,O)}$, and $\|\cdot\|^1_{(O,O,S)}$. For all the learning models with these norms, we cross-validated their regularization parameters ranging from 0.01 to 5.0 with intervals of 0.05. We ran our experiments with 10 random selections and plotted the mean square error (MSE) for the test samples.

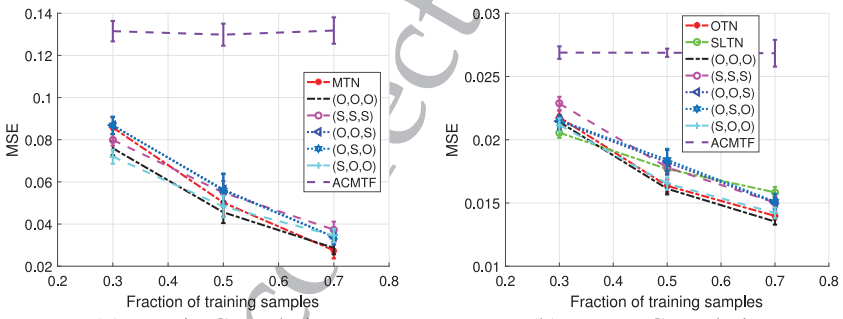
As benchmark methods, we used the overlapped trace norm (OTN) and the scaled latent trace norm (SLTN) for individual tensors and the matrix trace norm (MTN) for individual matrices. For all these norms, we cross-validated the regularization parameters ranging from 0.01 to 5.0 with intervals of 0.05. We compared our results with those of advanced coupled matrix-tensor factorization ACMTF (Acar, Papalexakis et al., 2014b), for which the regularization parameters were selected using cross-validation in the range $0, 0.0001, 0.001, \ldots, 1$. To select ranks to use with the ACMTF method, we first ran experiments using ranks of $1, 3, 5, \ldots, 19$ and selected the rank that gave the best performance. Due to the nonconvex nature of ACMTF, we ran experiments with five random initializations to select the best local optimal solution.

We first ran experiments on coupled tensor completion based on CP rank in different settings. In the first experiment, we considered coupled tensors with no shared components. In this experiment, we created a tensor with CP rank 5 in which the component vectors were nonorthogonal and generated from a normal distribution. We also created a matrix of rank 5 and without any components in common with the tensor. Figure 2 shows that the coupled norms did not perform better than individual matrix completion using the matrix trace norm. However, for tensor completion, the coupled norm $\|\cdot\|^1_{(O,O,O)}$ had performance comparable to that of the overlapped trace norm.




(a) Matrix Completion (b) Tensor Completion.

Figure 2: Completion performance of a matrix with dimension $20 \times 30$ and rank 5 with no sharing and of a tensor with dimension $20 \times 20 \times 20$ and CP rank 5 with nonorthogonal component vectors.




(a) Matrix Completion (b) Tensor Completion.

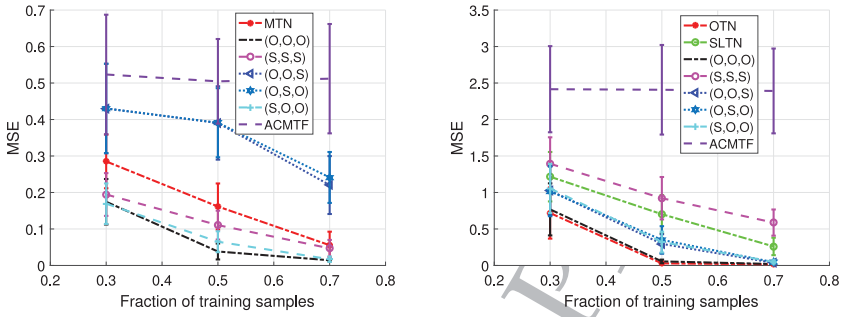
Figure 3: Completion performance of a matrix with dimension $20 \times 30$ and rank 5 with all components shared and of a tensor with dimension $20 \times 20 \times 20$ and CP rank 5 with orthogonal component vectors.



We next ran experiments on coupled tensors with some components in common and with both orthogonal and nonorthogonal component vectors. We created coupled tensors with CP rank of 5 and both the tensor and matrix shared all components along mode 1. We generated the tensor with orthogonal component vectors. As shown in Figure 3, the coupled norm $\|\cdot\|^1_{(O,O,O)}$ had good performance for both the matrix and tensor.

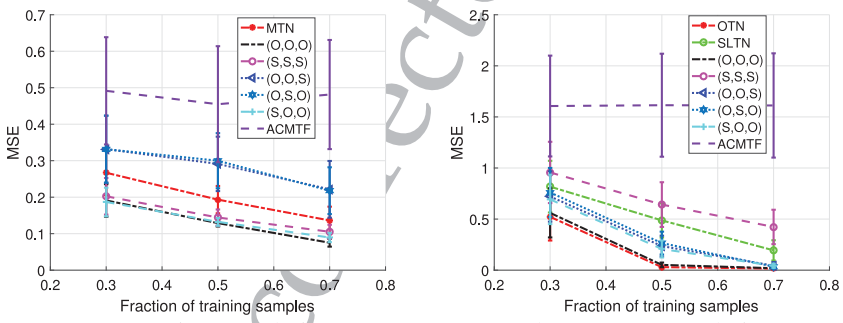
Figure 4 shows the performance of coupled tensors with the same rank as in the previous experiment with tensors created from nonorthogonal component vectors. Again, the coupled norm $\|\cdot\|^1_{(O,O,O)}$ had better performance than individual matrix and tensor completions.

In our final experiment, we created tensors with CP rank 5 and coupled them with a matrix of rank 10 sharing all five component vectors along




(a) Matrix Completion (b) Tensor Completion.

Figure 4: Completion performance of a matrix with dimension $20 \times 30$ and rank 5 with all component vectors shared and of a tensor with dimension $20 \times 20 \times 20$ and CP rank 5 and nonorthogonal component vectors.



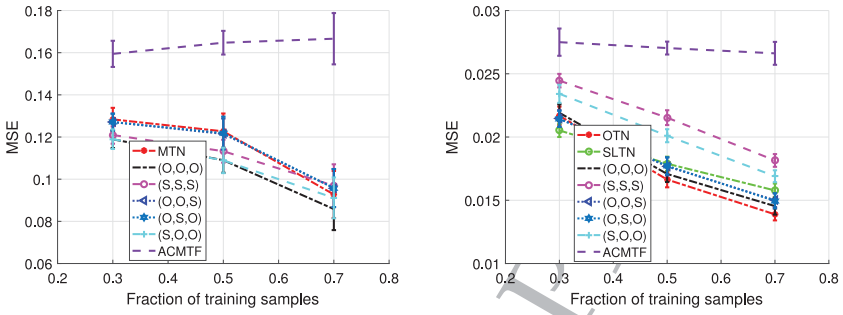

(a) Matrix Completion (b) Tensor Completion.

Figure 5: Completion performance of a matrix with dimension $20 \times 30$ and rank 5 and of a tensor with dimension $20 \times 20 \times 20$ with CP rank 10 and nonorthogonal component vectors that shared five components.



mode 1. Figures 5 and 6 show the results for tensors created with orthogonal and nonorthogonal component vectors, respectively. In both cases, the coupled norms $\|\cdot\|^1_{(O,O,O)}$, $\|\cdot\|^1_{(S,S,S)}$, and $\|\cdot\|^1_{(S,O,O)}$ had better matrix completion performance than individual completion by the matrix trace norm. Similarly, as in the previous experiments, both the overlapped trace norm and the coupled norm $\|\cdot\|^1_{(O,O,O)}$ had comparable performances.

*6.1.2 Simulations Using Tucker Rank.* To create coupled tensors with the Tucker rank, we first generated a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ using Tucker decomposition (Kolda & Bader, 2009) as $\mathcal{T} = \mathcal{C} \times_1 U_1 \times_2 U_2 \times_3 U_3$, where $\mathcal{C} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ was the core tensor generated from a normal distribution specifying multilinear rank $(r_1, r_2, r_3)$ and component matrices $U_1 \in \mathbb{R}^{r_1 \times p_1}$,



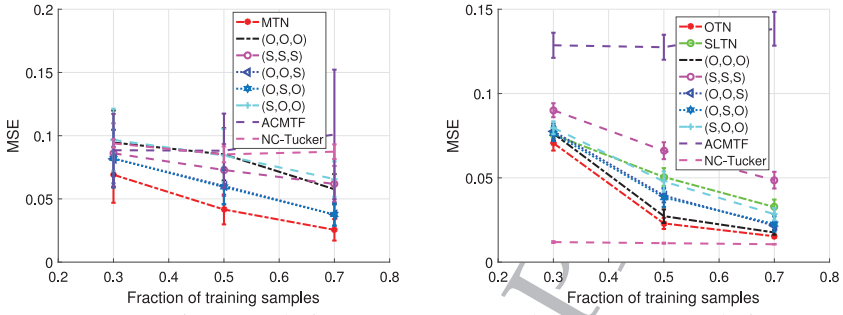

(a) Matrix Completion (b) Tensor Completion.

Figure 6: Completion performance of a matrix with dimension $20 \times 30$ and rank 5 and of a tensor with dimension $20 \times 20 \times 20$ and CP rank 10 and orthogonal component vectors that shared five components.

$U_2 \in \mathbb{R}^{r_2 \times p_2}$, and $U_3 \in \mathbb{R}^{r_3 \times p_3}$ were orthogonal matrices. Next, we generated a matrix that was coupled with mode 1 of the tensor using singular value decomposition $X = USV^\top$, where we specified its rank $r$ using diagonal matrix $S$ and generated matrices $U$ and $V$ as orthogonal matrices. For sharing between the matrix and the tensor, we computed $T_{(1)} = U_n S_n V_n^\top$ and replaced the first $s$ singular values of $S$ with the first $s$ singular values of $S_n$, replaced the first basis vectors $s$ of $U$ with the first $s$ basis vectors of $U_n$, and computed $X = USV^\top$ such that the coupled structure shared $s$ common components. We also added noise sampled from a gaussian distribution with mean zero and variance 0.01 to the elements of the coupled tensor.

As in the synthetic experiments using the CP rank, we considered coupled structures with tensors with dimension $20 \times 20 \times 20$ and matrices with dimension $20 \times 30$ coupled on their mode 1. We considered different multilinear ranks of tensors, ranks of matrices, and degrees of sharing among them. We used the same percentages in selecting the training, testing, and validation sets as we did in the CP rank experiments. We again compared our results with those of ACMTF.

We also used an additional nonconvex coupled learning model to incorporate multilinear ranks of the coupled tensor by considering Tucker decomposition under the assumption that the components of the coupled mode were shared between both the matrix and tensor. We used the Tensorlab framework (Vervliet, Debals, Sorber, Van Barel, & De Lathauwer, 2016) to implement this model. We regularized the factorized components of the tensor (including the core tensor) and the matrix using the Frobenius norm. We used a regularization parameter selected from the range 0.01 to 50 in logarithmic linear scale with five divisions (in Matlab syntax `exp(linspace(log(0.01), log(50), 5))`). We refer to this benchmark method as NC-Tucker. Due to the nonconvex nature of the model, we ran




(a) Matrix Completion (b) Tensor Completion.

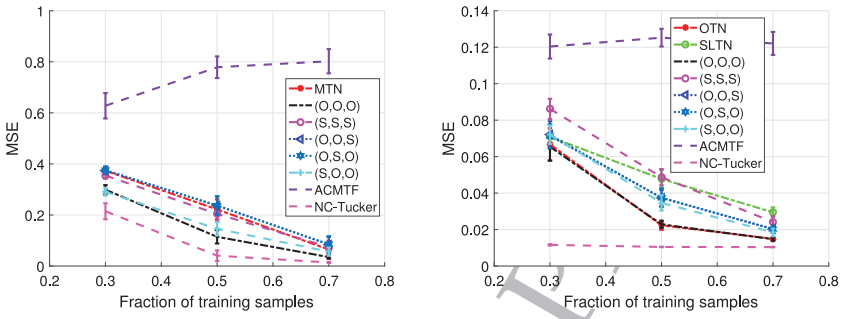
Figure 7: Completion performance of a matrix with dimension $20 \times 30$ and rank 5 and of a tensor with dimension $20 \times 20 \times 20$ and multilinear rank $(5, 5, 5)$ with no sharing.

5 to 10 simulations with different random initializations and selected the best local optimal solution. Specifying the multilinear rank a priori for this model would be challenging in real applications, but since we knew the rank in our simulations, we could specify the multilinear ranks to be used to create the tensors.

In our first simulations, we considered a coupled tensor with a matrix rank of 5 and a tensor multilinear rank $(5, 5, 5)$ with no shared components. Figure 7 shows that with this setting, individual matrix and tensor completion had better performance than that of the coupled norms. The nonconvex NC-Tucker benchmark method had the best performance for the tensor but performed poorly in matrix completion compared to the coupled norms.

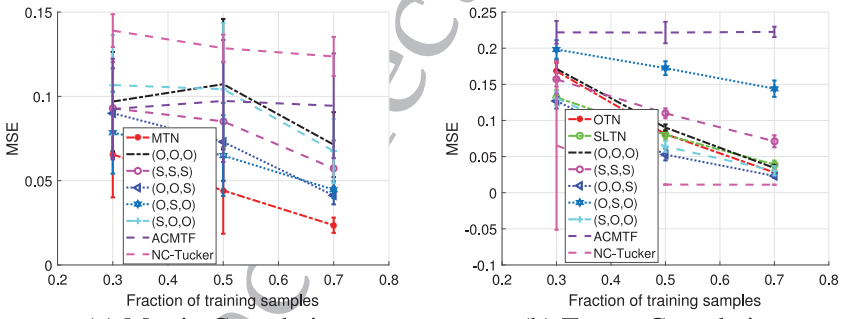
In our next simulation, we considered coupling of tensors and matrices with some degree of sharing among them. We created a matrix of rank 5 and a tensor of multilinear rank $(5, 5, 5)$ and let them share all five singular components along mode 1. Figure 8 shows that the coupled norm $\|\cdot\|^1_{(O,O,O)}$ had the best performance among the coupled norms for both matrix and tensor completion. Individual tensor completion with the overlapped trace norm had the same performance as $\|\cdot\|^1_{(O,O,O)}$. The NC-Tucker method performed better than the coupled norms for tensor and matrix completion.

In our next simulation, we considered a matrix of rank 5 and a tensor of multilinear rank $(5, 15, 5)$ that shared all five singular components along mode 1. Figure 9 shows that with this setting, although the coupled norm $\|\cdot\|^1_{(O,O,S)}$ had the best performance among the coupled norms and individual tensor completion, it was outperformed by the NC-Tucker method. However, the NC-Tucker method performed poorly in matrix completion compared to the coupled norms. For the matrix completion, individual matrix completion by the matrix trace norm had the best performance, while coupled norms $\|\cdot\|^1_{(O,O,S)}$ and $\|\cdot\|^1_{(S,O,O)}$ had the next best performance.




(a) Matrix Completion (b) Tensor Completion.

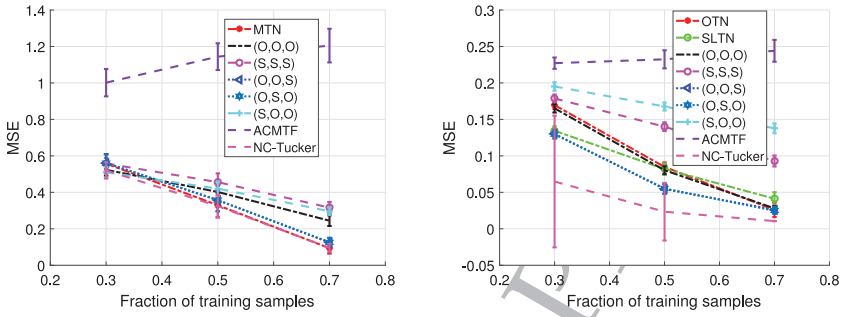
Figure 8: Completion performances of completion of a matrix with dimension $20 \times 30$ and rank 5 and of a tensor with dimension $20 \times 20 \times 20$ and multilinear rank $(5, 5, 5)$ that shared five components.




(a) Matrix Completion (b) Tensor Completion.

Figure 9: Completion performance of a matrix with dimension $20 \times 30$ and rank 5 and of a tensor with dimension $20 \times 20 \times 20$ and multilinear rank $(5, 15, 5)$ that shared five components.

For our final simulation, we created a coupled matrix with rank 5 and a tensor with multilinear rank $(15, 5, 5)$, all sharing five singular components along mode 1. Figure 10 shows that the mixed coupled norms $\|\cdot\|^1_{(O,S,O)}$ and $\|\cdot\|^1_{(O,O,S)}$ performed equally and had better performance for tensor completion than the individual tensor completion. The NC-Tucker method had better performance than the coupled norms for tensor completion, while the performance was comparable for matrix completion. For matrix completion when the percentage of training samples was small, coupled norms $\|\cdot\|^1_{(O,O,O)}$ and $\|\cdot\|^1_{(S,O,O)}$ had better performance. As the percentage of training samples was increased, the performance of individual matrix




(a) Matrix Completion (b) Tensor Completion.

Figure 10: Completion performance of completion of a matrix with dimension $20 \times 30$ and rank 5 and of a tensor with dimension $20 \times 20 \times 20$ and multilinear rank $(15, 5, 5)$ that shared five components.

completion improved, while those of $\|\cdot\|^1_{(O,S,O)}$ and $\|\cdot\|^1_{(O,O,S)}$ were close but second best.

The results of these simulations show that the ACMTF performed poorly compared to our proposed methods.

**6.2 Real-World Data.** As a real-world data experiment, we applied our proposed method to the UCLAF data set (Zheng, Cao, Zheng, Xie, & Yang, 2010), which consists of GPS data for 164 users in 168 locations performing five activities, resulting in a sparse user-location-activity tensor $\mathcal{T} \in \mathbb{R}^{164\times168\times5}$. This data set also has a user-location matrix $X \in \mathbb{R}^{164\times168}$, which we used as side information coupled to the user mode of $\mathcal{T}$. Using similar observed element percentages as in the synthetic data simulations, we performed completion experiments on $\mathcal{T}$. We considered all the elements of the user-location matrix as observed elements and used them as training data. We repeated the evaluation for 10 random sample selections. We cross-validated the regularization parameters from 0.01 to 500 divided into 50 in logarithmic linear scale. As a baseline method, we again used the ACMTF method (Acar, Papalexakis et al., 2014) with CP rank 5. Additionally, we used the coupled (Tucker) method (Ermis et al., 2015) and the NC-Tucker method with multilinear rank $(3, 3, 3)$, where we selected the best performances among 5 random initializations. Figure 11 shows the completion performances for the coupled tensor.

We can see that the best performance among coupled norms was that of mixed coupled norm $\|\cdot\|^1_{(S,O,O)}$, indicating that learning with side information as a coupled structure improves tensor completion performance compared to completion using only tensor norms. This also indicates that mode 1 may have a lower rank than the other modes and that modes 2 and

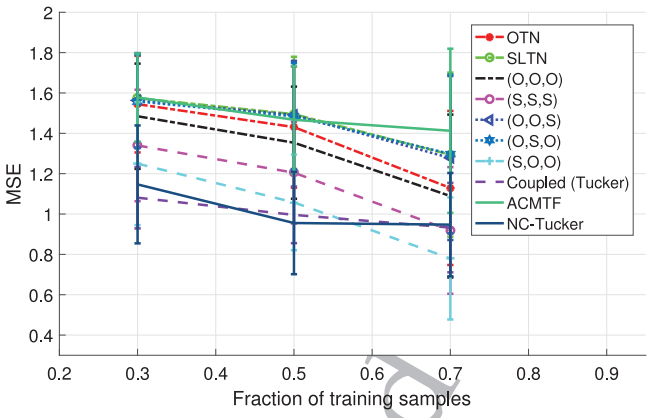



Figure 11: Completion performance for UCLAF data.



3 may have ranks closer to each other. The nonconvex coupled (Tucker) method and the NC-Tucker method had better performance than $\|\cdot\|^1_{(S,O,O)}$ when the number of observed samples was less than 70 percent of the total elements.

## 7 Conclusion and Future Work

We have proposed a new set of convex norms for the completion problem of coupled tensors. We restricted our study to coupling a three-way tensor with a matrix and defined low-rank inducing norms by extending trace norms such as the overlapped trace norm and scaled latent trace norm of tensors and the matrix trace norm. We also introduced the concept of mixed norms, which combines the features of both overlapped and latent trace norms. We looked at the theoretical properties of our convex completion model and evaluated it using synthetic and real-world data. We found that the proposed coupled norms perform comparably to existing nonconvex ones. However, our norms lead to global optimal solutions and eliminate the need for specifying the ranks of the coupled tensors beforehand. While there are still many aspects to be studied, we believe that our work is the first step in modeling convex norms for coupled tensors.

Although coupling can occur among many tensors with different dimensions and multiple matrices on different modes, this study focused on a three-mode tensor and a single matrix. The methodology used to create coupled norms can be extended to any of those settings, but mere extensions may not lead to the optimal design of norms for those settings. Particularly, the square tensor norm (Mn, Huang, Wright, & Goldfarb, 2014) has shown to be better suited to tensors beyond three modes and thus can also be used

to model novel coupled norms in the future. Furthermore, theoretical analysis using methods such as the gaussian width (Amelunxen, Lotz, McCoy, & Tropp, 2014) may provide a deeper understanding of coupled tensors, which should enable the design of better norms. Such studies could be interesting directions for future research.

## Appendix A: Proofs of Dual Norms

We first provide the proofs of the dual norms of theorems 1 and 2.

**Proof of Theorem 1.** We use lemma 3 of Tomioka and Suzuki (2013) to prove the duality. Consider a linear operator $\Phi$ such that $\Phi(\mathcal{T}, M) = [\text{vec}(M); \text{vec}(T_{(1)}); \text{vec}(T_{(2)}); \text{vec}(T_{(3)})] \in \mathbb{R}^{d_1+3d_2}$, where $d_1 = n_1 m$ and $d_2 = n_1 n_2 n_3$. We define

$$\|z\|_* = \left( \|[Z^{(1)}_{(1)}; X]\|^q_{S_p} + \sum_{k=2}^{3} \|Z^{(k)}_{(k)}\|^q_{S_p} \right)^{1/q}, \tag{A.1}$$

where $\mathcal{Z}^{(k)}$ is the inverse vectorization of elements $z_{(d_1+(k-1)d_2+1):(d_1+kd_2)}$ and $X$ is the inverse vectorization of $z_{1:d_1}$. The dual of the above norm is expressed as

$$\|z\|_{*^*} = \left( \|[Z^{(1)}_{(1)}; X]\|^{q^*}_{S_{p^*}} + \sum_{k=2}^{3} \|Z^{(k)}_{(k)}\|^{q^*}_{S_{p^*}} \right)^{1/q^*}.$$

Let

$$\Phi^\top(z) = \{\mathcal{T}, M\} = \left\{ \sum_{k=1}^{3} \mathcal{Z}^{(k)}, X \right\}.$$

Then following lemma 3 of Tomioka and Suzuki (2013), we write

$$|||[\mathcal{T}, M]|||_{\overline{*}(\Phi)} = \inf \|z\| \quad s.t \quad \Phi^\top(z) = \{\mathcal{T}, M\}.$$

Given that

$$|||[\mathcal{T}, M]|||_{\underline{*}(\Phi)} := \|[\mathcal{T}, M]\|^1_{(\text{O,O,O}),\underline{S_p/q}},$$

and following lemma 3 in Tomioka and Suzuki (2013) we obtain the dual of $\|[\mathcal{T}, M]\|^1_{(\text{O,O,O}),\underline{S_p/q}}$ as $\|[\mathcal{T}, M]\|^1_{(\text{L,L,L}),\overline{S_{p^*}/q^*}}$. □

**Proof of Theorem 2.** We can apply theorem 1 to latent tensors $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$, as well as the dual of the overlapping norm to $\mathcal{T}$. First, consider the dual with respect to $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$. By applying theorem 1, we obtain

$$\|\mathcal{T}, M\|^1_{(\mathrm{L,O,O}),\overline{S_{p^*/q^*}}} = \left(\left(\sum_i^{r_1} \sigma_i\left([T_{(1)}; M]\right)^{p^*}\right)^{\frac{q^*}{p^*}} + \|\mathcal{T}\|^{q^*}_{(-,\mathrm{O,O}),S^*_p}\right)^{\frac{1}{q^*}}.$$

Next, by applying lemma 1 of Tomioka and Suzuki (2013) to $\|\mathcal{T}\|_{(-,\mathrm{O,O})}$, we obtain

$$\|\mathcal{T}, M\|^1_{(\mathrm{L,O,O}),\overline{S_{p^*/q^*}}} = \left(\left(\sum_i^{r_1} \sigma_i\left([T_{(1)}; M]\right)^{p^*}\right)^{\frac{q^*}{p^*}} + \inf_{\hat{\mathcal{T}}^{(1)}+\hat{\mathcal{T}}^{(2)}=\mathcal{T}} \left(\left(\sum_j^{r_2} \sigma_j\left(\hat{T}^{(1)}_{(2)}\right)^{p^*}\right)^{\frac{q^*}{p^*}} + \left(\sum_k^{r_3} \sigma_k\left(\hat{T}^{(2)}_{(3)}\right)^{p^*}\right)^{\frac{q^*}{p^*}}\right)\right)^{\frac{1}{q^*}}.$$

□

## Appendix B: Proofs of Excess Risk Bounds

Here we derive the excess risk bounds for the coupled completion problem.

From previous work (El-Yaniv & Pechyony, 2007; Shamir & Shalev-Shwartz, 2014), we know that for a loss function $l(\cdot,\cdot)$ that is, a $\Lambda$-Lipschitz loss function and bounded as $\sup_{i_1,i_2,i_3} |l(\boldsymbol{X}_{i_1,i_2,i_3}, \boldsymbol{W}_{i_1,i_2,i_3})| \leq b_l$ and with the assumption that $|\mathrm{S}_{\mathrm{Train}}| = |\mathrm{S}_{\mathrm{Test}}| = |\mathrm{S}|/2$, we have the following bound for equation 5.1 based on transductive Rademacher complexity theory (El-Yaniv and Pechyony, 2007; Shamir & Shalev-Shwartz, 2014) with probability $1-\delta$,

$$\frac{1}{|\mathrm{S}_{\mathrm{Test}}|} \sum_{(i_1,i_2,i_3)\in \mathrm{S}_{\mathrm{Test}}} l(\boldsymbol{X}_{i_1,i_2,i_3}, \boldsymbol{W}_{i_1,i_2,i_3}) - \frac{1}{|\mathrm{S}_{\mathrm{Train}}|} \sum_{(i_1,i_2,i_3)\in \mathrm{S}_{\mathrm{Train}}} l(\boldsymbol{X}_{i_1,i_2,i_3}, \boldsymbol{W}_{i_1,i_2,i_3})$$
$$\leq\ 4R(\boldsymbol{W}) + b_l\left(\frac{11+4\sqrt{\log\frac{1}{\delta}}}{\sqrt{|S_{\mathrm{Train}}|}}\right),$$

where $R(\boldsymbol{W})$ is transductive Rademacher complexity defined as

$$R(\boldsymbol{W}) = \frac{1}{|\mathrm{S}|}\mathbb{E}_\sigma\left[\sup_{\|\boldsymbol{W}\|_{\mathrm{cn}}\leq B} \sum_{(i_1,i_2,i_3)\in \mathrm{S}} \sigma_{i_1,i_2,i_3} l(\boldsymbol{W}_{i_1,i_2,i_3}, \boldsymbol{X}_{i_1,i_2,i_3})\right], \tag{B.1}$$

where $\sigma_{i_1,i_2,i_3} \in \{-1, 1\}$ with probability 0.5 if $(i_1, i_2, i_3) \in \mathrm{S}$, or 0 otherwise.

We can rewrite equation B.1 as

$$R(\boldsymbol{W}) = \frac{1}{|\mathrm{S}|}\mathbb{E}_\sigma\left[\sup_{\|\boldsymbol{W}\|_{\mathrm{cn}}\leq B_M+B_\mathcal{T}}\sum_{(i_1,i_2,i_3)\in\mathrm{S}}\sigma_{i_1,i_2,i_3}l(\boldsymbol{W}_{i_1,i_2,i_3},\boldsymbol{X}_{i_1,i_2,i_3})\right]$$
$$\leq \frac{\Lambda}{|\mathrm{S}|}\mathbb{E}_\sigma\sup_{\|\boldsymbol{W}\|_{\mathrm{cn}}\leq B_M+B_\mathcal{T}}\sum_{(i_1,i_2,i_3)\in\mathrm{S}}\sigma_{i_1,i_2,i_3}\boldsymbol{W}_{i_1,i_2,i_3}$$
$$\text{(Rademacher contraction)},$$
$$\leq \frac{\Lambda}{|\mathrm{S}|}\mathbb{E}_\sigma\sup_{\|\boldsymbol{W}\|_{\mathrm{cn}}\leq B_M+B_\mathcal{T}}\|\boldsymbol{W}\|_{\mathrm{cn}}\|\Sigma\|_{\mathrm{cn}^*}\text{(Holder's inequality)},$$

where we have used that $\|\mathcal{W}\|_{\mathrm{F}} \leq B_\mathcal{T}$ and $\|W_M\|_{\mathrm{F}} \leq B_M$, and $\Sigma$ is of dimensions of the coupled tensor consisting Rademacher variables ($\Sigma_{i_1,i_2,i_3} = \sigma_{i_1,i_2,i_3}$ if $(i_1, i_2, i_3) \in \mathrm{S}$, else $\Sigma_{i_1,i_2,i_3} = 0$).

**Proof of Theorem 3.** Let $\boldsymbol{W} = \mathcal{W} \cup W_M$, where $\mathcal{W}$ and $W_M$ are the completed tensors of $\mathcal{T}$ and $M$, and let $\Sigma = \Sigma_\mathcal{T} \cup \Sigma_M$, where $\Sigma_\mathcal{T}$ and $\Sigma_M$ consist of the corresponding Rademacher variables ($\sigma_{i_1,i_2,i_3}$) for $\mathcal{T}$ and $M$. Since we use an overlapping norm, we have $\|\boldsymbol{W}\|_{\mathrm{cn}} = \|\mathcal{W}, W_M\|^1_{(\mathrm{O},\mathrm{O},\mathrm{O})}$ from which we obtain

$$\|\mathcal{W}, W_M\|^1_{(\mathrm{O},\mathrm{O},\mathrm{O})} = \|[W_{(1)}; W_M]\|_{\mathrm{tr}} + \sum_{k=2}^{3}\|W_{(k)}\|_{\mathrm{tr}}$$
$$\leq \sqrt{r_{(1)}}(B_\mathcal{T} + B_M) + \sum_{k=2}^{3}\sqrt{r_k}B_\mathcal{T},$$

where $(r_1, r_2, r_3)$ is the multilinear rank of $\mathcal{W}$ and $r_{(1)}$ is the rank of the concatenated matrix of unfolding tensors on mode 1. To obtain the above inequality, we used the fact that for any matrix $U$ with rank $r$, we have $\|U\|_{\mathrm{tr}} \leq \sqrt{r}\|U\|_{\mathrm{F}}$ (Tomioka & Suzuki, 2013).

Using Latała's theorem (Latała, 2005; Shamir & Shalev-Shwartz, 2014) for the mode $k$ unfolding, we can bound $\|\Sigma_{\mathcal{T}(k)}\|_{\mathrm{op}}$

$$\mathbb{E}\|\Sigma_{\mathcal{T}(k)}\|_{\mathrm{op}} \leq C_1\left(\sqrt{n_k} + \sqrt{\textstyle\prod_{j\neq k}^{3} n_j} + \sqrt[4]{|\Sigma_{\mathcal{T}(k)}|}\right),$$

and since $\sqrt[4]{|\Sigma_{\mathcal{T}(k)}|} \leq \sqrt[4]{\prod_{i=1}^{3} n_i} \leq \frac{1}{2}\left(\sqrt{n_k} + \sqrt{\prod_{j\neq k}^{3} n_j}\right)$, we have

$$\mathbb{E}\|\Sigma_{\mathcal{T}(k)}\|_{\mathrm{op}} \leq \frac{3C_1}{2}\left(\sqrt{n_k} + \sqrt{\prod_{j\neq k}^{3} n_j}\right).$$

Similarly, using Latała's theorem, we obtain

$$\mathbb{E}\|[\Sigma_{\mathcal{T}(1)}; \Sigma_M]\|_{\mathrm{op}} \leq \frac{3C_2}{2}\left(\sqrt{n_1} + \sqrt{\prod_{j=2}^{3} n_j + m}\right).$$

To bound $\mathbb{E}\|\Sigma_{\mathcal{T}}, \Sigma_M\|^1_{(\mathrm{O},\mathrm{O},\mathrm{O})^*}$, we use the duality relationship from theorem 1 and corollary 1:

$$\|\Sigma_{\mathcal{T}}, \Sigma_M\|^1_{(\mathrm{O},\mathrm{O},\mathrm{O})^*} = \inf_{\Sigma^{(1)}_{\mathcal{T}}+\Sigma^{(2)}_{\mathcal{T}}+\Sigma^{(3)}_{\mathcal{T}}=\Sigma_{\mathcal{T}}} \max\left\{\|[\Sigma^{(1)}_{\mathcal{T}(1)}; \Sigma_M]\|_{\mathrm{op}}, \|\Sigma^{(2)}_{\mathcal{T}(2)}\|_{\mathrm{op}}, \|\Sigma^{(3)}_{\mathcal{T}(3)}\|_{\mathrm{op}}\right\}.$$

Since we can take any $\Sigma^{(k)}_{\mathcal{T}}$ to be equal to $\Sigma_{\mathcal{T}}$, the above norm can be upper-bounded:

$$\|\Sigma_{\mathcal{T}}, \Sigma_M\|^1_{(\mathrm{O},\mathrm{O},\mathrm{O})^*} \leq \max\left\{\|[\Sigma_{\mathcal{T}(1)}; \Sigma_M]\|_{\mathrm{op}}, \min\left\{\|\Sigma_{\mathcal{T}(2)}\|_{\mathrm{op}}, \|\Sigma_{\mathcal{T}(3)}\|_{\mathrm{op}}\right\}\right\}.$$

Taking the expectation leads to

$$\begin{aligned}
&\mathbb{E}\|\Sigma_{\mathcal{T}}, \Sigma_M\|^1_{(\mathrm{O},\mathrm{O},\mathrm{O})^*} \\
&\quad\leq \mathbb{E}\max\left\{\|[\Sigma_{\mathcal{T}(1)}; \Sigma_M]\|_{\mathrm{op}}, \min\left\{\|\Sigma_{\mathcal{T}(2)}\|_{\mathrm{op}}, \|\Sigma_{\mathcal{T}(3)}\|_{\mathrm{op}}\right\}\right\} \\
&\quad\leq \max\left\{\mathbb{E}\|[\Sigma_{\mathcal{T}(1)}; \Sigma_M]\|_{\mathrm{op}}, \min\left\{\mathbb{E}\|\Sigma_{\mathcal{T}(2)}\|_{\mathrm{op}}, \mathbb{E}\|\Sigma_{\mathcal{T}(3)}\|_{\mathrm{op}}\right\}\right\}.
\end{aligned}$$

Finally, we have

$$\begin{aligned}
R(\boldsymbol{W}) \leq \frac{3\Lambda}{2|\mathrm{S}|}&\left[\sqrt{r_{(1)}}(B_{\mathcal{T}} + B_M) + \sum_{k=2}^{3}\sqrt{r_k}B_{\mathcal{T}}\right] \\
&\max\left\{C_2\left(\sqrt{n_1} + \sqrt{\prod_{j=2}^{3} n_j + m}\right), \min_{k\in 2,3} C_1\left(\sqrt{n_k} + \sqrt{\prod_{j\neq k}^{3} n_j}\right)\right\}.
\end{aligned}$$

□

Before we give the excess risk bound for the $\|\cdot\|^1_{(\mathrm{S},\mathrm{S},\mathrm{S})}$, in the following theorem, we give the excess risk of coupled completion with the $\|\cdot\|^1_{(\mathrm{L},\mathrm{L},\mathrm{L})}$.

**Theorem 6.** *Let* $\|\cdot\|_{\text{cn}} = \|\cdot\|^1_{(\text{L},\text{L},\text{L})}$. *Then, with probability* $1-\delta$,

$$R(\boldsymbol{W}) \leq \frac{3\Lambda}{2|\text{S}|}\left[\sqrt{r_{(1)}}B_M + \min\left(\sqrt{r_{(1)}}, \min_{k=2,3}\sqrt{r_k}\right)B_{\mathcal{T}}\right]$$
$$\max\left\{C_2\left(\sqrt{n_1}+\sqrt{\prod_{j=2}^3 n_j + m}\right), \max_{k=2,3}\left\{C_2\left(\sqrt{n_k}+\sqrt{\prod_{j\neq k}^3 n_j}\right)\right\}\right\},$$

*where* $(r_1, r_2, r_3)$ *is the multilinear rank of* $\mathcal{W}$, $r_{(1)}$ *is the rank of the coupled unfolding on mode 1, and* $B_M$, $B_{\mathcal{T}}$, $C_1$, *and* $C_2$ *are constants.*

**Proof.** Let $\boldsymbol{W} = \mathcal{W} \cup W_M$, where $\mathcal{W}$ and $W_M$ are the completed tensors of $\mathcal{T}$ and $M$ and $\Sigma = \Sigma_{\mathcal{T}} \cup \Sigma_M$, where $\Sigma_{\mathcal{T}}$ and $\Sigma_M$ consist of the corresponding Rademacher variables. We can see that

$$\|\boldsymbol{W}\|^1_{(\text{L},\text{L},\text{L})} = \inf_{\mathcal{W}^{(1)}+\mathcal{W}^{(2)}+\mathcal{W}^{(3)}=\mathcal{W}}\left(\|[W^{(1)}_{(1)}; W_M]\|_{\text{tr}} + \sum_{k=2}^3 \|W^{(k)}_{(k)}\|_{\text{tr}}\right),$$

which can be bounded as

$$\|\boldsymbol{W}\|^1_{(\text{L},\text{L},\text{L})} \leq \sqrt{r_{(1)}}(B_M + B_{\mathcal{T}}) + \min_{k=2,3}\sqrt{r_k}B_{\mathcal{T}},$$

where the last term is derived by considering the infimum with respect to $\mathcal{W}^{(2)}$ and $\mathcal{W}^{(3)}$.

Using the duality result given in theorem 1 (corollary 1) and Latała's theorem, we obtain

$$\begin{aligned}\|\Sigma_{\mathcal{T}}, \Sigma_M\|^1_{(\text{L},\text{L},\text{L})^*} &\leq \max\left\{\mathbb{E}\|[\Sigma_{\mathcal{T}(1)}; \Sigma_M]\|_{\text{op}}, \mathbb{E}\|\Sigma_{\mathcal{T}(2)}\|_{\text{op}}, \mathbb{E}\|\Sigma_{\mathcal{T}(3)}\|_{\text{op}}\right\} \\ &\leq \frac{3}{2}\max\left\{C_2\left(\sqrt{n_1}+\sqrt{\prod_{j=2}^3 n_j + m}\right),\right. \\ &\qquad \left.\max_{k=2,3}\left\{C_1\left(\sqrt{n_k}+\sqrt{\prod_{j\neq k}^3 n_j}\right)\right\}\right\}.\end{aligned}$$

Finally, we have

$$R(\boldsymbol{W}) \leq \frac{3\Lambda}{2|\text{S}|}\left[\sqrt{r_{(1)}}(B_M + B_{\mathcal{T}}) + \min_{k=2,3}\sqrt{r_k}B_{\mathcal{T}}\right]$$
$$\max\left\{C_2\left(\sqrt{n_1}+\sqrt{\prod_{j=2}^3 n_j + m}\right), \max_{k=2,3}\left\{C_1\left(\sqrt{n_k}+\sqrt{\prod_{j\neq k}^3 n_j}\right)\right\}\right\}.$$

□

**Proof of Theorem 4.** By definition, we have

$$\|\boldsymbol{W}\|^1_{(\mathrm{S,S,S})} = \inf_{\mathcal{W}^{(1)}+\mathcal{W}^{(2)}+\mathcal{W}^{(3)}=\mathcal{W}} \left( \frac{1}{\sqrt{n_1}} \|[W^{(1)}_{(1)}, W_M]\|_{\mathrm{tr}} + \sum_{k=2,3} \frac{1}{\sqrt{n_k}} \|W^{(k)}_{(k)}\|_{\mathrm{tr}} \right),$$

which results in

$$\|\boldsymbol{W}\|^1_{(\mathrm{S,S,S})} \leq \sqrt{\frac{r_{(1)}}{n_1}}(B_M + B_{\mathcal{T}}) + \min_{k\in 2,3} \sqrt{\frac{r_k}{n_k}} B_{\mathcal{T}}.$$

Using the duality result given in theorem 1 and Latała's theorem, we obtain

$$\begin{aligned}
&\mathbb{E}\|\Sigma_{\mathcal{T}}, \Sigma_M\|^1_{(\mathrm{S,S,S})^*} \\
&= \mathbb{E} \max \left\{ \sqrt{n_1} \|[\Sigma_{\mathcal{T}(1)}; \Sigma_M]\|_{\mathrm{op}}, \sqrt{n_2} \|\Sigma_{\mathcal{T}(2)}\|_{\mathrm{op}}, \sqrt{n_3} \|\Sigma_{\mathcal{T}(3)}\|_{\mathrm{op}} \right\} \\
&\leq \frac{3}{2} \max \left\{ C_2 \left( n_1 + \sqrt{\textstyle\prod_{i=1}^3 n_i + n_1 m} \right), C_1 \max_{k=2,3} \left( n_k + \sqrt{\textstyle\prod_{i\neq k}^3 n_i} \right) \right\}.
\end{aligned}$$

Finally, we have

$$\begin{aligned}
R(\boldsymbol{W}) \leq \frac{3\Lambda}{2|\mathrm{S}|} &\left[ \sqrt{\frac{r_{(1)}}{n_1}}(B_M + B_{\mathcal{T}}) + \min_{k\in 2,3} \sqrt{\frac{r_k}{n_k}} B_{\mathcal{T}} \right] \\
&\max \left\{ C_2 \left( n_1 + \sqrt{\textstyle\prod_{i=1}^3 n_i + n_1 m} \right), C_1 \max_{k=2,3} \left( n_k + \sqrt{\textstyle\prod_{i=1}^3 n_i} \right) \right\}.
\end{aligned}$$

□

**Proof of Theorem 5.** First, let us look at $\|\boldsymbol{W}\|^1_{(\mathrm{S,O,O})}$, which is expressed as

$$\|\boldsymbol{W}\|^1_{(\mathrm{S,O,O})} = \inf_{\mathcal{W}^{(1)}+\mathcal{W}^{(2)}=\mathcal{W}} \left( \frac{1}{\sqrt{n_1}} \|[W^{(1)}_{(1)}; W_M]\|_{\mathrm{tr}} + \|W^{(2)}_{(2)}\|_{\mathrm{tr}} + \|W^{(2)}_{(3)}\|_{\mathrm{tr}} \right).$$

This norm can be upper-bounded:

$$\|\boldsymbol{W}\|^1_{(\mathrm{S,O,O})} \leq \sqrt{\frac{r_{(1)}}{n_1}}(B_M + B_{\mathcal{T}}) + \sum_{i=2,3} \sqrt{r_i} B_{\mathcal{T}}.$$

Now we are left with bounding $\|\Sigma_{\mathcal{T}}, \Sigma_M\|^1_{(\mathrm{S,O,O})^*}$. Using theorem 2, we obtain

$$\|\Sigma_{\mathcal{T}}, \Sigma_M\|^1_{(\mathrm{S,O,O})^*} \leq \max\left(\sqrt{n_1}\|[\Sigma_{\mathcal{T}(1)}; \Sigma_M]\|_{\mathrm{op}}, \min\left(\|\Sigma_{\mathcal{T}(2)}\|_{\mathrm{op}}, \|\Sigma_{\mathcal{T}(3)}\|_{\mathrm{op}}\right)\right).$$

We then have

$$\mathbb{E}\|\Sigma_{\mathcal{T}}, \Sigma_M\|^1_{(\mathrm{S,O,O})^*} \leq \frac{3}{2}\max\left\{C_2\left(n_1 + \sqrt{\textstyle\prod_{i=1}^3 n_i + n_1 m}\right), \min_{k=2,3} C_1\left(\sqrt{n_k} + \sqrt{\textstyle\prod_{i\neq k}^3 n_i}\right)\right\}.$$

The final resulting bound is

$$R(\boldsymbol{W}) \leq \frac{3\Lambda}{2|\mathrm{S}|}\left[\sqrt{\frac{r_{(1)}}{n_1}}(B_M + B_{\mathcal{T}}) + \sum_{i=2,3}\sqrt{r_i}B_{\mathcal{T}}\right] \max\left\{C_2\left(n_1 + \sqrt{\textstyle\prod_{i=1}^3 n_i + n_1 m}\right), \min_{k=2,3} C_1\left(\sqrt{n_k} + \sqrt{\textstyle\prod_{i\neq k}^3 n_i}\right)\right\}.$$

□

In addition to the above transductive bounds for completion with coupled norms, we also provide the bounds for individual tensor completion with tensor norms such as the overlapped trace norm, the latent trace norm, and the scaled latent trace norm. We can consider equation 5.1 only for a matrix or a tensor without coupling and with low-rank regularization. Therefore, we may have the transductive bounds for a matrix $M \in \mathbb{R}^{n_1 \times m}$ (Shamir & Shalev-Shwartz, 2014) as

$$R(W_M) \leq c\frac{B_M\Lambda}{|\mathrm{S}^M|}\sqrt{\hat{r}}\left(\sqrt{n_1} + \sqrt{m}\right), \tag{B.2}$$

where $\mathrm{S}^M$ is the index set of observed samples of matrix $M$, $\hat{r}$ is the rank induced by matrix trace norm regularization, and $c$ is a constant.

Next we can consider the transductive bounds for tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with regularization using norms such as the overlapped trace norm (Tomioka & Suzuki, 2013), the latent trace norm (Tomioka & Suzuki, 2013), and the scaled latent trace norm (Wimalawarne et al., 2014) in the following three theorems. We denote the index set of observed sample of $\mathcal{T}$ by $\mathrm{S}^{\mathcal{T}}$.

**Theorem 7.** *Using the overlapped trace norm regularization given as* $\|\mathcal{W}\|_{\mathrm{overlap}} = \|\mathcal{W}\|_{(\mathrm{O,O,O})}$, *we obtain*

$$R(\mathcal{W}) \leq c_1 \frac{B_{\mathcal{T}} \Lambda}{|S^{\mathcal{T}}|} \left( \sum_{k=1}^{3} \sqrt{\hat{r}_k} \right) \min_k \left( \sqrt{n_k} + \sqrt{\prod_{j \neq k}^{3} n_j} \right),$$

*for some constant* $c_1$*;* $(\hat{r}_1, \hat{r}_2, \hat{r}_3)$ *is the multilinear rank of* $\mathcal{W}$*.*

**Proof.** Using the same procedure as for theorem 3, we obtain

$$\mathbb{E}\|\Sigma_{\mathcal{T}}\|_{\text{overlap}^*} \leq \mathbb{E} \min_k \|\Sigma_{\mathcal{T}(k)}\|_{\text{op}} \leq \min_k \mathbb{E}\|\Sigma_{\mathcal{T}(k)}\|_{\text{op}}$$

$$\leq \frac{3c_1}{2} \min_k \left( \sqrt{n_k} + \sqrt{\prod_{j \neq k}^{3} n_j} \right).$$

Since $\|\mathcal{W}\|_{\text{overlap}} \leq \left( \sum_{k=1}^{3} \sqrt{\hat{r}_k} \right) B_{\mathcal{T}}$, where $\|\mathcal{W}\|_{\text{F}} \leq B_{\mathcal{T}}$ (Tomioka & Suzuki, 2013), we have

$$R(\mathcal{W}) \leq c_1 \frac{B_{\mathcal{T}} \Lambda}{|S^{\mathcal{T}}|} \left( \sum_{k=1}^{3} \sqrt{\hat{r}_k} \right) \min_k \left( \sqrt{n_k} + \sqrt{\prod_{j \neq k}^{3} n_j} \right).$$

□

**Theorem 8.** *Using the latent trace norm regularization given by* $\|\mathcal{W}\|_{\text{latent}} = \|\mathcal{W}\|_{(\text{L,L,L})}$*, we obtain*

$$R(\mathcal{W}) \leq c_2 \Lambda B_{\mathcal{T}} \frac{\min_k \sqrt{\hat{r}_k}}{|S^{\mathcal{T}}|} \max_k \left( \sqrt{n_k} + \sqrt{\prod_{j \neq k}^{3} n_j} \right),$$

*for some constant* $c_2$*;* $(\hat{r}_1, \hat{r}_2, \hat{r}_3)$ *is the multilinear rank of* $\mathcal{W}$*.*

**Proof.** Using the duality result from Wimalawarne et al. (2014), we have

$$\|\Sigma_{\mathcal{T}}\|_{\text{latent}^*} = \max_k \|\Sigma_{\mathcal{T}(k)}\|_{\text{op}}.$$

Using Latała's theorem, we obtain

$$\mathbb{E}\|\Sigma_{\mathcal{T}}\|_{\text{latent}^*} \leq \frac{3c_2}{2} \max_k \left( \sqrt{n_k} + \sqrt{\prod_{j \neq k}^{3} n_j} \right).$$

Finally, using the known bound $\|\mathcal{W}\|_{\text{latent}} \leq \min_i \sqrt{\hat{r}_i} B_{\mathcal{T}}$ (Wimalawarne et al., 2014), where $\|\mathcal{W}\|_{\text{F}} \leq B_{\mathcal{T}}$, we obtain the excess risk:

$$R(\mathcal{W}) \leq \frac{3c_2 \Lambda B_{\mathcal{T}} \min_i \sqrt{\hat{r}_i}}{2|S^{\mathcal{T}}|} \max_k \left( \sqrt{n_k} + \sqrt{\prod_{j \neq k}^{3} n_j} \right).$$

□

**Theorem 9.** *Using the scaled latent trace norm regularization given by* $\|\mathcal{W}\|_{\text{scaled}} = \|\mathcal{W}\|_{(\text{S,S,S})}$*, we obtain*

$$R(\mathcal{W}) \leq \frac{3c_3 \Lambda B_{\mathcal{T}}}{2|\text{S}^{\mathcal{T}}|} \min_i \left( \sqrt{\frac{\hat{r}_i}{n_i}} \right) \max_k \left( n_k + \sqrt{\prod_{j=1}^{3} n_j} \right).$$

*for some constant* $c_3$*;* $(\hat{r}_1, \hat{r}_2, \hat{r}_3)$ *is the multilinear rank of* $\mathcal{W}$*.*

**Proof.** From previous work (Wimalawarne et al., 2014), we can derive

$$\|\Sigma_{\mathcal{T}}\|_{\text{scaled}^*} = \max_k \sqrt{n_k} \|\Sigma_{\mathcal{T}(k)}\|_{\text{op}}.$$

Using an approach similar to that for theorem 8 with the additional scaling of $\sqrt{n_k}$ and using Latała's theorem, we arrive at the following bound:

$$R(\mathcal{W}) \leq \frac{3c_3 \Lambda B_{\mathcal{T}}}{2|\text{S}^{\mathcal{T}}|} \min_i \left( \sqrt{\frac{\hat{r}_i}{n_i}} \right) \max_k \left( n_k + \sqrt{\prod_{j=1}^{3} n_j} \right).$$

□

**Acknowledgments**

M.Y. was supported by the JST PRESTO program JPMJPR165A. H.M. has been partially supported by JST ACCEL grant JPMJAC1503 (Japan), MEXT Kakenhi 16H02868 (Japan), FiDiPro by Tekes (currently Business Finland), and AIPSE programme by Academy of Finland.

**References**

Acar, E., Bro, R., & Smilde, A. K. (2015). Data fusion in metabolomics using coupled matrix and tensor factorizations. *Proceedings of the IEEE*, *103*(9), 1602–1620.

Acar, E., Kolda, T. G., & Dunlavy, D. M. (2011). All-at-once optimization for coupled matrix and tensor factorizations. *CoRR*, abs/1105.3422.

Acar, E., Nilsson, M., & Saunders, M. (2014). A flexible modeling framework for coupled matrix and tensor factorizations. In *Proceedings of the 22nd Signal Processing Conference* (pp. 111–115). Piscataway, NJ: IEEE.

Acar, E., Papalexakis, E. E., Gürdeniz, G., Rasmussen, M. A., Lawaetz, A. J., Nilsson, M., & Bro, R. (2014). Structure-revealing data fusion. *BMC Bioinformatics*, *15*, 239.

Amelunxen, D., Lotz, M., McCoy, M. B., & Tropp, J. A. (2014). Living on the edge: Phase transitions in convex programs with random data. *Information and Inference*, *3*, 224–294.

Argyriou, A., Evgeniou, T., & Pontil, M. (2006). Multi-task feature learning. In B. Schölkopf, J. C. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems, 19* (pp. 41–48). Cambridge, MA: MIT Press.

Banerjee, A., Basu, S., & Merugu, S. (2007). Multi-way clustering on relation graphs. In *Proceedings of the 2007 SIAM International Conference on Data Mining* (pp. 145–156). Philadelphia: SIAM.

Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Tren. Mach. Learn.*, *1*, 1–122.

Carroll, J. D., & Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika*, *35*(3), 283–319.

El-Yaniv, R., & Pechyony, D. (2007). Transductive Rademacher complexity and its applications. In N. H. Bshouty & C. Gentile (Eds.), *Lecture Notes in Computer Science: Vol. 4539. Learning Theory COLT 2007* (pp. 157–171). Berlin: Springer.

Ermis, B., Acar, E., & Cemgil, A. T. (2015). Link prediction in heterogeneous data via generalized coupled tensor factorization. *Data Mining and Knowledge Discovery*, *29*(1), 203–236.

Gunasekar, S., Yamada, M., Yin, D., & Chang, Y. (2015). Consistent collective matrix completion under joint low rank structure. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*. JMLR.org.

Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an explanatory multimodal factor analysis. *UCLA Working Papers in Phonetics*, *16*, 1–84.

Hitchcock, F. L. (1927). The expression of a tensor or a polyadic as a sum of products. *J. Math. Phys.*, *6*(1), 164–189.

Khan, S. A., & Kaski, S. (2014). Bayesian multi-view tensor factorization. In T. Calders, F. Esposito, E. Hüllermeier, & R. Meo (Eds.), *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases* (pp. 656–671). Berlin: Springer.

Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, *51*(3), 455–500.

Latała, R. (2005). Some estimates of norms of random matrices. *Proc. Amer. Math. Soc.*, *133*(5), 1273–1282.

Li, C., Zhao, Q., Li, J., Cichocki, A., & Guo, L. (2015). Multi-tensor completion with common structures. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* (pp. 2743–2749). Palo Alto, CA: AAAI Press.

Liu, G., Lin, Z., & Yu, Y. (2010). Robust subspace segmentation by low-rank representation. In *Proceedings of the International Conference on Machine Learning*. Madison, WI: Omnipress.

Liu, J., Musialski, P., Wonka, P., & Ye, J. (2009). Tensor completion for estimating missing values in visual data. In *Proceedings of the International Conference on Computer Vision* (pp. 2114–2121). Piscataway, NJ: IEEE.

Mn, C., Huang, B., Wright, J., & Goldfarb, D. (2014). Square deal: Lower bounds and improved relaxations for tensor recovery. In *Proceedings of the International Conference on Machine Learning* (pp. 73–81). JMLR.org.

Narita, A., Hayashi, K., Tomioka, R., & Kashima, H. (2011). *Tensor factorization using auxiliary information*. Berlin: Springer.

Romera-Paredes, B., Aung, H., Bianchi-Berthouze, N., & Pontil, M. (2013). Multilinear multitask learning. In *Proceedings of the International Conference on Machine Learning* (pp. 1444–1452). JMLR.org.

Shamir, O., & Shalev-Shwartz, S. (2014). Matrix completion with the trace norm: Learning, bounding, and transducing. *Journal of Machine Learning Research*, *15*, 3401–3423.

Signoretto, M., Dinh, Q. T., De Lathauwer, L., & Suykens, J. A. K. (2013). Learning with tensors: A framework based on convex optimization and spectral regularization. *Machine Learning*, *94*(3), 303–351.

Singh, A. P., & Gordon, G. J. (2008). Relational learning via collective matrix factorization. In *Proceedings of the SIGKKD International Conference on Knowledge Discovery and Data Mining*. New York: ACM.

Takeuchi, K., Tomioka, R., Ishiguro, K., Kimura, A., & Sawada, H. (2013). Non-negative multiple tensor factorization. In *Proceedings of the International Conference on Data Mining* (pp. 1199–1204). Piscataway, NJ: IEEE.

Tomioka, R., & Suzuki, T. (2013). Convex tensor decomposition via structured Schatten norm regularization. In C. J. C. Burges, L. Bottou, M. Welling. Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems*, *26* Red Hook, NY: Curran.

Tomioka, R., Suzuki, T., Hayashi, K., & Kashima, H. (2011). Statistical performance of convex tensor decomposition. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems*, *24* (pp. 972–980). Red Hook, NY: Curran.

Vervliet, N., Debals, O., Sorber, L., Van Barel, M., & De Lathauwer, L. (2016). *Tensorlab 3.0*. https://www.tensorlab.net/

Wimalawarne, K., Sugiyama, M., & Tomioka, R. (2014). Multitask learning meets tensor factorization: task imputation via convex optimization. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems*, *27*. Red Hook, NY: Curran.

Wimalawarne, K., Tomioka, R., & Sugiyama, M. (2016). Theoretical and experimental analyses of tensor-based regression and classification. *Neural Computation*, *28*(4), 686–715.

Zheng, V. W., Cao, B., Zheng, Y., Xie, X., & Yang, Q. (2010). Collaborative filtering meets mobile recommendation: A user-centered approach. In *AAAI*.

Received May 15, 2017; accepted June 8, 2018.