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

Scaled Coupled Norms and Coupled Higher-Order Tensor Completion

Kishan Wimalawarne
kishanwn@gmail.com
Bioinformatics Center, Institute for Chemical Research, Kyoto University,
Uji, Kyoto 611-0011, Japan

Makoto Yamada
makoto.yamada@riken.jp
Graduate School of Informatics, Kyoto University, Yoshida-honmachi, Sakyo-ku,
Kyoto 606-8501, Japan; RIKEN, Center for Advanced Intelligence Project,
Tokyo 103-0027, Japan; Institute of Statistical Mathematics, Tokyo 190-8562, Japan;
and PRESTO, Japan Science and Technological Agency, Japan

Hiroshi Mamitsuka
mami@kuicr.kyoto-u.ac.jp
Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho,
Uji, Kyoto 611-0011, Japan, and Department of Computer Science, Aalto University,
Espoo F1-00076, Finland

Recently, a set of tensor norms known as *coupled norms* has been proposed as a convex solution to coupled tensor completion. Coupled norms have been designed by combining low-rank-inducing tensor norms with the matrix trace norm. Though coupled norms have shown good performances, they have two major limitations: they do not have a method to control the regularization of coupled modes and uncoupled modes, and they are not optimal for couplings among higher-order tensors. In this letter, we propose a method that scales the regularization of coupled components against uncoupled components to properly induce the low-rankness on the coupled mode. We also propose coupled norms for higher-order tensors by combining the square norm to coupled norms. Using the excess risk-bound analysis, we demonstrate that our proposed methods lead to lower risk bounds compared to existing coupled norms. We demonstrate the robustness of our methods through simulation and real-data experiments.

1 Introduction

In recent years, learning from multiple data sources has gained considerable interest. One such field that has gained interest is coupled tensor completion

(also referred to as collective tensor completion), where we impute missing elements of a partially observed tensor by sharing information from its coupled tensors (Acar, Papalexakis et al., 2014; Acar, Nilsson, & Saunders, 2014; Acar, Bro, & Smilde, 2015; Bouchard, Yin, & Guo, 2013). A coupling between two tensors occurs when they share a common mode, where one tensor can be side information (Narita, Hayashi, Tomioka, & Kashima, 2011) to the other or both mutually share information (Acar, Papalexakis et al., 2014). Coupled tensor completion has been useful in several real-world applications such as link prediction (Ermis, Acar, & Cemgil, 2015), recommendation systems (Acar, Kolda, & Dunlavy, 2011; Acar, Papalexakis et al., 2014; Acar, Nilsson et al., 2014; Acar et al., 2015; Jeon, Jeon, Sael, & Kang, 2016), and computer vision (Li, Zhao, Li, Cichocki, & Guo, 2015; Zhou, Qian, Shen, Zhang, & Xu, 2017).

Recently, Wimalawarne, Yamada, and Mamitsuka (2018) proposed a set of norms known as *coupled norms* to solve coupled completion. One of the main advantages of these norms is that they are convex and lead to global solutions, while many of the existing coupled completion models are nonconvex factorization methods (Acar, Nilsson et al., 2014; Ermis et al., 2015). Furthermore, most factorization-based methods are restricted to the CANDECOMP/PARAFAC (CP) rank (Acar, Nilsson et al., 2014) of tensors while others are restricted to nonnegative factorization (Ermis et al., 2015). Coupled norms are able to learn using the multilinear rank of tensors and are applicable to heterogeneous tensor data. Theoretical analysis on completion with coupled norms has shown that proper regularization of low-rankness along the coupled modes leads to better performance compared to individual tensor completion. Except for the computational challenge of using trace norm regularization, these norms can be easily extended to multiple couplings of multiple tensors, thus making them a promising approach to coupled tensor learning.

Although coupled norms have several favorable qualities, we find that they have two limitations. One limitation is that there is no control on regularization of coupled modes with respect to uncoupled modes. In the existing design of coupled norms, there is always an assumption that all modes are low-ranked. This is not an optimal design since the the shared low-rankness induced by the concatenation of the tensors on the coupled modes could be different compared to low-rankness induced by the tensors independently. Another limitation with coupled norms is that their design is limited to coupling of three-mode tensors. A naive application of these coupled norms to higher-order tensors (tensors with more than three dimensions) may not be optimal since these norms have been designed using norms robust for three-mode tensors such as the overlapped trace norm (Liu, Musialski, Wonka, & Ye, 2009; Tomioka & Suzuki, 2013) and scaled latent trace norm (Wimalawarne, Sugiyama, & Tomioka, 2014). The recently proposed square norm (Mu, Huang, Wright, & Goldfarb, 2014) has been

shown to be better for completion of higher-order tensors, which would be more suitable for coupled higher-order tensors.

In this letter, we propose extensions to coupled norms to overcome the limitations we have noted. We introduce scaling of the regularization on the coupled mode with respect to uncoupled modes. Additionally, we integrate the square norm to coupled tensors to create extensions for higher-order tensors. We derive excess risk bounds for coupled completion based on regularization by scaled coupled norms and their higher-order extensions. We provide details of simulation and real-data experiments and show that our methods lead to better performance for coupled completion.

Before we move on to the core concepts of our research, we introduce some notations that we use in this letter. Following Kolda and Bader (2009), we write a K-mode tensor as $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$, and its mode-$k$ unfolding is obtained by $T_{(k)} \in \mathbb{R}^{n_k \times \prod_{j \neq k}^{K} n_j}$, which is obtained by concatenating all slices along mode-k. Given two matrices $M \in \mathbb{R}^{n_1 \times n_2}$ and $N \in \mathbb{R}^{n_1 \times n_2'}$, the notation $[M; N] \in \mathbb{R}^{n_1 \times (n_2 + n_2')}$ represents their concatenation on the common mode-1.

2 A Short Review on Completion with Coupled Norms

Coupled tensors completion using coupled norms was introduced by Wimalawarne et al. (2018). They considered a partially observed three-way tensor $\hat{\mathcal{T}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and a matrix $\hat{M} \in \mathbb{R}^{n_1 \times n_2'}$ having a common mode (mode-1 in this case) with the number of partially observed elements of m_1 and m_2, respectively. Given mappings of $\Omega_{\hat{\mathcal{T}}} : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^{m_1}$ and $\Omega_{\hat{M}} : \mathbb{R}^{n_1 \times n_2'} \to \mathbb{R}^{m_2}$, a collective completion model was proposed as

$$\min_{\mathcal{T}, M} \frac{1}{2} \|\Omega_{\hat{M}}(M - \hat{M})\|_{\mathrm{F}}^2 + \frac{1}{2} \|\Omega_{\hat{\mathcal{T}}}(\mathcal{T} - \hat{\mathcal{T}})\|_{\mathrm{F}}^2 + \lambda \|\mathcal{T}, M\|_{\mathrm{cn}}, \tag{2.1}$$

where $\|\mathcal{T}, M\|_{\mathrm{cn}}$ represents a coupled norm, which can be constructed by using tensor norms such as the overlapped trace norm (Tomioka & Suzuki, 2013), the scaled latent trace norm (Wimalawarne et al., 2014), and the matrix trace norm (Fazel, 2002; Candès & Recht, 2009).

We now review basic constructions of coupled norms. Given a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and a matrix $M \in \mathbb{R}^{n_1 \times n_2'}$, the general definition of a coupled norm takes the following format,

$$\|\mathcal{T}, M\|_{(b,c,d)}^a, \tag{2.2}$$

where the superscript a specifies the mode in which the tensor and the matrix are coupled, and the subscripts indicated by $b, c, d \in \{O, L, S, -\}$ specify the regularization method on each mode of the tensor. The notation O indicates overlapping trace norm-based regularization, L indicates

latent trace norm-based regularization, S indicates scaled latent trace norm-based regularization, and $-$ indicates no regularization with respect to the specified mode. The core building block to construct coupled norms is the matrix trace norm also known as the nuclear norm (Fazel, 2002; Candès & Recht, 2009), which is defined for a matrix $M \in \mathbb{R}^{n_1 \times n'_2}$ with rank J as $\|M\|_{\mathrm{tr}} = \sum_{j=1}^{J} \sigma_j$, where σ_j is the jth nonzero singular value of the matrix M. The matrix trace norm is a convex relaxation to minimizing the rank of a matrix (Fazel, 2002; Candès & Recht, 2009), and it is used to define all low-rank tensor norms (Liu, Lin, & Yu, 2010; Wimalawarne et al., 2014; Mu et al., 2014).

In order to look into few of the norms introduced in Wimalawarne et al. (2018), we first consider the coupled norm $\|\mathcal{T}, M\|_{(O,O,O)}^1$, which takes following format:

$$\|\mathcal{T}, M\|_{(O,O,O)}^1 := \|[T_{(1)}; M]\|_{\mathrm{tr}} + \sum_{k=2}^{3} \|T_{(k)}\|_{\mathrm{tr}}. \tag{2.3}$$

In the above norm, the tensor $\mathcal{T}$ is unfolded on each mode and regularized with the trace norm, and on the coupled mode, the concatenated matrix $[T_{(1)}; M]$ is regularized. The trace norm for the concatenation between the matrix and the tensor induces low-rankness for both the tensor and the matrix with respect to the coupled mode, allowing collective regularization.

A coupled norm with all the modes regularized with the scaled latent norm is defined as $\|\mathcal{T}, M\|_{(S,S,S)}^1$, which takes the following format,

$$\|\mathcal{T}, M\|_{(S,S,S)}^1 = \inf_{\mathcal{T}^{(1)} + \mathcal{T}^{(2)} + \mathcal{T}^{(3)} = \mathcal{T}} \left(\frac{1}{\sqrt{n_1}} \|[T_{(1)}^{(1)}; M]\|_{\mathrm{tr}} + \sum_{k=2}^{3} \frac{1}{\sqrt{n_k}} \|T_{(k)}^{(k)}\|_{\mathrm{tr}} \right), \tag{2.4}$$

where $\mathcal{T}^{(1)}$, $\mathcal{T}^{(2)}$, and $\mathcal{T}^{(3)}$ are latent tensors (Wimalawarne et al., 2014, 2018). The above norm is created by extending the scaled latent trace norm (Wimalawarne et al., 2014) with the addition of the concatenation of M to the unfolded latent tensor $\mathcal{T}^{(1)}$. The use of unfolded latent tensors on each mode allows the norm to regularize the rank of each mode independent of the ranks of other modes compared to overlapped regularization. For an example, if all the modes of a tensor are full rank except one mode, the latent trace norm will be able to induce low-rankness with respect to low-rank mode since each mode is regularized independently using latent tensors. However, the overlapped trace norm will regularize all the modes equally to induce low-rankness. The use of latent tensors benefits the above coupled norm, equation 2.4, since concatenated regularization on the coupled mode does not depend on the ranks of the uncoupled modes. The division of latent tensor regularization by the inverse of its mode dimension leads to

the inducement of low-rankness with relative ranks with respect to mode dimensions (Wimalawarne et al., 2014).

In addition to the homogeneous regularization methods given above, mixed norms such as (S, O, O), (O, S, O), and (O, O, S) have also been proposed (Wimalawarne et al., 2018). An example of a mixed norm is

$$\|\mathcal{T}, M\|_{(\text{S,O,O})}^1 = \inf_{\mathcal{T}^{(1)} + \mathcal{T}^{(2)} = \mathcal{T}} \left(\frac{1}{\sqrt{n_1}} \|[T_{(1)}^{(1)}; M]\|_{\text{tr}} + \sum_{k=2}^{3} \|T_{(k)}^{(2)}\|_{\text{tr}} \right),$$

where the scaled latent trace norm is regularized on the coupled mode and the rest of the modes are regularized using the overlapped trace norm.

3 Limitations of Coupled Norms

Though coupled norms have favorable properties such as convexity and better performance in coupled tensor completion (Wimalawarne et al., 2018) compared to individual tensor completion, they are not optimal for coupled tensor completion. We identify two major limitations with coupled norms.

3.1 Lack of Control on Shared Low-Rankness. The basic design principle of coupled norms is to combine two tensor norms by having a single trace norm regularization on the concatenated unfolding of tensors along the coupled mode. The underlying assumption with this formulation is that the concatenation of unfolded tensors is low rank; in other words, we can have a low-rank factorization of the concatenated matrix. This indicates that both tensors have a common left component matrix indicating shared low-rankness along the coupled modes. Though this is a reasonable assumption, in practice, the degree of shared low-rankness needs to be controlled when regularizing a learning model. Since the coefficient in front of the regularization of concatenation of unfolded tensors is equal to one, similar to other trace norm regularizations for other modes in a coupled norm, it induces an equal amount of regularization for both coupled and uncoupled components. This makes existing coupled norms suboptimal. A better design of coupled norms with theoretical guarantees should be developed.

3.2 Inefficiency with Higher-Order Tensors. The coupled norms proposed by Wimalawarne et al. (2018) are confined to the overlapped trace norm and latent trace norms. Though these norms can be applied as low-rank-inducing norms for any tensor, they may not be efficient with higher-order tensors. The square norm proposed by Mu et al. (2014) has been shown to be more efficient as a low-rank-inducing norm for higher-order tensors. More specifically, for a higher-order tensor with K modes each of dimensions n with a multilinear rank of $(r, \ldots, r)$, the excess risk bound using

the overlapped norm is bounded as $O(K\sqrt{r}(\sqrt{n^{K-1}} + \sqrt{n}))$ (Wimalawarne et al., 2018), while the use of the square norm leads to an excessive risk bound of $O(r^{\lfloor K/2 \rfloor} n^{\lceil K/2 \rceil})$ (see theorem 11 in the appendix and Mu et al., 2014). Thus, the existing coupled norms would not give the best performance for coupled higher orders, which creates a need to incorporate the square norm in coupled norms.

4 Proposed Methods

In this section, we propose new approaches to overcome the limitations we have described. We propose a new coupled completion model and discuss extensions to coupled norms.

First, we give the main problem that we investigate in this letter: generalized coupled tensor completion for higher-order tensors with control of regularization of coupled components. We define this problem by considering two partially observed tensors $\hat{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$, $K \geq 3$ and $\hat{\mathcal{Y}} \in \mathbb{R}^{n_1 \times n_2' \times \cdots \times n_{K'}'}$, $K' \geq 2$. Let m_1 and m_2 be the number of observed elements of $\hat{\mathcal{X}}$ and $\hat{\mathcal{Y}}$, respectively. We define $\Omega_{\hat{\mathcal{X}}} : \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K} \to \mathbb{R}^{m_1}$ and $\Omega_{\hat{\mathcal{Y}}} : \mathbb{R}^{n_1 \times n_2' \times \cdots \times n_{K'}'} \to \mathbb{R}^{m_2}$ as mappings to observed elements. Then the proposed coupled completion model is

$$\min_{\mathcal{X}, \mathcal{Y}} \frac{1}{2} \|\Omega_{\hat{\mathcal{X}}}(\mathcal{X} - \hat{\mathcal{X}})\|_F^2 + \frac{1}{2} \|\Omega_{\hat{\mathcal{Y}}}(\mathcal{Y} - \hat{\mathcal{Y}})\|_F^2 + \lambda \|\mathcal{X}, \mathcal{Y}\|_{\text{hcn}}^{(1,\gamma)}, \tag{4.1}$$

where $\|\mathcal{X}, \mathcal{Y}\|_{\text{cn}}^{(a,\gamma)}$ is an extended definition of coupled norms, with γ indicating scaling of the concatenated regularization on the coupled mode a. In sections 4.1 and 4.2, we propose our extensions to coupled norms with scaling to previously defined coupled norms (Wimalawarne et al., 2018) with three-mode tensors and higher-order coupled tensors.

4.1 Scaled Coupled Norms. We propose to explicitly control the regularization on concatenated components on the coupled mode. We achieve this by introducing a scaling factor $\gamma \in \mathbb{R}_+$ for the coupled regularization of the coupled norm. To include the scaling parameter, we extend the existing definition of the coupled norm as $\| \cdot \|_{(b,c,d)}^{(a,\gamma)}$, which we hereafter refer to as a *scaled coupled norm*, where the superscript (a, γ) indicates that the regularization on components on coupled mode a is scaled by γ.

Using the new definition, we redefine all the coupled norms in Wimalawarne et al. (2018). As an example, we can show the norm $\|\mathcal{T}, M\|_{(\text{O},\text{O},\text{O})}^{(1,\gamma)}$ for a three-mode tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $M \in \mathbb{R}^{n_1 \times n_2'}$ as follows:

$$\|\mathcal{T}, M\|_{(\text{O},\text{O},\text{O})}^{(1,\gamma)} := \gamma \|[T_{(1)}; M]\|_{\text{tr}} + \sum_{j=2}^{3} \|T_{(j)}\|_{\text{tr}}. \tag{4.2}$$

If $\gamma = 1$, then scaled coupled norms coincide with original coupled norms. In practice, the optimal scaling parameter for γ needs to be selected using an appropriate parameter selection method such as cross-validation. Though this creates an additional computational cost, we show theoretically (see section 5.1) and experimentally (see section 6.1) that this scaling leads to better performance.

We can also regularize each of the trace norms separately (e.g., the right-hand side of equation 4.2 can be $\gamma_1 \|[T_{(1)}; M]\|_{\mathrm{tr}} + \gamma_2 \sum_{j=2}^{3} \|T_{(j)}\|_{\mathrm{tr}}$), which would add more computational costs to solve the completion model. Our definition of scaled coupled norms is more convenient during optimization due to fewer parameters and also helps in theoretical analysis and interpretation, as we show in section 5.

4.2 Completion of Coupled Higher-Order Tensors. Now we propose coupled norms for higher-order tensors by combining the square norm with the coupled norms.

We first consider a higher-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ with $K \geq 4$ coupled with another tensor, $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2' \times \cdots \times n_{K'}'}$ with $K' \geq 2$. Without losing generality, we assume that they are coupled on a specific mode a. We introduce a modified scaled coupled norm notation as

$$\|\mathcal{X}, \mathcal{Y}\|_{(b,c),(d,e,f)}^{(a,\gamma)}, \tag{4.3}$$

where a indicates the coupled mode, and (b, c) and (d, e, f) indicate regularization methods for each tensor. As defined in Wimalawarne et al. (2018), if $\mathcal{Y}$ is a three-mode tensor, then $c, d, e, f \in \{\mathrm{O, L, S, -}\}$. The notation $b := [v]$ indicates that the particular higher-order tensor ($\mathcal{X}$ in this case) should be regularized using the square norm (Mu et al., 2014), given as

$$\|\mathcal{X}_{[v]}\|_{\mathrm{tr}} = \left\| \mathrm{reshape}\left(X_{(1)}, \prod_{i=1}^{v} n_i, \prod_{j=v+1}^{K} n_j \right) \right\|_{\mathrm{tr}},$$

where v specifies the modes to be considered for reshaping and the function reshape() (Mu et al., 2014) reshapes the tensor to a matrix of dimensions $n_1 \ldots n_v \times n_{v+1} \ldots n_K$. Further, the second parameter c in the subscript indicates how the tensor $\mathcal{X}$ should be regularized coupled to $\mathcal{Y}$.

4.2.1 Coupled Norms for a Higher-Order Tensor and a Matrix. Let us look at a few examples of possible higher-order coupled norms that we can construct using the new extension, equation 4.3. Given that $\mathcal{Y}$ is a matrix $\mathcal{Y} := M \in \mathbb{R}^{n_1 \times n_2'}$, we can define a coupled norm using only the overlapped regularization as

$$\|\mathcal{X}, M\|_{([v],O),(O,O)}^{(1,\gamma)} := \gamma \|[X_{(1)}; M]\|_{\mathrm{tr}} + \|X_{[v]}\|_{\mathrm{tr}} + \|M\|_{\mathrm{tr}}. \qquad (4.4)$$

Now let us consider making the regularization with respect to the coupled mode less dependent on the regularization of other modes of the tensor. To achieve this, we can add latent tensor regularization to the coupled modes. By changing O to S in equation 4.4 for the tensor $\mathcal{Y}$, we have the following definition,

$$\|\mathcal{X}, M\|_{([v],S),(O,O)}^{(1,\gamma)} := \inf_{\mathcal{X}^{(1)}+\mathcal{X}^{(2)}=\mathcal{X}} \frac{\gamma}{\sqrt{n_1}} \|[X_{(1)}^{(1)}; M]\|_{\mathrm{tr}} + \|X_{[v]}^{(2)}\|_{\mathrm{tr}} + \|M\|_{\mathrm{tr}},$$

$$(4.5)$$

where the tensor $\mathcal{X}$ is considered as a summation of latent tensors $\mathcal{X}^{(1)}$ and $\mathcal{X}^{(2)}$ with $X^{(1)}$ coupled to M and $X^{(2)}$ regularized independently.

4.2.2 Coupled Norms for a Higher-Order Tensor and a Three-Way Tensor. If $\mathcal{Y}$ is a three-mode tensor $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2' \times n_3'}$, we can define several coupled norms using the overlapped regularization (notation O) and latent and scaled latent regularizations (notations L and S) for the three-mode tensor, while applying regularization to the $\mathcal{X}$ tensor using the square norm. Applying overlapped regularization and scaled latent regularization to $\|\mathcal{X}, \mathcal{Y}\|_{([v],O),(O,O,O)}^{(1,\gamma)}$ to $\mathcal{Y}$ gives us

$$\|\mathcal{X}, \mathcal{Y}\|_{([v],O),(O,O,O)}^{(1,\gamma)} := \gamma \|[X_{(1)}; Y_{(1)}]\|_{\mathrm{tr}} + \sum_{k=2}^{3} \|Y_{(k)}\|_{\mathrm{tr}} + \|X_{[v]}\|_{\mathrm{tr}}$$

and

$$\|\mathcal{X}, \mathcal{Y}\|_{([v],O),(S,S,S)}^{(1,\gamma)} = \inf_{\mathcal{Y}^{(1)}+\mathcal{Y}^{(2)}=\mathcal{Y}} \left(\frac{\gamma}{\sqrt{n_1}} \|[X_{(1)}; Y_{(1)}^{(1)}]\|_{\mathrm{tr}} \right.$$

$$\left. + \sum_{k=2}^{3} \frac{1}{\sqrt{n_k}} \|Y_{(k)}^{(2)}\|_{\mathrm{tr}} + \|X_{[v]}\|_{\mathrm{tr}} \right).$$

Additionally, we can create mixed norms by applying a mixture of regularizations to the three-mode tensor. Due to space limitations, we will give only one example of the mixed higher-order coupled norm $\|\mathcal{X}, \mathcal{Y}\|_{([v],O),(O,S,O)}^{(1,\gamma)}$, which is defined as

$$\|\mathcal{X}, \mathcal{Y}\|_{([v],O),(O,S,O)}^{(1,\gamma)} = \inf_{\mathcal{Y}^{(1)}+\mathcal{Y}^{(2)}=\mathcal{Y}} \left(\gamma \|[X_{(1)}; Y_{(1)}^{(1)}]\|_{\mathrm{tr}} + \frac{1}{\sqrt{n_2}} \|Y_{(2)}^{(2)}\|_{\mathrm{tr}} \right.$$

$$\left. + \|Y_{(3)}^{(1)}\|_{\mathrm{tr}} + \|X_{[v]}\|_{\mathrm{tr}} \right),$$

where the second mode of $\mathcal{Y}$ is regularized with the scaled latent trace norm and the rest of the modes are regularized using the overlapped trace norm.

4.2.3 Coupled Norms for Two Higher-Order Tensors. Let us now consider $\mathcal{Y}$ also as a higher-order tensor $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2' \times \cdots \times n_{K'}'}$, $K' \geq 4$. In this case, we propose a coupled norm with both tensors regularized using the square norm as $\|\mathcal{X}, \mathcal{Y}\|_{([v],O),([v'],O)}^{(1,\gamma)}$, where v' indicates the modes that are used for the square reshaping norm for $\mathcal{Y}$. Then the definition of this norm is

$$\|\mathcal{X}, \mathcal{Y}\|_{([v],O),([v'],O)}^{(1,\gamma)} := \gamma \|[X_{(1)}; Y_{(1)}]\|_{\mathrm{tr}} + \|X_{[v]}\|_{\mathrm{tr}} + \|Y_{[v']}\|_{\mathrm{tr}}. \tag{4.6}$$

We can also define coupled norms with latent trace norm regulation for higher-order tensors as

$$\|\mathcal{X}, \mathcal{Y}\|_{([v],L),([v'],O)}^{(1,\gamma)} := \inf_{\mathcal{X}^{(1)}+\mathcal{X}^{(2)}=\mathcal{X}} \gamma \|[X_{(1)}^{(1)}; Y_{(1)}]\|_{\mathrm{tr}} + \|X_{[v]}^{(2)}\|_{\mathrm{tr}} + \|Y_{[v']}\|_{\mathrm{tr}}$$

and

$$\|\mathcal{X}, \mathcal{Y}\|_{([v],L),([v'],L)}^{(1,\gamma)} := \inf_{\mathcal{X}^{(1)}+\mathcal{X}^{(2)}=\mathcal{X}} \inf_{\mathcal{Y}^{(1)}+\mathcal{Y}^{(2)}=\mathcal{Y}} \gamma \|[X_{(1)}^{(1)}; Y_{(1)}^{(2)}]\|_{\mathrm{tr}}$$
$$+ \|X_{[v]}^{(1)}\|_{\mathrm{tr}} + \|Y_{[v']}^{(2)}\|_{\mathrm{tr}}.$$

4.2.4 Coupled Norms for Tensors Coupled on Multiple Modes. Our proposed norms need not be restricted to coupling of tensors by a single mode. The proposed norms can be extended to tensors that are coupled by more than one mode by specifying the multiple modes on the superscript a of equation 4.3 and changing the definition of the norm accordingly. As an example, let us consider two higher-order tensors $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ and $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2' \times \cdots \times n_{K'}'}$, $K, K' \geq 4$, which are coupled on their first two modes. Using overlapped regularization for both the tensors, we can define the norm $\|\mathcal{X}, \mathcal{Y}\|_{([v],O),([v'],O)}^{((1,2),\gamma)}$ as

$$\|\mathcal{X}, \mathcal{Y}\|_{([v],O),([v'],O)}^{((1,2),\gamma)} := \gamma \|[X_{(1,2)}; Y_{(1,2)}]\|_{\mathrm{tr}} + \|X_{[v]}\|_{\mathrm{tr}} + \|Y_{[v']}\|_{\mathrm{tr}}, \tag{4.7}$$

where coupling with modes 1 and 2 leads to an unfolding of $\mathcal{X}$ and $\mathcal{Y}$ by combining modes 1 and 2 for the coupled regularization. Further, if we consider scaled latent norm regularization for both coupled tensors, we can define the following coupled norm:

$$\|\mathcal{X}, \mathcal{Y}\|_{([v],\mathrm{S}),([v'],\mathrm{S})}^{((1,2),\gamma)} := \inf_{\mathcal{X}^{(1)}+\mathcal{X}^{(2)}=\mathcal{X}} \ \inf_{\mathcal{Y}^{(1)}+\mathcal{Y}^{(2)}=\mathcal{Y}} \frac{\gamma}{\sqrt{n_1 n_2}} \|[X_{(1,2)}^{(1)}; Y_{(1,2)}^{(1)}]\|_{\mathrm{tr}}$$

$$+ \frac{1}{\sqrt{n_1 \cdots n_v}} \|X_{[v]}^{(2)}\|_{\mathrm{tr}} + \frac{1}{\sqrt{n_1 \cdots n_{v'}'}} \|Y_{[v']}^{(2)}\|_{\mathrm{tr}}. \quad (4.8)$$

4.3 Optimization of the Coupled Completion Model. The objective function in equation 4.1 can be solved using convex optimization methods. We have used the alternating direction method of multipliers (ADMM) method (Boyd, Parikh, Chu, Peleato, & Eckstein, 2011) to solve the above objective function. We do not give the details of the optimization procedure since it is similar to the optimization of coupled norms given in Wimalawarne et al. (2018).

5 Theoretical Analysis

In this section, we present a theoretical analysis of the proposed coupled completion models regularized by the scaled coupled norms and the higher-order coupled norms.

Taking a similar approach to that of Wimalawarne et al. (2018), we derive excess risk bounds (El-Yaniv & Pechyony, 2007) for coupled completion using the Rademacher complexity. For our analysis, we consider two tensors $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ and $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2' \times \cdots \times n_{K'}'}$ coupled on their first mode. We represent indexes of observed elements of $\mathcal{X}$ by the set P, where $(i_1, \ldots, i_K) \in \mathrm{P}$ refers to the element $\mathcal{X}_{i_1, \ldots, i_K}$. Similarly, the set Q represents the set of observed elements in $\mathcal{Y}$. Further, we consider observed elements separated as training and test sets as $\mathrm{P_{Train}}$, $\mathrm{Q_{Train}}$, $\mathrm{P_{Test}}$, and $\mathrm{Q_{Test}}$, such that $\mathrm{P} = \mathrm{P_{Train}} \cup \mathrm{P_{Test}}$ and $\mathrm{Q} = \mathrm{Q_{Train}} \cup \mathrm{Q_{Test}}$.

We denote any coupled norm by $\|\mathcal{W}, \mathcal{V}\|_{\mathrm{hcn}}^{(1,\gamma)}$ and the hypothesis class constrained by it with some constant B as $\mathsf{W} = \{\mathcal{W}, \mathcal{V} : \|\mathcal{W}, \mathcal{V}\|_{\mathrm{hcn}}^{(1,\gamma)} \le B\}$. Then the training error of the coupled completion model, equation 4.1, for a given hypothesis class W is expressed as

$$L(\mathcal{W}, \mathcal{V}) := \frac{1}{|\mathrm{P_{Train}}| + |\mathrm{Q_{Train}}|} \left[\sum_{(i_1, \ldots, i_K) \in \mathrm{P_{Train}}} l(\mathcal{X}_{i_1, \ldots, i_K}, \mathcal{W}_{i_1, \ldots, i_K}) \right.$$

$$\left. + \sum_{(j_1, \ldots, j_{K'}) \in \mathrm{Q_{Train}}} l(\mathcal{Y}_{j_1, \ldots, j_{K'}}, \mathcal{V}_{j_1, \ldots, j_{K'}}) \right], \quad (5.1)$$

where $l(a, b) = (a - b)^2$. When $K = 3$ and $K' = 2$, we deal with coupled three-mode tensors based on the scaled coupled norms proposed in section 4.1. When $K \ge 4$ and $K' \ge 2$, we consider completion of coupled higher-order tensors. We can also define the test error for a given hypothesis class

$W, \bar{L}(\mathcal{W}, \mathcal{V})$ as follows:

$$\bar{L}(\mathcal{W}, \mathcal{V}) := \frac{1}{|P_{\text{Test}}| + |Q_{\text{Test}}|} \left[\sum_{(i_1,\ldots,i_K) \in P_{\text{Test}}} l(\mathcal{X}_{i_1,\ldots,i_K}, \mathcal{W}_{i_1,\ldots,i_K}) \right.$$
$$\left. + \sum_{(j_1,\ldots,j_{K'}) \in Q_{\text{Test}}} l(\mathcal{Y}_{j_1,\ldots,j_{K'}}, \mathcal{V}_{j_1,\ldots,j_{K'}}) \right]. \tag{5.2}$$

Using the standard conditions, let $l(\cdot, \cdot)$ be a Λ-Lipschitz continuous loss function bounded by $\sup_{i_1,\ldots,i_K} |l(\mathcal{X}_{i_1,\ldots,i_K}, \mathcal{W}_{i_1,\ldots,i_K})| \leq b_l$ and $\sup_{j_1,\ldots,j_{K'}} |l(\mathcal{Y}_{j_1,\ldots,j_{K'}}, \mathcal{V}_{j_1,\ldots,j_{K'}})| \leq b_l$. We consider the special case where $|P_{\text{Train}}| = |P_{\text{Test}}| = |P|/4$ and $|Q_{\text{Train}}| = |Q_{\text{Test}}| = |Q|/4$. This leads to equal-sized training and test sets, $|P_{\text{Train}}| + |Q_{\text{Train}}| = |P_{\text{Test}}| + |Q_{\text{Test}}| = |P|/2 = |Q|/2 = d$, similar to the assumption made in Shamir and Shalev-Shwartz (2014). Following the transductive Rademacher complexity theory (El-Yaniv & Pechyony, 2007; Shamir & Shalev-Shwartz, 2014), the following excess risk bound holds with probability $1 - \delta$,

$$\bar{L}(\mathcal{W}, \mathcal{V}) - L(\mathcal{W}, \mathcal{V}) \leq 4R_{P,Q}(l \circ \mathcal{W}, l \circ \mathcal{V}) + b_l \left(\frac{11 + 4\sqrt{\log \frac{1}{\delta}}}{\sqrt{|P_{\text{Train}}| + |Q_{\text{Train}}|}} \right), \tag{5.3}$$

where $R_{P,Q}(l \circ \mathcal{W}, l \circ \mathcal{V})$ is the transductive Rademacher complexity (Wimalawarne et al., 2018; El-Yaniv & Pechyony, 2007; Shamir & Shalev-Shwartz, 2014), defined as follows,

$$R_{P,Q}(l \circ \mathcal{W}, l \circ \mathcal{V}) = \frac{1}{d} \mathbb{E}_\sigma \left[\sup_{\mathcal{W}, \mathcal{V} \in W} \sum_{(i_1,\ldots,i_K) \in P} \sigma_{i_1,\ldots,i_K} l(\mathcal{X}_{i_1,\ldots,i_K}, \mathcal{W}_{i_1,\ldots,i_K}) \right.$$
$$\left. + \sum_{(j_1,\ldots,j_{K'}) \in Q} \sigma_{j_1,\ldots,j_{K'}} l(\mathcal{Y}_{j_1,\ldots,j_{K'}}, \mathcal{V}_{j_1,\ldots,j_{K'}}) \right],$$

where Rademacher variables $\sigma_{i_1,\ldots,i_K}$ and $\sigma_{j_1,\ldots,j_{K'}}$ consist of values of $\{-1, 1\}$ with probability 0.5. Let us define $\Sigma \in \mathbb{R}^{n_1 \times n_2 \times \cdots n_K}$ with $\Sigma_{i_1,\ldots,i_K} = \sigma_{i_1,\ldots,i_K}$ if $(i_1, \ldots, i_K) \in P$ and $\Sigma_{i_1,\ldots,i_K} = 0$ otherwise, and $\Sigma' \in \mathbb{R}^{n_1 \times n'_2 \times \cdots n'_{K'}}$ with $\Sigma'_{j_1,\ldots,j_{K'}} = \sigma_{j_1,\ldots,j_{K'}}$ if $(j_1, \ldots, j_{K'}) \in Q$ and $\Sigma'_{j_1,\ldots,j_{K'}} = 0$ otherwise. Then using the assumption of Rademacher contraction principle (Meir & Zhang,

2003) we can bound $R_{P,Q}(l \circ \mathcal{W}, l \circ \mathcal{V})$ as follows:

$$R_{P,Q}(l \circ \mathcal{W}, l \circ \mathcal{V}) \leq \frac{\Lambda}{d} \mathbb{E}_\sigma \left[\sup_{\mathcal{W}, \mathcal{V} \in \mathsf{W}} \sum_{i_1, \ldots, i_K} \Sigma_{i_1, \ldots, i_K} \mathcal{W}_{i_1, \ldots, i_K} \right.$$

$$\left. + \sum_{j_1, \ldots, j_{K'}} \Sigma'_{j_1, \ldots, j_{K'}} \mathcal{V}_{j_1, \ldots, j_{K'}} \right]. \tag{5.4}$$

Using the primal-dual relationship of the norm $\| \cdot \|_{\mathrm{hcn}^\star}^{(1,\gamma)}$ we can further bound equation 5.4 as follows,

$$R_{P,Q}(l \circ \mathcal{W}, l \circ \mathcal{V}) \leq \frac{\Lambda}{d} \left[\mathbb{E}_\sigma \sup_{\mathcal{W}, \mathcal{V} \in \mathsf{W}} \|\mathcal{W}, \mathcal{V}\|_{\mathrm{hcn}}^{(1,\gamma)} \| \Sigma, \Sigma' \|_{\mathrm{hcn}^\star}^{(1,\gamma)} \right], \tag{5.5}$$

where $\| \cdot \|_{\mathrm{hcn}^\star}^{(1,\gamma)}$ is the dual norm of $\| \cdot \|_{\mathrm{hcn}}^{(1,\gamma)}$. Using equation 5.5, we bound the Rademacher complexities for each of the scaled coupled norms (see appendixes B and C), which we discuss next.

5.1 Excess Risk of Scaled Three-Mode Couple Norms. Excess risk bounds for three-mode tensors coupled with a matrix based on unscaled coupled norms have been derived in Wimalawarne et al. (2018). Since the scaling parameter γ affects only the concatenated components of the coupled norm, the excess risk bounds in Wimalawarne et al. (2018) can be easily updated for scaled norms (see appendix A). The updated Rademacher complexities $(R_{P,Q}(l \circ \mathcal{W}, l \circ \mathcal{V})$ in equation 5.5 for completion of a coupled three-mode tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and a matrix $M := \mathcal{Y} \in \mathbb{R}^{n_1 \times n_2'}$ are shown in Table 1. We show only the $\| \cdot \|_{(O,O,O)}^{(1,\gamma)}$, $\| \cdot \|_{(S,S,S)}^{(1,\gamma)}$, and $\| \cdot \|_{(S,O,O)}^{(1,\gamma)}$ norms due to space limitations.

In Table 1, the parameter γ scales the induced low-rankness related to the rank of the coupled unfolding $\sqrt{r_{(1)}}$. Note that the parameter γ inversely scales the components $\gamma^{-1}C_2(\sqrt{n_1} + \sqrt{\prod_{j=2}^3 n_j + n_2'})$ and $\gamma^{-1}C_2(n_1 + \sqrt{\prod_{i=1}^3 n_i + n_1 n_2'})$. This behavior of scaling with γ tells us that if $0 < \gamma < 1$, then the shared low-rankness among the two tensors on the coupled mode is less, and excess risk would be bonded with larger terms of $\gamma^{-1}C_2(\sqrt{n_1} + \sqrt{\prod_{j=2}^3 n_j + n_2'})$ and $\gamma^{-1}C_2(n_1 + \sqrt{\prod_{i=1}^3 n_i + n_1 n_2'})$. On the other hand, if $\gamma > 1$, it indicates more shared low-rankness among coupled tensors, and the maximum term would select a smaller value. This analysis allows us to conclude that in order to obtain less excess risk, coupled tensor should share an adequate amount of low-rankness on the coupled mode.

Table 1: Rademacher Complexity Bounds of Scaled Coupled Norms for Three-Mode Tensors.

Norm	Rademacher Complexity $R_{P,Q}(l \circ \mathcal{W}, l \circ \mathcal{V})$
$\lVert \cdot \rVert_{(O,O,O)}^{(1,\gamma)}$	$\frac{3\Lambda}{2d}\left[\gamma\sqrt{r_{(1)}}(B_{\mathcal{W}} + B_{\mathcal{V}}) + \sum_{k=2}^{3}\sqrt{r_k}B_{\mathcal{V}} \right]$
	$\max\left\{ \gamma^{-1}C_2\left(\sqrt{n_1} + \sqrt{\prod_{j=2}^{3}n_j + n_2'}\right), \right.$
	$\left. \min_{k \in 2,3} C_1\left(\sqrt{n_k} + \sqrt{\prod_{j \neq k}^{3}n_j}\right) \right\}$
$\lVert \cdot \rVert_{(S,S,S)}^{(1,\gamma)}$	$\frac{3\Lambda}{2d}\left[\gamma\sqrt{\frac{r_{(1)}}{n_1}}(B_{\mathcal{W}} + B_{\mathcal{V}}) + \min_{k \in 2,3}\sqrt{\frac{r_k}{n_k}}B_{\mathcal{V}} \right]$
	$\max\left\{ \gamma^{-1}C_2\left(n_1 + \sqrt{\prod_{i=1}^{3}n_i + n_1 n_2'}\right), \right.$
	$\left. C_1 \max_{k=2,3}\left(n_k + \sqrt{\prod_{i \neq k}^{3}n_i}\right) \right\}$
$\lVert \cdot \rVert_{(S,O,O)}^{(1,\gamma)}$	$\frac{3\Lambda}{2d}\left[\gamma\sqrt{\frac{r_{(1)}}{n_1}}(B_{\mathcal{W}} + B_{\mathcal{V}}) + \sum_{i=2,3}\sqrt{r_i}B_{\mathcal{V}} \right]$
	$\max\left\{ \gamma^{-1}C_2\left(n_1 + \sqrt{\prod_{i=1}^{3}n_i + n_1 n_2'}\right), \right.$
	$\left. \min_{k=2,3} C_1\left(\sqrt{n_k} + \sqrt{\prod_{i \neq k}^{3}n_i}\right) \right\}$

Note: Coupled completion of $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2'}$ resulting in a hypothesis class $\mathsf{W} = \{\mathcal{W}, \mathcal{V} : \lVert \mathcal{W}, \mathcal{V} \rVert_{\mathrm{hcn}}^{(1,\gamma)} \leq B\}$ where $\lVert \cdot \rVert_{\mathrm{hcn}}^{(1,\gamma)}$ is any of the three-mode tensor-based coupled norms. The multilinear rank of $\mathcal{W}$ is (r_1, r_2, r_3), and the rank on coupled mode unfolding is $r_{(1)}$. $B_{\mathcal{W}}, B_{\mathcal{V}}, C_1$, and C_2 are constants.

5.2 Excess Risk of Coupled Higher-Order Tensors. Now we look into excess risk bounds for completion models regularized by the proposed higher-order coupled norms. Due to the large number of coupled norms we can define using the proposed norm, we analyze excess risk bounds for only a few coupled norms in this section.

The following theorem gives the Rademacher complexity for a coupling between a K-mode tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ with $K \geq 4$ and a matrix $\mathcal{Y} := M \in \mathbb{R}^{n_1 \times n_2'}$.

Theorem 1. *Let $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ and $\mathcal{Y} := M \in \mathbb{R}^{n_1 \times n_2'}$ be coupled on their first modes with sets of observed elements* P *and* Q*, respectively, with* $|P|/2 = |Q|/2 = d$*. Given a hypothesis class* $\mathsf{W} = \{\mathcal{W}, \mathcal{V} : \lVert \mathcal{W}, \mathcal{V} \rVert_{([v],O),(O,O)}^{(1,\gamma)} \leq B\}$*, the coupled completion using* $\lVert \mathcal{W}, \mathcal{V} \rVert_{([v],O),(O,O)}^{(1,\gamma)}$ *leads to the following Rademacher complexity of equation 5.5 with probability* $1 - \delta$*,*

$$R_{\mathrm{P},\mathrm{Q}}(l \circ \mathcal{W}, l \circ \mathcal{V})$$

$$\leq \frac{3\Lambda}{2d}\left[\gamma\sqrt{r_{(1)}}(B_{\mathcal{W}}+B_{\mathcal{V}}) + \min\left(\sqrt{\prod_{i=1}^{v}r_i}, \sqrt{\prod_{j=v+1}^{K}r_j}\right)B_{\mathcal{W}}+\sqrt{r'}B_V\right]$$

$$\min\left\{\gamma^{-1}C_1\left(\sqrt{n_1}+\sqrt{\prod_{j=2}^{K}n_j+n_2'}\right), \max\left\{C_2\left(\sqrt{\prod_{i=1}^{v}n_i}+\sqrt{\prod_{j=v+1}^{K}n_j}\right),\right.\right.$$

$$\left.\left. C_3\left(\sqrt{n_1}+\sqrt{n_2'}\right)\right\}\right\}.$$

where $(r_1, \ldots, r_K)$ is the multilinear rank of $\mathcal{W}$, r' is the rank of $\mathcal{V}$, $r_{(1)}$ is the rank of the coupled unfolding on the first mode, $B_{\mathcal{W}}$, $B_{\mathcal{V}}$, C_1, C_2, and C_3 are constants.

Next, three theorems consider a coupling between a K-mode tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ with $K \geq 4$ and a three-mode tensor $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2' \times n_3'}$.

Theorem 2. *Let $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ and $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2' \times n_3'}$ be coupled on their first modes with sets of observed elements* P *and* Q*, respectively, with* $|\mathrm{P}|/2 = |\mathrm{Q}|/2 = d$. *Given a hypothesis class* $\mathbf{W} = \{\mathcal{W}, \mathcal{V} : \|\mathcal{W}, \mathcal{V}\|_{([v],\mathrm{O}),(\mathrm{O},\mathrm{O},\mathrm{O})}^{(1,\gamma)} \leq B\}$, *the coupled completion using* $\|\mathcal{W}, \mathcal{V}\|_{([v],\mathrm{O}),(\mathrm{O},\mathrm{O},\mathrm{O})}^{(1,\gamma)}$ *leads to the following Rademacher complexity of equation 5.5 with probability* $1 - \delta$,

$$R_{\mathrm{P},\mathrm{Q}}(l \circ \mathcal{W}, l \circ \mathcal{V})$$

$$\leq \frac{3\Lambda}{2d}\left[\gamma\sqrt{r_{(1)}}(B_{\mathcal{W}}+B_{\mathcal{V}}) + \sum_{i=2}^{3}\sqrt{r_i'}B_{\mathcal{W}}+\min\left(\sqrt{\prod_{i=1}^{k}r_i}, \sqrt{\prod_{j=v+1}^{K}r_j}\right)B_{\mathcal{V}}\right]$$

$$\min\left\{C_1\gamma^{-1}\left(\sqrt{n_1}+\sqrt{\prod_{a=2}^{K}n_a+n_2'n_3'}\right), \max\left\{C_2\left(\sqrt{\prod_{i=1}^{v}n_i}+\sqrt{\prod_{j=v+1}^{K}n_j}\right),\right.\right.$$

$$\left.\left. \min\left\{C_4\left(\sqrt{n_2'}+\sqrt{n_1n_3'}\right), C_5\left(\sqrt{n_3'}+\sqrt{n_1n_2'}\right)\right\}\right\}\right\},$$

where $(r_1, \ldots, r_K)$ is the multilinear rank of $\mathcal{W}$; (r_1', r_2', r_3') is the multilinear rank of $\mathcal{V}$; $r_{(1)}$ is the rank of the coupled unfolding on the first mode; and $B_{\mathcal{W}}$, $B_{\mathcal{V}}$, C_1, C_2, C_4, and C_5 are constants.

Theorem 3. *Let $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ and $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2' \times n_3'}$ be coupled on their first modes with set of observed elements* P *and* Q*, respectively, with* $|\mathrm{P}|/2 = |\mathrm{Q}|/2 = d$.

Given a hypothesis class $W = \{\mathcal{W}, \mathcal{V} : \|\mathcal{W}, \mathcal{V}\|_{([v],O),(S,S,S)}^{(1,\gamma)} \leq B\}$, the coupled completion using $\|\mathcal{W}, \mathcal{V}\|_{([v],O),(S,S,S)}^{(1,\gamma)}$ leads to the following Rademacher complexity of equation 5.5 with probability $1 - \delta$,

$$R_{P,Q}(l \circ \mathcal{W}, l \circ \mathcal{V})$$

$$\leq \frac{3\Lambda}{2d}\left[\gamma\sqrt{\frac{r_{(1)}}{n_1}}(B_\mathcal{W} + B_\mathcal{V}) + min\left(\sqrt{\prod_{i=1}^{v} r_i}, \sqrt{\prod_{j=v+1}^{K} r_j}\right)B_\mathcal{W}\right.$$

$$+ min\left(\sqrt{\frac{r_2'}{n_2}}, \sqrt{\frac{r_3'}{n_3}}\right)B_\mathcal{V}\right] max\left\{\gamma^{-1}C_1\left(n_1 + \sqrt{\prod_{a=1}^{K} n_a + n_1 n_2' n_3'}\right),\right.$$

$$\left. C_2\left(\sqrt{\prod_{i=1}^{v} n_i} + \sqrt{\prod_{j=v+1}^{K} n_j}\right), C_4\left(n_2' + \sqrt{n_1 n_2' n_3'}\right), C_5\left(n_3' + \sqrt{n_1 n_2' n_3'}\right)\right\},$$

where $(r_1, \ldots, r_K)$ is the multilinear rank of $\mathcal{W}$; (r_1', r_2', r_3') is the multilinear rank of $\mathcal{V}$; $r_{(1)}$ is the rank of the coupled unfolding on the first mode; and $B_\mathcal{W}, B_\mathcal{V}, C_1, C_2, C_4,$ and C_5 are constants.

Inspection of the bounds of theorem 1 to 3 leads us to similar conclusions as in section 5.1 that more shared low-rankness among the coupled tensors leads to lower excess risks due to the scaling of γ. In order to make easier comparisons, we consider a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ with $n_1 = n_2 = \cdots = n_K = n$ and matrix $M \in \mathbb{R}^{n_1 \times n_2'}$ with $n_1 = n_2' = n$. Assuming that the multilinear rank of $\mathcal{T}$ is $(r, \ldots, r)$, the rank of M is r, and the rank on the concatenation of unfolded tensors is also r, then the Rademacher complexity of $\|\cdot\|_{([v],O),(O,O)}^{(1,\gamma)}$ is bounded by $O((\gamma\sqrt{r} + r^{\lceil K/2\rceil})n^{\lfloor K/2\rfloor})$ given that γ is sufficiently large ($\gamma < n^{(K/4)}$).

Theorem 4. *Let $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ and $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2' \times n_3'}$ be coupled on their first modes with the set of observed elements P and Q, respectively, with $|P|/2 = |Q|/2 = d$. Given a hypothesis class $W = \{\mathcal{W}, \mathcal{V} : \|\mathcal{W}, \mathcal{V}\|_{([v]),(S,O,O)}^{(1,\gamma)} \leq B\}$, the coupled completion using $\|\mathcal{W}, \mathcal{V}\|_{([v],O),(S,O,O)}^{(1,\gamma)}$ leads to the following Rademacher complexity of equation 5.5 with probability $1 - \delta$,*

$$R_{P,Q}(l \circ \mathcal{W}, l \circ \mathcal{V})$$

$$\leq \frac{3\Lambda}{2d}\left[\gamma\sqrt{\frac{r_{(1)}}{n_1}}(B_\mathcal{W} + B_\mathcal{V}) + \sum_{a=2}^{3}\sqrt{r_a'}B_\mathcal{V} + min\left(\sqrt{\prod_{i=1}^{v} r_i}, \sqrt{\prod_{j=v+1}^{K} r_j}\right)B_\mathcal{W}\right]$$

$$max \left\{ \gamma^{-1}C_1 \left(n_1 + \sqrt{\prod_{a=1}^{K} n_a + n_1 n_2' n_3'} \right), C_2 \left(\sqrt{\prod_{i=1}^{v} n_i} + \sqrt{\prod_{j=v+1}^{K} n_j} \right) \right.$$

$$\left. min \left(C_4 \left(\sqrt{n_2'} + \sqrt{n_1 n_3'} \right), C_5 \left(\sqrt{n_3'} + \sqrt{n_1 n_2'} \right) \right) \right\},$$

where $(r_1, \ldots, r_K)$ is the multilinear rank of $\mathcal{W}$; (r_1', r_2', r_3') is the multilinear rank of $\mathcal{V}$; $r_{(1)}$ is the rank of the coupled unfolding on the first mode; and $B_{\mathcal{W}}$, $B_{\mathcal{V}}$, C_1, C_2, C_4, and C_5 are constants.

Theorems 1, 2, and 3 also show that combining the square norm with the overlapped trace norm and the scaled latent norm leads to a lower Rademacher complexity. Again, if we consider the special case of $n_1 = n_2 = \cdots = n_K = n$ and $n_1 = n_2' = n_3' = n$, we have a Rademacher complexity of $\mathcal{O}((\gamma \sqrt{r_1} + r^{\lceil K/2 \rceil}) n^{\lceil K/2 \rceil})$. However, if we apply coupled norm in Wimalawarne et al. (2018), we end up with a larger Rademacher complexity of $\mathcal{O}((\gamma \sqrt{r_1} + \sqrt{r}) \sqrt{n^{K-1}})$. Hence, our proposed method leads to a better theoretical guarantee.

Finally, we consider the coupling of two higher-order tensors $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ and $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2' \times \cdots \times n_{K'}'}$ where $K, K' \geq 4$.

Theorem 5. *Let $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ and $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2' \times \cdots \times n_{K'}'}$ be coupled on their first modes with set of observed elements P and Q, respectively, with $|P|/2 = |Q|/2 = d$. Given a hypothesis class $W = \{\mathcal{W}, \mathcal{V} : \|\mathcal{W}, \mathcal{V}\|_{([v],O),([v'],O)}^{(1,\gamma)} \leq B\}$, the coupled completion using $\|\mathcal{W}, \mathcal{V}\|_{([v],O),([v'],O)}^{(1,\gamma)}$ leads to the following Rademacher complexity of equation 5.5 with probability $1 - \delta$,*

$$R_{P,Q}(l \circ \mathcal{W}, l \circ \mathcal{V})$$

$$\leq \frac{3\Lambda}{2d} \left[\gamma \sqrt{r_{(1)}}(B_{\mathcal{W}} + B_{\mathcal{V}}) + min \left(\sqrt{\prod_{i1=1}^{v} r_{i1}}, \sqrt{\prod_{j1=v+1}^{K} r_{j1}} \right) B_{\mathcal{W}} \right.$$

$$\left. + min \left(\sqrt{\prod_{i2=1}^{v'} r_{i2}'}, \sqrt{\prod_{j2=v'+1}^{K'} r_{j2}'} \right) B_{\mathcal{V}} \right] min \left\{ \gamma^{-1}C_6 \left(\sqrt{n_1} + \sqrt{\prod_{a=2}^{K} n_a + \prod_{b=2}^{K} n_b'} \right), \right.$$

$$\left. max \left\{ C_2 \left(\sqrt{\prod_{i=1}^{v} n_i} + \sqrt{\prod_{j=v+1}^{K} n_j} \right), C_7 \left(\sqrt{\prod_{i'=1}^{v'} n_{i'}'} + \sqrt{\prod_{j'=v'+1}^{K'} n_{j'}'} \right) \right\} \right\},$$

where $(r_1, \ldots, r_K)$ is the multilinear rank of $\mathcal{W}$; $(r_1', \ldots, r_{K'}')$ is the multilinear rank of $\mathcal{V}$; $r_{(1)}$ is the rank of the coupled unfolding on the first mode; and $B_{\mathcal{W}}$; $B_{\mathcal{V}}$, C_2, C_6, and C_7 are constants.

We can draw a similar conclusion for coupling two higher-order tensors as in previous theorems that the proposed extension leads to a lower Rademacher complexity compared to applying coupled norms in Wimalawarne et al. (2018). If we extend norms from Wimalawarne et al. (2018) or section 4.1 for higher-order tensors (e.g., $\| \cdot \|_{(O,O,...,O)}^{(1,\gamma)}$), the excess risk will be bounded by a larger term, such as $O(K\sqrt{r}(\sqrt{n^{K-1}} + \sqrt{n}))$, which is larger than the excess risk bounds achievable from theorem 5. This indicates that the integration of the square norm to coupled norms leads to better performance for coupled higher-order tensors.

Finally, we point out that Rademacher complexities for all coupled norms are bounded by $O(1/d)$, where d is the total number of observed elements of both coupled tensors. If completion of tensors were performed separately, the resulting Rademacher complexities for each tensor would be bounded with respect to the number of observed elements of that tensor. Since the Rademacher complexity is bounded by $1/d$, it may lead to lower bounds compared to the sum of individual Rademacher complexities for each tensor. Furthermore, since we used the transductive Rademacher complexity analysis, we obtained faster rates of decrease by $1/d$ compared to an analysis under inductive settings (Shamir & Shalev-Shwartz, 2014), which could lead to a bounding by $1/\sqrt{d}$.

6 Experiments

In this section, we present details of simulation experiments that we carried out for coupled tensor completion.

6.1 Simulation Experiments. We organized our simulation experiments into two sections. In the first section, we give a simulation experiment based on scaled coupled norm regularized coupled completion models for a coupled three-mode tensor and a matrix. In the following section, we give simulation experiments to evaluate the proposed higher-order coupled norms for coupled higher-order tensor completion.

6.1.1 Experiment with Coupled a Three-Mode Tensor and a Matrix. To create coupled tensors for our simulations, we used a similar approach as in Wimalawarne et al. (2018). All our coupled tensors were created using multilinear ranks. To generate a K-mode tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ with multilinear rank $(r_1, \ldots, r_K)$, we created a core tensor $\mathcal{C} \in \mathbb{R}^{r_1 \times \cdots \times r_K}$ sampled from a Normal distribution and orthogonal component matrices $U_i \in \mathbb{R}^{r_i \times n_i}$, $i = 1, \ldots, K$ and compute $\mathcal{X} = \mathcal{C} \times_1 U_1 \cdots \times_K U_K$ where $\times_k$ is the k-mode product (Kolda & Bader, 2009). We coupled two tensors $\mathcal{X}$ and $\mathcal{Y}$ along a mode a by sharing b left singular vectors on their mode-a, unfolding $X_{(a)} = M_1 P_1 N_1^\top$ and $Y_{(a)} = M_2 P_2 N_2^\top$ with $M_1(1:b, 1:n_a) = M_2(1:b, 1:n_a)$. We added noise sampled from a gaussian distribution of zero mean and variance of 0.01 to

(a) Matrix Completion (M)   (b) Tensor Completion ($\mathcal{T}$)
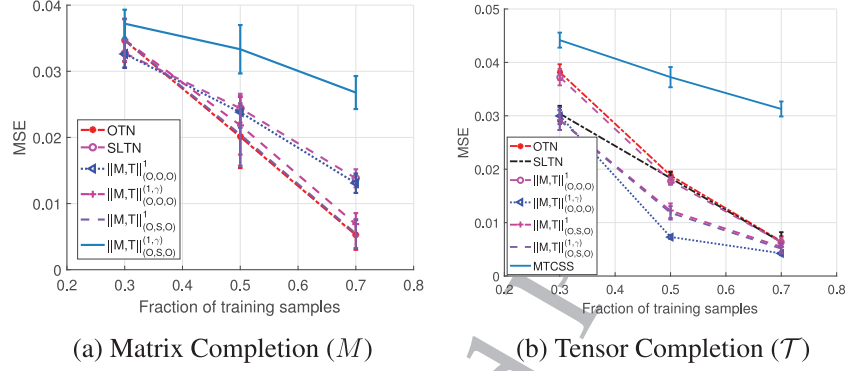
Figure 1: Performances of completion of a tensor with dimensions of $20 \times 20 \times 20$ with a multilinear rank of $(15, 5, 5)$ and a matrix with dimensions of 20×30 with a rank of 5.

all elements of the tensors. We randomly sampled training sets of percentages of 30, 50, and 70 from the total number of elements of each tensor and another 10% as validation sets. The remaining elements were taken as test sets. We repeated the experiments with three random selections and calculated the mean squared error (MSE) on the test data.

For our simulation experiment in this section, we created a three-mode tensor $\mathcal{T} \in \mathbb{R}^{20 \times 20 \times 20}$ and a matrix $M \in \mathbb{R}^{20 \times 30}$ coupled on their first modes. We specified the multilinear rank of $\mathcal{T}$ as $(15, 5, 5)$ and the rank of M as 5. We explicitly shared five left singular vectors among the tensor and the matrix on the coupled mode such that all the left singular vectors of the matrix are shared with the tensor. We cross-validated the regularization parameters from the range of 0.01 to 1 in intervals of 0.05 and the scaling parameters from the set $2^{-8}, 2^{-7}, \ldots, 2^{8}$.

Figure 1 shows the performance of the simulation experiment. We experimented with all the coupled norms for a three-mode tensor (Wimalawarne et al., 2018) and their scaled norms; however, for clear plotting, we show only the coupled norms and scaled coupled norms that gave the best performances for the coupled matrix and the tensor. As baseline methods, we used individual completion models regularized by the overlapped trace norm (OTN), the scaled latent trace norm (SLTN), and the matrix trace norm (MTN). As a further baseline method we used MTCSS proposed by Li et al. (2015).

Figure 1 shows that for matrix completion, none of the coupled norms outperformed individual matrix completion. The scaled coupled norm $\|M, \mathcal{T}\|_{(O,S,O)}^{(1,\gamma)}$ has performance equal to that of the matrix trace norm, while its unscaled norm has given a poor performance. For tensor completion,
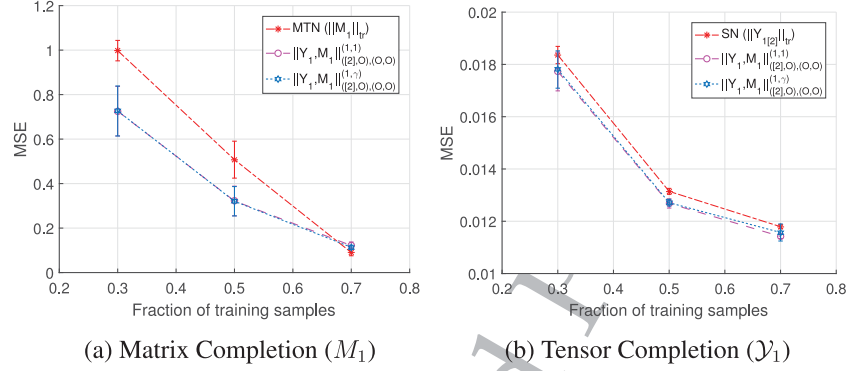
Figure 2: Coupled completion of a four-mode tensor and a matrix.

several norms, such as $\|M, \mathcal{T}\|_{(O,O,O)}^{(1,\gamma)}$, $\|M, \mathcal{T}\|_{(O,S,O)}^{1}$, and $\|M, \mathcal{T}\|_{(O,S,O)}^{(1,\gamma)}$, performed better than individual tensor completion by the overlapped trace norm and the scaled latent trace norm. In addition, the coupled norm $\|M, \mathcal{T}\|_{(O,O,O)}^{1}$ had weaker performance than individual tensor completion, while its scaled norm had the best performance, and the MTCSS method had poor performance compared to coupled norms.

6.1.2 Experiments with Higher-Order Coupled Norms. In this section, we consider four-mode tensors of dimensions $20 \times 20 \times 20 \times 20$ coupled to other tensors. We used the same procedure to create coupled tensors as in the section 6.1.1.

For all the experiments in this section, we used the coupled completion models regularized by higher-order norms introduced in section 4.2. To evaluate individual completion of the higher-order tensors, we used the square norm (SN; Mu et al., 2014. Further, we used the OTN and the SLTN for individual three-mode tensor completion and the matrix trace norm for individual matrix completion. For all models, we used regularization parameters from a range of 0.01 to 2 in intervals of 0.025 and scaling parameter set $2^{-8}, 2^{-7}, \ldots, 2^{8}$.

For our first simulation experiment with coupled higher-order tensors, we designed a coupled tensor with a four-mode tensor $\mathcal{Y}_1 \in \mathbb{R}^{20 \times 20 \times 20 \times 20}$ and a matrix $M_1 \in \mathbb{R}^{20 \times 20}$ coupled on their first modes. We specified the multilinear rank of the $\mathcal{Y}_1$ to be $(3, 6, 6, 6)$ and the rank of M_1 to be 3 where all the left singular vectors along mode-1 were shared among both the tensor and the matrix. For this experiment, we used the coupled norm $\|\mathcal{Y}_1, M_1\|_{([2],O),(O,O)}^{(1,\gamma)}$ for coupled completion. From Figure 2, we can see that both the scaled and unscaled norms of $\|\mathcal{Y}_1, M_1\|_{([2],O),(O,O)}^{(1,\gamma)}$ gave the best performances for both the matrix and the tensor completion.

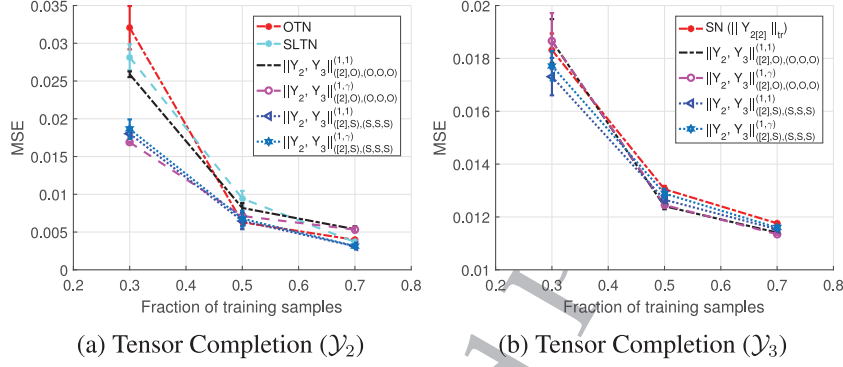(a) Tensor Completion ($\mathcal{Y}_2$)   (b) Tensor Completion ($\mathcal{Y}_3$)

Figure 3: Coupled completion of a four-mode tensor and a three-mode tensor.

Next, we look into a coupled tensor consisting of a four-mode tensor $\mathcal{Y}_2 \in \mathbb{R}^{20 \times 20 \times 20 \times 20}$ and a three-mode tensor $\mathcal{Y}_3 \in \mathbb{R}^{20 \times 20 \times 20}$. We specified the multilinear rank of $\mathcal{Y}_2$ to be $(3, 6, 6, 6)$ and the multilinear rank of $\mathcal{Y}_3$ to be $(3, 6, 6)$ with all the left singular vectors along mode-1 shared among the tensors. As baseline methods, we used the square norm for $\mathcal{Y}_2$ and the overlapped trace norm and the scaled latent trace norm for $\mathcal{Y}_3$. We experimented with different coupled norms that can be applied for coupled four-mode and three-mode tensors; however, for convenience, we plot in Figure 3 only results from the norms that gave the best performance. We observe that for $\mathcal{Y}_1$, the best performance is given by the scaled and unscaled norms of $\|\mathcal{Y}_2, \mathcal{Y}_3\|_{([2],O),(O,O,O)}$ and $\|\mathcal{Y}_2, \mathcal{Y}_3\|_{([2],S),(S,S,S)}$. For the tensor $\mathcal{Y}_2$, the coupled norms $\|\mathcal{Y}_2, \mathcal{Y}_3\|_{([2],O),(O,O,O)}^{(1,\gamma)}$, $\|\mathcal{Y}_2, \mathcal{Y}_3\|_{([2],S),(S,S,S)}^{(1,1)}$, and $\|\mathcal{Y}_2, \mathcal{Y}_3\|_{([2],S),(S,S,S)}^{(1,\gamma)}$ outperformed the OTN and the SLTN for individual tensor completion.

Finally, we look into two coupled four-mode tensors. Here, we considered two tensors $\mathcal{Y}_4 \in \mathbb{R}^{20 \times 20 \times 20 \times 20}$ and $\mathcal{Y}_5 \in \mathbb{R}^{20 \times 20 \times 20 \times 20}$. We constrained the multilinear ranks of $\mathcal{Y}_4$ and $\mathcal{Y}_5$ to be $(3, 6, 6, 6)$ and $(3, 8, 8, 8)$, respectively, and coupled them on their first modes by making all the left singular vectors common to both the tensors. We used the coupled norms $\|\mathcal{Y}_4, \mathcal{Y}_5\|_{([2],O),([2],O)}^{(1,1)}$ and $\|\mathcal{Y}_4, \mathcal{Y}_5\|_{([2],O),([2],O)}^{(1,\gamma)}$ for coupled completion. Additionally, we used the scaled overlapped norm extended from Wimalawarne et al. (2018) as $\|\mathcal{Y}_4, \mathcal{Y}_5\|_{(O,O,O,O)}^{(1,\gamma)}$, which indicates that both tensors are regularized with respect to each mode unfolding and the concatenated tensor unfolding on mode-1.

Figure 4 shows that both higher-order coupled norms outperformed individual tensor learning with the square norm. For the tensor $\mathcal{Y}_5$, the scaled higher-order norm $\|\mathcal{Y}_4, \mathcal{Y}_5\|_{([2],O),([2],O)}^{(1,\gamma)}$ further improved the performance compared to the unscaled norm. We can also see that $\|\mathcal{Y}_4, \mathcal{Y}_5\|_{(O,O,O,O)}^{(1,\gamma)}$ gave

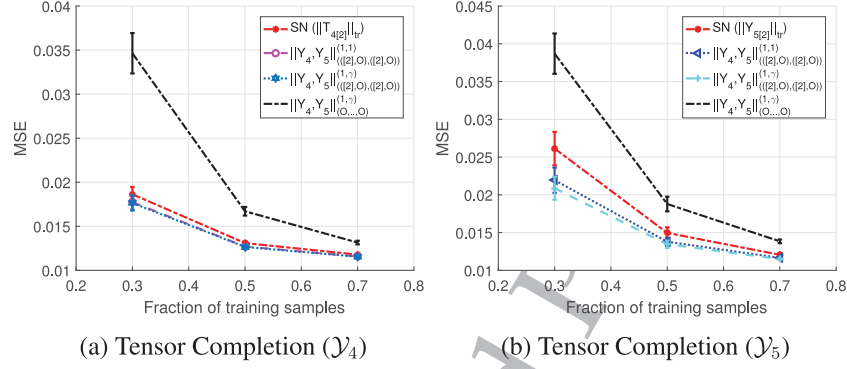(a) Tensor Completion ($\mathcal{Y}_4$) (b) Tensor Completion ($\mathcal{Y}_5$)

Figure 4: Coupled completion two four-mode tensors.

a weaker performance compared to coupled higher-order norms, agreeing with our theoretical analysis in section 5.

6.2 Multiview Video Completion Experiment. As a real-data experiment we applied our proposed methods for multiview video completion using the EPFL data set: multicamera pedestrian Videos data (Berclaz, Fleuret, Turetken, & Fua, 2011). The data set consists of movements of four people in a room captured from synchronized cameras. For our experiments, we used two videos; one video was considered to be corrupted and the other not. To create the video data, we sampled 50 frames with equal time splits from each video. We then downsampled each frame to a height and width of 76 and 102, respectively. The dimensions of both videos were the same $\mathcal{V}_1, \mathcal{V}_2 \in \mathbb{R}^{\text{frames} \times \text{channels} \times \text{width} \times \text{height}}$, where frames $= 50$, channels $= 3$ representing RBG color channels, width $= 76$, and height $= 102$. Since frames and RGB channels are common to both videos, we considered that the two videos are coupled on both of these modes. We considered the video $\mathcal{V}_1$ to be corrupted and sampled percentages of 10, 30, 50, and 70 of the total number of elements as its observed elements (training sets). From the remaining elements, we considered 10% of the total number of elements as validation sets; the rest were taken as test sets. To recover missing elements of the corrupted video, we completed it coupled to the uncorrupted video as side information using our proposed completion models regularized using higher-order coupled norms.

We completed coupled videos using the coupled norms $\|\mathcal{V}_1, \mathcal{V}_2\|_{([2],O),([2],O)}^{((1,2),\gamma)}$ in equation 4.7 and $\|\mathcal{V}_1, \mathcal{V}_2\|_{([2],S),([2],S)}^{((1,2),\gamma)}$ in equation 4.8. We performed individual completion of $\mathcal{V}_1$ using the square norm similar to the coupled norm using the first two modes to reshape the tensor as a matrix. We applied cross-validation on regularization parameters of 10^x ranging x from 0 to 7, with intervals of 0.25. We experimented with
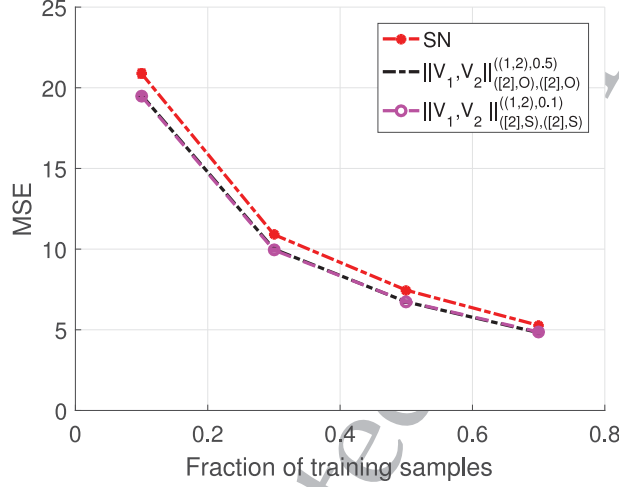
Figure 5: Performances of multiview video completion.

different values for γ ranging from 0.1 to 1, with intervals of 0.1. We found that best performance for $\|\mathcal{V}_1, \mathcal{V}_2\|_{([2],O),([2],O)}^{((1,2),\gamma)}$ was given, with $\gamma = 0.5$ and for $\|\mathcal{V}_1, \mathcal{V}_2\|_{([2],S),([2],S)}^{((1,2),\gamma)}$ with $\gamma = 0.1$. Figure 5 shows that the proposed coupled norms gave better performance compared to the individual tensor completion using square norm regularization. We provide further experimental results to compare the performances of the proposed methods with baseline methods in appendix D.

7 Conclusion

In this letter, we have investigated two limitations of coupled norms and proposed scaled coupled norms and coupled norms for higher-order tensors. Through theoretical analysis and experiments, we demonstrated that our proposed methods are more robust for coupled completion compared to existing coupled norms. However, we feel that coupled norms should be further investigated to be used widely in real-world applications.

One drawback of the scaling of coupled norms is that it requires more computations to find the optimal scaling parameters (γ). Though cross-validation can be employed to find the optimal scaling parameter, it can become computationally infeasible in real-world applications, especially with tensors with large dimensions. Future research in coupled norms should be directed toward finding better optimization strategies and parameter selection methods to overcome these computational issues. Further, our theoretical analysis was focused on excess risk bounds for tensor

completion. In future research, a more suitable yet rigorous theoretical analysis would be to derive exact recovery bounds (Yuan & Zhang, 2016) for coupled completion.

Appendix A: Dual Norms for Scaled Coupled Norms

The dual norms of our proposed coupled norms are important to prove excess risk bounds given in this letter. The dual norms of scaled coupled norms for three-mode tenors are similar to dual norms given in Wimalawarne et al. (2018). The only difference comes from the scaling parameter γ, which is inversely multiplied in the dual norm. For instance, the dual norm of $\|\mathcal{Y}, M\|_{(O,O,O)}^{(1,\gamma)}$ is

$$\|\mathcal{Y}, M\|_{(O,O,O)^\star}^{(1,\gamma)} = \max\{\gamma^{-1}\|[Y_{(1)}; M]\|_{op}, \|Y_{(2)}\|_{op}\|Y_{(3)}\|_{op}\}.$$

Dual norms of coupled higher-order tensors can also be derived by using a similar approach as in Wimalawarne et al. (2018). We give a brief overview of how to derive dual norms for higher-order coupled norms, starting with the dual norm of $\|\mathcal{X}, \mathcal{Y}\|_{([v],O),(O,O,O)}^{(1,\gamma)}$, which we derive next.

Theorem 6. *Let a matrix $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times n_4}$ and a tensor $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2' \times n_3'}$ be coupled on their first modes. The dual norm of $\|\mathcal{X}, \mathcal{Y}\|_{([v],O),(O,O,O)}^{(1,\gamma)}$ is*

$$\|\mathcal{X}, \mathcal{Y}\|_{([v],O),(O,O,O)^\star}^{(1,\gamma)} = \inf_{\mathcal{X}^{(1)}+\mathcal{X}^{(2)}=\mathcal{X}} \ \inf_{\mathcal{Y}^{(1)}+\mathcal{Y}^{(2)}+\mathcal{Y}^{(3)}=\mathcal{Y}} \ \max\left\{\gamma^{-1}\|[\mathcal{X}_{(1)}^{(1)}; \mathcal{Y}_{(1)}^{(1)}]\|_{op}, \right.$$
$$\left. \|\mathcal{X}_{[v]}^{(2)}\|_{op}, \|\mathcal{Y}_{(2)}^{(2)}\|_{op}, \|\mathcal{Y}_{(3)}^{(3)}\|_{op}\right\}.$$

Proof. We adopt the method of deriving dual norms for tensor norms in Tomioka and Suzuki (2013) and Wimalawarne et al. (2018) to derive the dual norm $\|\mathcal{X}, \mathcal{Y}\|_{([v],O),(O,O,O)^\star}^{(1,\gamma)}$. First, we derive the unscaled dual norm $\|\mathcal{X}, \mathcal{Y}\|_{([v],O),(O,O,O)^\star}^{(1,1)}$.

Let us consider a linear operator Φ similar to Tomioka and Suzuki (2013) and Wimalawarne et al. (2018) such that $z := \Phi(\mathcal{X}, \mathcal{Y}) = [\text{vec}([X_{(1)}; Y_{(1)}]); \text{vec}(X_{([v])}); \text{vec}(Y_{(2)}); \text{vec}(Y_{(3)})] \in \mathbb{R}^{2d_1+3d_2}$ where $d_1 = n_1 n_2 n_3 n_4$ and $d_2 = n_1 n_2' n_3'$.

Now, we consider the Schatten norm of the coupled norm $\|\mathcal{X}, \mathcal{Y}\|_{([v],O),(O,O,O),S_p^q}^{(1,1)}$, which is defined as

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

where $\mathcal{Z}_1^{(k)}$ is the inverse vectorization of elements $z_{((k-1)d_2+1):(kd_2)}$ for $k = 1, 2, 3$, and $\mathcal{Z}_2^{(1)}$ and $\mathcal{Z}_2^{(2)}$ are inverse vectorizations of $z_{(3d_1+1):(3d_1+d_2)}$ and $z_{(3d_1+d_2+1):(3d_1+2d_2)}$, respectively. Then the dual of the above norm is

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

where $1/p + 1/p^* = 1$ and $1/q + 1/q^* = 1$.

Since $\Phi^\top$ is the inverse operator of Φ Tomioka and Suzuki (2013), we find that

$$\Phi^\top(z) = \{\mathcal{X}, \mathcal{Y}\} = \left\{ \sum_{j=1}^{2} \mathcal{Z}_2^{(j)}, \sum_{i=1}^{3} \mathcal{Z}_1^{(i)} \right\}.$$

Then we have a norm such as

$$|||[\mathcal{X}, \mathcal{Y}]|||_{\overline{*}(\Phi)} = \inf_{\mathcal{Z}_1^{(1)} + \mathcal{Z}_1^{(2)} = \mathcal{X}} \inf_{\mathcal{Z}_2^{(1)} + \mathcal{Z}_2^{(2)} + \mathcal{Z}_2^{(3)} = \mathcal{Y}} \left(\|[\mathcal{Z}_{1(1)}^{(1)}; \mathcal{Z}_{2(1)}^{(1)}]\|_{S_p}^{q} + \|\mathcal{Z}_{1[v]}^{(2)}\|_{S_p}^{q} \right.$$
$$\left. + \|\mathcal{Z}_{2(2)}^{(2)}\|_{S_p}^{q} + \|\mathcal{Z}_{2(3)}^{(3)}\|_{S_p}^{q} \right)^{1/q}. \tag{A.2}$$

Now if we take $|||\mathcal{X}, \mathcal{Y}|||_{*(\Phi)} = \|\mathcal{X}, \mathcal{Y}\|_{([v],O),(O,O,O),S_p^q}^{(1,1)}$ and

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

then from lemma 3 in Tomioka and Suzuki (2013), we can conclude that the dual norm of $\|\mathcal{X}, \mathcal{Y}\|_{([v],O),(O,O,O)}^{(1,1)}$ is $|||[\mathcal{X}, \mathcal{Y}]|||_{\overline{*}(\Phi)^\star}$, as given in equation A.2.

If we consider the case special where $p = 1$ and $q = 1$, then we will obtain the following dual norm:

$$\|\mathcal{X}, \mathcal{Y}\|_{([v],O),(O,O,O)^\star}^{(1,1)} = \inf_{\mathcal{X}^{(1)} + \mathcal{X}^{(2)} = \mathcal{X}} \inf_{\mathcal{Y}^{(1)} + \mathcal{Y}^{(2)} + \mathcal{Y}^{(3)} = \mathcal{Y}} \max \left\{ \|[\mathcal{X}_{(1)}^{(1)}; \mathcal{Y}_{(1)}^{(1)}]\|_{\mathrm{op}}, \right.$$
$$\left. \|\mathcal{X}_{[v]}^{(2)}\|_{\mathrm{op}}, \|\mathcal{Y}_{(2)}^{(2)}\|_{\mathrm{op}}, \|\mathcal{Y}_{(3)}^{(3)}\|_{\mathrm{op}} \right\}.$$

Using the duality relationship (Boyd & Vandenberghe, 2004), we find that the scaled norm $\|\mathcal{X}, \mathcal{Y}\|_{([v],O),(O,O,O)^\star}^{(1,\gamma)}$ has the following dual norm:

$$\|\mathcal{X}, \mathcal{Y}\|_{([v],O),(O,O,O)^\star}^{(1,\gamma)} = \inf_{\mathcal{X}^{(1)} + \mathcal{X}^{(2)} = \mathcal{X}} \inf_{\mathcal{Y}^{(1)} + \mathcal{Y}^{(2)} + \mathcal{Y}^{(3)} = \mathcal{Y}} \max \left\{ \gamma^{-1} \|[\mathcal{X}_{(1)}^{(1)}; \mathcal{Y}_{(1)}^{(1)}]\|_{\mathrm{op}}, \right.$$
$$\left. \|\mathcal{X}_{[v]}^{(2)}\|_{\mathrm{op}}, \|\mathcal{Y}_{(2)}^{(2)}\|_{\mathrm{op}}, \|\mathcal{Y}_{(3)}^{(3)}\|_{\mathrm{op}} \right\}.$$

$\square$

We can use theorem 6 to deduce dual norms for other norms. For example, if we consider $\|\mathcal{X}, \mathcal{Y}\|_{([v],O),([v],O)^\star}^{(1,\gamma)}$, we can extend theorem 6 to arrive at the dual norm,

$$\|\mathcal{X}, \mathcal{Y}\|_{([v],O),([v],O)^\star}^{(1,\gamma)} = \inf_{\mathcal{X}^{(1)}+\mathcal{X}^{(2)}=\mathcal{X}} \inf_{\mathcal{Y}^{(1)}+\mathcal{Y}^{(2)}=\mathcal{Y}} \max \left\{ \gamma^{-1} \| [\mathcal{X}_{(1)}^{(1)}; \mathcal{Y}^{(1)}] \|_{\mathrm{op}}, \right.$$
$$\left. \|\mathcal{X}_{[v]}^{(2)}\|_{\mathrm{op}}, \|\mathcal{Y}_{[v]}^{(2)}\|_{\mathrm{op}} \right\}.$$

Similarly, dual norms of other scaled higher-order coupled norms can be derived.

Appendix B: Excess Risk Bounds for Coupled Three-Mode Tensor Completion

The excess risk bounds for coupled completion using scaled coupled norms given in section 5.1 can be derived in an identical way, as Wimalawarne et al. (2018) proved for unscaled norms. As a guide for proof and completeness, we give the detailed proof of excess risk bounds for the norm $\| \cdot \|_{(O,O,O)}^{(1,\gamma)}$.

Theorem 7. *Let $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{Y} := M \in \mathbb{R}^{n_1 \times n_2'}$ be coupled on their first modes with sets of observed elements P and Q, respectively, with $|P|/2 = |Q|/2 = d$. Given a hypothesis class $\mathbf{W} = \{\mathcal{W}, W_M : \|\mathcal{W}, \mathcal{W}_M\|_{(O,O,O)}^{(1,\gamma)} \le B\}$, the coupled completion using $\|\mathcal{W}, \mathcal{V}\|_{(O,O,O)}^{(1,\gamma)}$ leads to the following Rademacher complexity of equation 5.5 with probability $1 - \delta$,*

$$R_{P,Q}(l \circ \mathcal{W}, l \circ \mathcal{V})$$
$$\le \frac{3\Lambda}{2d} \left[\gamma^{-1} \sqrt{r_{(1)}}(B_\mathcal{W} + B_M) + \sum_{k=2}^{3} \sqrt{r_k} B_\mathcal{W} \right]$$
$$\max \left\{ C_2 \left(\sqrt{n_1} + \sqrt{\prod_{j=2}^{3} n_j + n_2'} \right), \min_{k \in 2,3} C_1 \left(\sqrt{n_k} + \sqrt{\prod_{j \neq k}^{3} n_j} \right) \right\}.$$

where (r_1, r_2, r_3) is the multilinear rank of $\mathcal{W}$; r' is the rank of M; $r_{(1)}$ is the rank of the coupled unfolding on the first mode; and $B_M, B_\mathcal{W}, C_1, C_2,$ and C_3 are constants.

Proof. We first bound $\|\mathcal{W}, W_M\|_{(O,O,O)}^{(1,\gamma)}$ as

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

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, $\|\mathcal{W}_M\|_F \leq B_M$, and $\|\mathcal{W}\|_F \leq B_{\mathcal{W}}$.

To bound $\mathbb{E}\|\Sigma, \Sigma'\|^{(1,\gamma)}_{(O,O,O)^*}$, we use the following duality relationship from section 8:

$$\|\Sigma, \Sigma'\|^{(1,\gamma)}_{(O,O,O)^*} = \inf_{\Sigma^{(1)}+\Sigma^{(2)}+\Sigma^{(3)}=\Sigma} \max \left\{ \gamma^{-1}\|[\Sigma^{(1)}_{(1)}; \Sigma']\|_{op}, \right.$$
$$\left. \|\Sigma^{(2)}_{(2)}\|_{op}, \|\Sigma^{(3)}_{(3)}\|_{op} \right\}.$$

Since we can take any $\Sigma^{(k)}$ to be equal to Σ, the above norm can be upper-bounded as

$$\|\Sigma, \Sigma'\|^{(1,\gamma)}_{(O,O,O)^*} \leq \max \left\{ \gamma^{-1}\|[\Sigma_{(1)}; \Sigma']\|_{op}, \min \left\{ \|\Sigma_{(2)}\|_{op}, \|\Sigma_{(3)}\|_{op} \right\} \right\}.$$

Now, taking the expectation leads to

$$\mathbb{E}\|\Sigma, \Sigma'\|^{(1,\gamma)}_{(O,O,O)^*} \leq \mathbb{E} \max \left\{ \gamma^{-1}\|[\Sigma_{(1)}; \Sigma']\|_{op}, \min \left\{ \|\Sigma_{(2)}\|_{op}, \|\Sigma_{(3)}\|_{op} \right\} \right\}$$
$$\leq \max \left\{ \gamma^{-1}\mathbb{E}\|[\Sigma_{(1)}; \Sigma']\|_{op}, \min \left\{ \mathbb{E}\|\Sigma_{(2)}\|_{op}, \mathbb{E}\|\Sigma_{(3)}\|_{op} \right\} \right\}.$$

From Wimalawarne et al. (2018) we know that $\mathbb{E}\|\Sigma_{(k)}\|_{op} \leq \frac{3C_1}{2}\left(\sqrt{n_k} + \sqrt{\prod_{j\neq k}^3 n_j}\right)$ and $\mathbb{E}\|[\Sigma_{(1)}; \Sigma']\|_{op} \leq \frac{3C_2}{2}\left(\sqrt{n_1} + \sqrt{\prod_{j=2}^3 n_j + n_2'}\right)$ for come constants C_1 and C_2, which give us the final bound as

$$R_{P,Q}(l \circ \mathcal{W}, l \circ \mathcal{V})$$
$$\leq \frac{3\Lambda}{2d} \left[\gamma^{-1}\sqrt{r_{(1)}}(B_{\mathcal{W}} + B_M) + \sum_{k=2}^3 \sqrt{r_k}B_{\mathcal{W}} \right]$$
$$\max \left\{ C_2 \left(\sqrt{n_1} + \sqrt{\prod_{j=2}^3 n_j + n_2'} \right), \min_{k\in 2,3} C_1 \left(\sqrt{n_k} + \sqrt{\prod_{j\neq k}^3 n_j} \right) \right\}.$$
$$\square$$

From the proof of theorem 7, we can see how the parameter γ changes the bounds compared to unscaled coupled norms in Wimalawarne et al. (2018).

Appendix C: Excess Risk Bounds for Coupled Higher-Order Tensor Completion

In this section, we give the proofs for the theorems in section 5.2.

Proof of Theorem 1. Let us denote Σ and Σ' corresponding Rademacher variables corresponding to $\mathcal{W}$ and V in equation 5.5.

We bound $\|\mathcal{W}, V\|^{(1,\gamma)}_{([v],O),(O,O)}$ as follows,

$$\|\mathcal{W}, V\|^{(1,\gamma)}_{([v],O),(O,O)} = \gamma\|[W_{(1)}; V]\|_{\mathrm{tr}} + \|W_{[v]}\|_{\mathrm{tr}} + \|V\|_{\mathrm{tr}}$$

$$\leq \gamma\sqrt{r_{(1)}}(B_{\mathcal{W}} + B_V) + \min\left(\sqrt{\prod_{i=1}^{k} r_i}, \sqrt{\prod_{j=v+1}^{K} r_j}\right) B_{\mathcal{W}}$$

$$+ \sqrt{r'}B_V,$$

where $\|\mathcal{W}\|_{\mathrm{F}} \leq B_{\mathcal{W}}$ and $\|V\|_{\mathrm{F}} \leq B_V$.

Using Latała's theorem (Latała, 2005; Shamir & Shalev-Shwartz, 2014), we can bound $\|\Sigma_{[v]}\|_{\mathrm{op}}$ as

$$\mathbb{E}\|\Sigma_{[v]}\|_{\mathrm{op}} \leq C_2\left(\sqrt{\prod_{i=1}^{k} n_i} + \sqrt{\prod_{j=v+1}^{K} n_j} + \sqrt[4]{|\Sigma_{[v]}|}\right),$$

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

$$\mathbb{E}\|\Sigma_{[v]}\|_{\mathrm{op}} \leq \frac{3C_2}{2}\left(\sqrt{\prod_{i=1}^{v} n_i} + \sqrt{\prod_{j=v+1}^{K} n_j}\right).$$

Similarly, using Latała's theorem, we have the following bound:

$$\mathbb{E}\|[\Sigma_{(1)}; \Sigma'_{(1)}]\|_{\mathrm{op}} \leq \frac{3C_1}{2}\left(\sqrt{n_1} + \sqrt{\prod_{j=2}^{K} n_j + n'_2}\right).$$

From Shamir and Shalev-Shwartz (2014), we know that $\mathbb{E}\|\Sigma'\|_{\mathrm{op}} \leq C_3(\sqrt{n_1} + \sqrt{n'_2})$.

To bound $\mathbb{E}\|\Sigma, \Sigma'\|^{(1,\gamma)}_{([v],O),(O,O)^\star}$, we extend the dual norms for coupled norms from Wimalawarne et al. (2018) as

$$\|\Sigma, \Sigma'\|^{(1,\gamma)}_{([v],O),(O,O)^\star} = \inf_{\Sigma^{(1)}+\Sigma^{(2)}=\Sigma} \inf_{\Sigma'^{(1)}+\Sigma'^{(2)}=\Sigma'} \max\left\{\gamma^{-1}\|[\Sigma^{(1)}_{(1)}; \Sigma'^{(1)}_{(1)}]\|_{\mathrm{op}},\right.$$

$$\left.\|\Sigma^{(2)}_{[v]}\|_{\mathrm{op}}, \|\Sigma'^{(2)}\|_{\mathrm{op}}\right\}.$$

Since we can take any $\Sigma^{(k)}$, $k = 1, 2$ to be equal to Σ and any $\Sigma'^{(l)}$, $l = 1, 2$ to be equal to Σ, the above norm, we have

$$\|\Sigma, \Sigma'\|^{(1,\gamma)}_{([v],O),(O,O)^\star} \le \min\left\{\gamma^{-1}\|[\Sigma_{(1)}; \Sigma']\|_{op}, \max\{\|\Sigma_{[v]}\|_{op}, \|\Sigma'\|_{op}\}\right\},$$

(C.1)

which can be understood as taking $\Sigma^{(1)} = \Sigma$ and $\Sigma^{(1)}_2 = \Sigma$, which lead to $\|\Sigma, \Sigma'\|^{(1,\gamma)}_{([v],O),(O,O)^\star} \le \max\{\|\Sigma_{[v]}\|_{op}, \|\Sigma'\|_{op}\}$, and taking $\Sigma^{(2)} = \Sigma$ and $\Sigma'^{(2)} = \Sigma'$, which lead to $\|\Sigma, \Sigma'\|^{(1,\gamma)}_{([v],O),(O,O)^\star} \le \gamma^{-1}\|[\Sigma_{(1)}; \Sigma']\|_{op}$, while the results by other combinations of $\Sigma^{(1)} = \Sigma$ and $\Sigma'^{(2)} = \Sigma'$ or $\Sigma^{(2)} = \Sigma$ and $\Sigma^{(1)}_2 = \Sigma'$ are also upper-bounded by equation C.1.

Furthermore, taking the expectation of equation C.1 leads to

$$\mathbb{E}\|\Sigma, \Sigma'\|^{(1,\gamma)}_{([v],O),(O,O)^\star}$$

$$\le \mathbb{E}\min\left\{\gamma^{-1}\|[\Sigma_{(1)}; \Sigma']\|_{op}, \max\{\|\Sigma_{[v]}\|_{op}, \|\Sigma'\|_{op}\}\right\}$$

$$\le \min\left\{\gamma^{-1}\mathbb{E}\|[\Sigma_{(1)}; \Sigma']\|_{op}, \max\{\mathbb{E}\|\Sigma_{[v]}\|_{op}, \mathbb{E}\|\Sigma'\|_{op}\}\right\}.$$

Finally, by using equation 5.5, we obtain

$$R_{P,Q}(l \circ \mathcal{W}, l \circ \mathcal{V})$$

$$\le \frac{3\Lambda}{2d}\left[\gamma\sqrt{r_{(1)}}(B_\mathcal{W} + B_V) + \min\left(\sqrt{\prod_{i=1}^{v}r_i}, \sqrt{\prod_{j=v+1}^{K}r_j}\right)B_\mathcal{W} + \sqrt{r'}B_V\right]$$

$$\min\left\{\gamma^{-1}C_1\left(\sqrt{n_1} + \sqrt{\prod_{j=2}^{K}n_j + n'_2}\right), \max\left\{C_2\left(\sqrt{\prod_{i=1}^{v}n_i} + \sqrt{\prod_{j=v+1}^{K}n_j}\right),\right.\right.$$

$$\left.\left. C_3\left(\sqrt{n_1} + \sqrt{n'_2}\right)\right\}\right\}.$$

$\square$

Proof of Theorem 2. Similar to theorem 1 due to all the norms regularized by overlapped norms, we have the following bound for $\|\mathcal{W}, \mathcal{V}\|^{1}_{([v],O),(O,O,O)}$ given that $\mathcal{W}$ and $\mathcal{V}$ are the corresponding learned elements with of $\mathcal{X}$ and $\mathcal{Y}$, respectively, as

$$\|\mathcal{W}, \mathcal{V}\|^{(1,\gamma)}_{([v],O),(O,O,O)} \le \left[\gamma\sqrt{r_{(1)}}(B_\mathcal{W} + B_\mathcal{V}) + \sum_{i=2}^{3}\sqrt{r'_i}B_\mathcal{W}\right.$$

$$+ \min \left(\sqrt{\prod_{i=1}^{v} r_i}, \sqrt{\prod_{j=v+1}^{K} r_j} \right) B_{\mathcal{V}} \right], \tag{C.2}$$

where $\|\mathcal{W}\|_{\mathrm{F}} \leq B_{\mathcal{W}}$ and $\|\mathcal{V}\|_{\mathrm{F}} \leq B_{\mathcal{V}}$.

Considering the dual norm, $\mathbb{E}\|\Sigma, \Sigma'\|_{([v]),(O,O,O)^*}^{(1,\gamma)}$ is given as

$$\|\Sigma, \Sigma'\|_{([v],O),(O,O,O)^*}^{(1,\gamma)}$$

$$= \inf_{\Sigma^{(1)}+\Sigma^{(2)}+\Sigma^{(3)}=\Sigma} \inf_{\Sigma'^{(1)}+\Sigma'^{(2)}=\Sigma'} \max \left\{ \gamma^{-1}\|[\Sigma_{(1)}^{(1)}; \Sigma'_{(1)}^{(1)}]\|_{\mathrm{op}}, \|\Sigma_{[v]}^{(2)}\|_{\mathrm{op}}, \right.$$

$$\left. \|\Sigma'_{(2)}^{(2)}\|_{\mathrm{op}}, \|\Sigma'_{(3)}^{(3)}\|_{\mathrm{op}} \right\},$$

which can be expressed by taking a similar argument as in theorem 1 as

$$\|\Sigma, \Sigma'\|_{([v],O),(O,O,O)^*}^{(1,\gamma)} \leq \min \left\{ \gamma^{-1}\|[\Sigma_{(1)}; \Sigma'_{(1)}]\|_{\mathrm{op}}, \right.$$

$$\left. \max \left\{ \|\Sigma_{[v]}\|_{\mathrm{op}}, \min \left\{ \|\Sigma'_{(2)}\|_{\mathrm{op}}, \|\Sigma'_{(3)}\|_{\mathrm{op}} \right\} \right\} \right\}.$$

The expectation of the above dual norm is bound as

$$\mathbb{E}\|\Sigma, \Sigma'\|_{([v],O),(O,O,O)^*}^{(1,\gamma)} \leq \min \left\{ \gamma^{-1}\mathbb{E}\|[\Sigma_{(1)}; \Sigma'_{(1)}]\|_{\mathrm{op}}, \max \left\{ \mathbb{E}\|\Sigma_{[v]}\|_{\mathrm{op}}, \right. \right.$$

$$\left. \left. \min \left\{ \mathbb{E}\|\Sigma'_{(2)}\|_{\mathrm{op}}, \mathbb{E}\|\Sigma'_{(3)}\|_{\mathrm{op}} \right\} \right\} \right\},$$

which can be bounded as

$$\mathbb{E}\|\Sigma, \Sigma'\|_{([v],O),(O,O,O)^*}^{(1,\gamma)} \leq \frac{3}{2} \min \left\{ C_1 \left(\sqrt{n_1} + \sqrt{\prod_{a=2}^{K} n_a + n_2' n_3'} \right), \right.$$

$$\max \left\{ C_2 \left(\sqrt{\prod_{i=1}^{v} n_i} + \sqrt{\prod_{j=v+1}^{K} n_j} \right), \min \left\{ C_4 \gamma^{-1} \left(\sqrt{n_2'} + \sqrt{n_1 n_3'} \right), \right. \right.$$

$$\left. \left. \left. C_5 \left(\sqrt{n_3'} + \sqrt{n_1 n_2'} \right) \right\} \right\} \right\}. \tag{C.3}$$

By combining equations C.2 and C.3 with 5.5, we arrive at the final bound. $\qquad \square$

Next, we look at the excess risk for the coupled norm $\|\mathcal{X}, \mathcal{Y}\|_{([v],O),(L,L,L)}^{(1,\gamma)}$.

Theorem 8. *Let $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ and $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2' \times n_3'}$ be coupled on their first modes with sets of observed elements* P *and* Q, *respectively, with* $|\mathrm{P}|/2 = |\mathrm{Q}|/2 = d$. *Given a hypothesis class* $\mathbf{W} = \{\mathcal{W}, \mathcal{V} : \|\mathcal{W}, \mathcal{V}\|_{([v],\mathrm{O}),(\mathrm{L},\mathrm{L},\mathrm{L})}^{(1,\gamma)} \leq B\}$, *the coupled completion using* $\|\mathcal{W}, \mathcal{V}\|_{([v],\mathrm{O}),(\mathrm{L},\mathrm{L},\mathrm{L})}^{(1,\gamma)}$ *leads to the following Rademacher complexity of equation 5.5 with probability* $1 - \delta$:

$$
R_{\mathrm{P},\mathrm{Q}}(l \circ \mathcal{W}, l \circ \mathcal{V})
$$

$$
\leq \frac{3\Lambda}{2d} \left[\gamma \sqrt{r_{(1)}}(B_{\mathcal{W}} + B_{\mathcal{V}}) + min \left(\sqrt{\prod_{i=1}^{v} r_i}, \sqrt{\prod_{j=v+1}^{K} r_j} \right) B_{\mathcal{W}} \right.
$$

$$
\left. + min\left(\sqrt{r_2'}, \sqrt{r_3'}\right)B_{\mathcal{V}} \right] max \left\{ \gamma^{-1}C_1 \left(\sqrt{n_1} + \sqrt{\prod_{a=2}^{K} n_a + n_2' n_3'} \right), \right.
$$

$$
\left. C_2 \left(\sqrt{\prod_{i=1}^{v} n_i} + \sqrt{\prod_{j=v+1}^{K} n_j} \right), C_4\left(\sqrt{n_2'} + \sqrt{n_1 n_3'}\right), C_5\left(\sqrt{n_3'} + \sqrt{n_1 n_2'}\right) \right\},
$$

where $(r_1, \ldots, r_K)$ *is the multilinear rank of* $\mathcal{W}$; (r_1', r_2', r_3') *is the multilinear rank of* $\mathcal{V}$; $r_{(1)}$ *is the rank of the coupled unfolding on the first mode; and* $B_{\mathcal{W}}$, $B_{\mathcal{V}}$, C_1, C_2, C_4, *and* C_5 *are constants.*

Proof. Let $\mathcal{W}$ and $\mathcal{V}$ be the completed tensors for $\mathcal{X}$ and $\mathcal{Y}$. Also let us denote Σ and Σ' consisting of corresponding Rademacher variables of $\mathcal{X}$ and $\mathcal{Y}$.

Here, we have $\|\mathcal{W}, \mathcal{V}\|_{([v],\mathrm{O}),(\mathrm{L},\mathrm{L},\mathrm{L})}^{(1,\gamma)}$, which can be explicitly written as

$$
\|\mathcal{W}, \mathcal{V}\|_{([v],\mathrm{O}),(\mathrm{L},\mathrm{L},\mathrm{L})}^{(1,\gamma)}
$$

$$
= \inf_{\mathcal{V}^{(1)}+\mathcal{V}^{(2)}+\mathcal{V}^{(3)}=\mathcal{V}} \left(\|[W_{(1)}^{(1)}, V_{(1)}]\|_{\mathrm{tr}} + \|W_{[v]}\|_{\mathrm{tr}} + \sum_{a=2}^{3} \|V_{(a)}^{(a)}\|_{\mathrm{tr}} \right), \quad \text{(C.4)}
$$

which we can bound as

$$
\|\mathcal{W}, \mathcal{V}\|_{([v],\mathrm{O}),(\mathrm{L},\mathrm{L},\mathrm{L})}^{(1,\gamma)} \leq \gamma \sqrt{r_{(1)}}(B_{\mathcal{W}} + B_{\mathcal{V}}) + min \left(\sqrt{\prod_{i=1}^{v} r_i}, \sqrt{\prod_{j=v+1}^{K} r_j} \right) B_{\mathcal{W}}
$$

$$
+ min\left(\sqrt{r_2'}, \sqrt{r_3'}\right)B_{\mathcal{V}}, \quad \text{(C.5)}
$$

where we have used the fact that we can take $\mathcal{W}^{(2)}$ or $\mathcal{W}^{(3)}$ to be equal to $\mathcal{W}$ in the last inequality and take $\|\mathcal{W}\|_{\mathrm{F}} \leq B_{\mathcal{W}}$ and $\|\mathcal{V}\|_{\mathrm{F}} \leq B_{\mathcal{V}}$.

To complete the proof, we need the dual norm, $\|\Sigma, \Sigma'\|^{(1,\gamma)}_{([v],O),(L,L,L)^\star}$, which can be written as follows,

$$\|\Sigma, \Sigma'\|^{(1,\gamma)}_{([v],O),(L,L,L)^\star} = \inf_{\Sigma^{(1)}+\Sigma^{(2)}=\Sigma} \max \left\{ \gamma^{-1}\|[\Sigma'_{(1)}; \Sigma^{(1)}_{(1)}]\|_{\mathrm{op}}, \|\Sigma^{(2)}_{[v]}\|_{\mathrm{op}}, \right.$$
$$\left. \|\Sigma'_{(2)}\|_{\mathrm{op}}, \|\Sigma'_{(3)}\|_{\mathrm{op}} \right\},$$

which can be bounded as follows,

$$\|\Sigma, \Sigma'\|^{(1,\gamma)}_{([v],O),(L,L,L)^\star} \leq \max \left\{ \gamma^{-1}\|[\Sigma'_{(1)}; \Sigma_{(1)}]\|_{\mathrm{op}}, \|\Sigma_{[v]}\|_{\mathrm{op}}, \right.$$
$$\left. \|\Sigma'_{(2)}\|_{\mathrm{op}}, \|\Sigma'_{(3)}\|_{\mathrm{op}} \right\},$$

where the infimum with respect to Σ does not have an effect.

The above dual norm can be bounded with its expectation as

$$\mathbb{E}\|\Sigma, \Sigma'\|^{(1,\gamma)1}_{([v],O),(L,L,L)^\star}$$
$$\leq \max \left\{ \gamma^{-1}\|[\Sigma'_{(1)}; \Sigma^{(1)}_{(1)}]\|_{\mathrm{op}}, \|\Sigma^{(2)}_{[v]}\|_{\mathrm{op}}, \|\Sigma'_{(2)}\|_{\mathrm{op}}, \|\Sigma'_{(3)}\|_{\mathrm{op}} \right\}$$
$$\leq \max \left\{ \gamma^{-1}\mathbb{E}\|[\Sigma'_{(1)}; \Sigma^{(1)}_{(1)}]\|_{\mathrm{op}}, \mathbb{E}\|\Sigma^{(2)}_{[v]}\|_{\mathrm{op}}, \mathbb{E}\|\Sigma'_{(2)}\|_{\mathrm{op}}, \mathbb{E}\|\Sigma'_{(3)}\|_{\mathrm{op}} \right\},$$

which leads to

$$\mathbb{E}\|\Sigma, \Sigma'\|^{(1,\gamma)}_{([v],O),(L,L,L)^\star} \leq \max \left\{ \gamma^{-1}C_1 \left(\sqrt{n_1} + \sqrt{\prod_{a=2}^{K} n_a + n'_2 n'_3} \right), \right.$$
$$C_2 \left(\sqrt{\prod_{i=1}^{v} n_i} + \sqrt{\prod_{j=v+1}^{K} n_j} \right), C_4\left(\sqrt{n'_2} + \sqrt{n_1 n'_3}\right),$$
$$\left. C_5\left(\sqrt{n'_3} + \sqrt{n_1 n'_2}\right) \right\}. \tag{C.6}$$

Combining equations C.5 and C.6 with 5.5 completes the proof.   □

Proof of Theorem 3. To derive the bounds for $\|\mathcal{W}, \mathcal{V}\|^{(1,\gamma)}_{([v],O),(S,S,S)}$ we use a similar approach as for theorem 4.

We have the following bound for $\|\mathcal{W}, \mathcal{V}\|^{(1,\gamma)}_{([v],O),(S,S,S)}$ as

$$\|\mathcal{W}, \mathcal{V}\|^{(1,\gamma)}_{([v],O),(S,S,S)} \leq \gamma \sqrt{\frac{r_{(1)}}{n_1}}(B_\mathcal{W} + B_\mathcal{V}) + \min \left(\sqrt{\prod_{i=1}^{v} r_i}, \sqrt{\prod_{j=v+1}^{K} r_j} \right) B_\mathcal{W}$$

$$+ \min \left(\sqrt{\frac{r'_2}{n_2}}, \sqrt{\frac{r'_3}{n_3}} \right) B_{\mathcal{V}}, \tag{C.7}$$

where $\|\mathcal{W}\|_{\mathrm{F}} \leq B_{\mathcal{W}}$ and $\|\mathcal{V}\|_{\mathrm{F}} \leq B_{\mathcal{V}}$.

The dual norm, $\|\Sigma, \Sigma'\|_{([v],O),(S,S,S)^*}^{(1,\gamma)}$, can be written as

$$\|\Sigma, \Sigma'\|_{([v],O),(S,S,S)^*}^{(1,\gamma)} = \inf_{\Sigma^{(1)}+\Sigma^{(2)}=\Sigma} \max \left\{ \gamma^{-1}\sqrt{n_1}\|[\Sigma'_{(1)}; \Sigma^{(1)}_{(1)}]\|_{\mathrm{op}}, \|\Sigma^{(2)}_{[v]}\|_{\mathrm{op}}, \right.$$
$$\left. \sqrt{n_2}\|\Sigma'_{(2)}\|_{\mathrm{op}}, \sqrt{n_3}\|\Sigma'_{(3)}\|_{\mathrm{op}} \right\}.$$

Using similar arguments as in theorem 5, we derive the bound for $\mathbb{E}\|\Sigma, \Sigma'\|_{([v],O),(S,S,S)^*}^{(1,\gamma)}$ as follows:

$$\mathbb{E}\|\Sigma, \Sigma'\|_{([v],O),(S,S,S)^*}^{(1,\gamma)} \leq \max \left\{ \gamma^{-1}C_1 \left(n_1 + \sqrt{\prod_{a=1}^{K} n_a + n_1 n'_2 n'_3} \right), \right.$$
$$C_2 \left(\sqrt{\prod_{i=1}^{v} n_i} + \sqrt{\prod_{j=v+1}^{K} n_j} \right), C_4 \left(n'_2 + \sqrt{n_1 n'_2 n'_3} \right),$$
$$\left. C_5 \left(n'_3 + \sqrt{n_1 n'_2 n'_3} \right) \right\}. \tag{C.8}$$

Combining equations C.7 and C.8 with 5.5 completes the proof.   □

Next, we give proofs for the mixed norms among higher-order tensors and three-mode tensors. First, we give the proof of theorem 4 for $\|\mathcal{X}, \mathcal{Y}\|_{([v],O),(S,O,O)}^{(1,\gamma)}$.

Proof of Theorem 4. We first write $\|\mathcal{W}, \mathcal{V}\|_{([v],O),(S,O,O)}^{(1,\gamma)}$ explicitly as

$$\|\mathcal{W}, \mathcal{V}\|_{([v],O),(S,O,O)}^{(1,\gamma)}$$
$$= \inf_{\mathcal{V}^{(1)}+\mathcal{V}^{(2)}=\mathcal{V}} \left(\frac{\gamma}{\sqrt{n_1}} \|[W_{(1)}; V^{(1)}_{(1)}]\|_{\mathrm{tr}} + \sum_{a=2}^{3} \|V^{(2)}_{(a)}\|_{\mathrm{tr}} + \|W_{[v]}\|_{\mathrm{tr}} \right),$$

and it can be bounded as

$$\|\mathcal{W}, \mathcal{V}\|_{([v],O),(S,O,O)}^{(1,\gamma)} \leq \gamma \sqrt{\frac{r_{(1)}}{n_1}}(B_{\mathcal{W}} + B_{\mathcal{V}}) + \sum_{a=2}^{3} \sqrt{r'_a} B_{\mathcal{V}}$$

$$+ \min \left(\sqrt{\prod_{i=1}^{v} r_i}, \sqrt{\prod_{j=v+1}^{K} r_j} \right) B_{\mathcal{W}}, \tag{C.9}$$

where $\|\mathcal{W}\|_F \leq B_{\mathcal{W}}$ and $\|\mathcal{V}\|_F \leq B_{\mathcal{V}}$.

The dual norm, $\|\Sigma, \Sigma'\|_{([v],O),(S,O,O)^\star}^{(1,\gamma)}$, can be written as

$$\|\Sigma, \Sigma'\|_{([v],O),(S,O,O)^\star}^{(1,\gamma)} = \inf_{\Sigma^{(1)}+\Sigma^{(2)}=\Sigma} \max \left\{ \gamma^{-1}\sqrt{n_1} \| [\Sigma_{1(1)}; \Sigma_{(1)}^{'(1)}] \|_{op}, \right.$$

$$\left. \| \Sigma_{1[v]} \|_{op}, \| \Sigma_{(2)}^{'(2)} \|_{op}, \| \Sigma_{2(3)}^{(2)} \|_{op} \right\},$$

which can be bounded as

$$\|\Sigma, \Sigma'\|_{([v],O),(S,O,O)^\star}^{(1,\gamma)} \leq \max \left\{ \gamma^{-1}\sqrt{n_1} \| [\Sigma_{1(1)}; \Sigma_{2(1)}] \|_{op}, \| \Sigma_{1[v]} \|_{op}, \right.$$

$$\left. \min\{ \| \Sigma_{2(2)} \|_{op}, \| \Sigma_{2(3)} \|_{op} \} \right\},$$

which leads to

$$\mathbb{E} \|\Sigma, \Sigma'\|_{([v]),(P,O,O)^\star}^{(1,\gamma)} \leq \max \left\{ \gamma^{-1} C_1 \left(n_1 + \sqrt{\prod_{a=1}^{K} n_a + n_1 n_2' n_3'} \right), \right.$$

$$C_2 \left(\sqrt{\prod_{i=1}^{v} n_i} + \sqrt{\prod_{j=v+1}^{K} n_j} \right) \min \left(C_4 \left(\sqrt{n_2'} + \sqrt{n_1 n_3'} \right), \right.$$

$$\left. \left. C_5 \left(\sqrt{n_3'} + \sqrt{n_1 n_2'} \right) \right) \right\}. \tag{C.10}$$

Combining equations C.9 and C.10 with 5.5 completes the proof. □

Following theorem 6, we can derive the bounds for other mixed norms as well. Next, we give the bounds for $\|\mathcal{X}, \mathcal{Y}\|_{([v],O),(O,S,O)}^{(1,\gamma)}$ and $\|\mathcal{X}, \mathcal{Y}\|_{([v],O),(O,O,S)}^{(1,\gamma)}$ without proofs.

Theorem 9. *Let $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ and $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2' \times n_3'}$ be coupled on their first modes with the set of observed elements* P *and* Q, *respectively, with $|P|/2 = |Q|/2 = d$. Given a hypothesis class $W = \{\mathcal{W}, \mathcal{V} : \|\mathcal{W}, \mathcal{V}\|_{([v],O),(O,S,O)}^{(1,\gamma)} \leq B\}$, the coupled completion using $\|\mathcal{W}, \mathcal{V}\|_{([v],O),(O,S,O)}^{(1,\gamma)}$ leads to the following Rademacher complexity of equation 5.5 with probability $1 - \delta$,*

$$R_{S,P}(l \circ \mathcal{W}, l \circ \mathcal{V})$$

$$\leq \frac{3\Lambda}{2d} \left[\gamma \sqrt{r_{(1)}}(B_{\mathcal{W}} + B_{\mathcal{V}}) + \min \left(\sqrt{\prod_{i=1}^{k} r_i}, \sqrt{\prod_{j=v+1}^{K} r_j} \right) B_{\mathcal{W}} + \sqrt{r_3'} B_{\mathcal{V}} \right.$$

$$\left. + \sqrt{\frac{r_2'}{n_2'}} B_{\mathcal{V}} \right] \max \left\{ \min \left(C_4 \left(n_2' + \sqrt{n_1 n_2' n_3'} \right), C_5 \left(\sqrt{n_3'} + \sqrt{n_1 n_2'} \right) \right), \right.$$

$$\left. C_1 \gamma^{-1} \left(\sqrt{n_1} + \sqrt{\prod_{a=2}^{K} n_a + n_2' n_3'} \right), C_2 \left(\sqrt{\prod_{i=1}^{v} n_i} + \sqrt{\prod_{j=v+1}^{K} n_j} \right) \right\},$$

where $(r_1, \ldots, r_K)$ is the multilinear rank of $\mathcal{W}$; (r_1', r_2', r_3') is the multilinear rank of $\mathcal{V}$; $r_{(1)}$ is the rank of the coupled unfolding on the first mode; $B_{\mathcal{W}}$, and $B_{\mathcal{V}}$, C_1, C_2, C_3, and C_4 are constants.

Theorem 10. *Let $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_K}$ and $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2' \times n_3'}$ be coupled on their first modes with the set of observed elements P and Q, respectively, with $|P|/2 = |Q|/2 = d$. Given a hypothesis class $W = \{\mathcal{W}, \mathcal{V} : \|\mathcal{W}, \mathcal{V}\|_{([v],O),(O,O,S)}^{(1,\gamma)} \leq B\}$, the coupled completion using $\|\mathcal{W}, \mathcal{V}\|_{([v],O),(O,O,S)}^{(1,\gamma)}$ leads to the following Rademacher complexity of equation 5.5 with probability $1 - \delta$:*

$$R_{P,Q}(l \circ \mathcal{W}, l \circ \mathcal{V})$$

$$\leq \frac{3\Lambda}{2d} \left[\gamma \sqrt{r_{(1)}}(B_{\mathcal{W}} + B_{\mathcal{V}}) + \min \left(\sqrt{\prod_{i=1}^{v} r_i}, \sqrt{\prod_{j=v+1}^{K} r_j} \right) B_{\mathcal{W}} + \sqrt{r_2'} B_{\mathcal{V}} \right.$$

$$\left. + \sqrt{\frac{r_3'}{n_3'}} B_{\mathcal{V}} \right] \max \left\{ \min \left(C_4 \left(\sqrt{n_2'} + \sqrt{n_1 n_3'} \right), C_5 \left(n_3' + \sqrt{n_1 n_2' n_3'} \right) \right), \right.$$

$$\left. C_1 \gamma^{-1} \left(\sqrt{n_1} + \sqrt{\prod_{a=2}^{K} n_a + n_2' n_3'} \right), C_2 \left(\sqrt{\prod_{i=1}^{v} n_i} + \sqrt{\prod_{j=v+1}^{K} n_j} \right) \right\},$$

where $(r_1, \ldots, r_K)$ is the multilinear rank of $\mathcal{W}$; (r_1', r_2', r_3') is the multilinear rank of $\mathcal{V}$; $r_{(1)}$ is the rank of the coupled unfolding on the first mode; and $B_{\mathcal{W}}$, $B_{\mathcal{V}}$, C_1, C_2, C_3, and C_4 are constants.

Next we give the bound for the coupled norm $\|\mathcal{W}, \mathcal{V}\|_{([v],O),([v'],O)}^{(1,\gamma)}$.

Proof of Theorem 5. Similar to theorem 1 (all norms are in the form of overlapping norms), we have the following for $\|\mathcal{W}, \mathcal{V}\|_{([v],O),([v'],O)}^{(1,\gamma)}$ given that $\mathcal{W}$ and $\mathcal{V}$ are the corresponding learned elements with respect to $\mathcal{X}$ and $\mathcal{Y}$,

receptively,

$$\|\mathcal{W}, \mathcal{V}\|_{([v],\mathrm{O}),([v'],\mathrm{O})}^{(1,\gamma)} \leq \gamma\sqrt{r_{(1)}}(B_{\mathcal{W}} + B_{\mathcal{V}}) + \min\left(\sqrt{\prod_{i1=1}^{k} r_{i1}}, \sqrt{\prod_{j1=v+1}^{K} r_{j1}}\right) B_{\mathcal{W}}$$

$$+ \min\left(\sqrt{\prod_{i2=1}^{v'} r'_{i2}}, \sqrt{\prod_{j2=k'+1}^{K'} r'_{j2}}\right) B_{\mathcal{V}}, \qquad (\mathrm{C}.11)$$

where $\|\mathcal{W}\|_{\mathrm{F}} \leq B_{\mathcal{W}}$ and $\|\mathcal{V}\|_{\mathrm{F}} \leq B_{\mathcal{V}}$.

It is easy to derive that

$$\|\Sigma, \Sigma'\|_{([v],\mathrm{O}),([v'],\mathrm{O})^*}^{(1,\gamma)} = \inf_{\Sigma^{(1)}+\Sigma^{(2)}=\Sigma} \inf_{\Sigma'^{(1)}+\Sigma'^{(2)}=\Sigma'} \max\left\{\gamma^{-1}\|[\Sigma_{(1)}^{(1)}, \Sigma_{(1)}'^{(1)}]\|_{\mathrm{op}}, \right.$$

$$\left. \|\Sigma_{[v]}^{(2)}\|_{\mathrm{op}}, \|\Sigma_{[v']}'^{(2)}\|_{\mathrm{op}}\right\},$$

which can be simplified using a similar argument in theorem 1 as follows:

$$\|\Sigma, \Sigma'\|_{([v],\mathrm{O}),([v'],\mathrm{O})^*}^{(1,\gamma)}$$

$$\leq \min\left\{\gamma^{-1}\|[\Sigma_{(1)}, \Sigma_{(1)}']\|_{\mathrm{op}}, \max\{\|\Sigma_{[v]}\|_{\mathrm{op}}, \|\Sigma_{[v']}'\|_{\mathrm{op}}\}\right\},$$

Its expectation leads to

$$\mathbb{E}\|\Sigma, \Sigma'\|_{([v],\mathrm{O}),([v'],\mathrm{O})^*}^{(1,\gamma)} \leq \frac{3}{2}\min\left\{\gamma^{-1}C_6\left(\sqrt{n_1} + \sqrt{\prod_{a=2}^{K} n_a + \prod_{b=2}^{K} n'_b}\right), \right.$$

$$\max\left\{C_2\left(\sqrt{\prod_{i=1}^{v} n_i} + \sqrt{\prod_{j=v+1}^{K} n_j}\right), \right.$$

$$\left.\left. C_7\left(\sqrt{\prod_{i'=1}^{v'} n'_{i'}} + \sqrt{\prod_{j'=v'+1}^{K'} n'_{j'}}\right)\right\}\right\}. \qquad (\mathrm{C}.12)$$

Combining equations C.11 and C.12 with 5.5 completes the proof. $\square$

Given higher-order tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$, the excess risk bound with the square reshaping norm.

Theorem 11. *Using the square reshaping norm regularization given by* $\|T_{[v]}\|_{\mathrm{tr}}$, *we have*
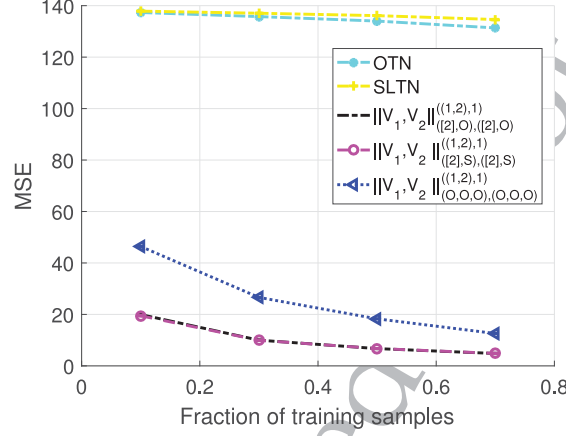
Figure 6: Further experiments for multiview video completion.

$$R_{\mathrm{P}}(l \circ \mathcal{W}) \leq \frac{3c_4 \Lambda B_{\mathcal{W}}}{2d} \, min \left[\sqrt{\prod_{i=1}^{v} r_i}, \sqrt{\prod_{j=v+1}^{K} r_j} \right] \left(\sqrt{\prod_{i=1}^{v} n_i} + \sqrt{\prod_{j=v+1}^{K} n_j} \right).$$

for some constant c_4 and $(r_1, \ldots, r_K)$ is the multilinear rank of $\mathcal{W}$.

Proof. The proof is direct and can be derived similarly to theorem 3 without considering the matrix coupling.   □

Appendix D: Further Experiments for Multiview Video Completion

We give further results for baseline methods for the multiview video completion experiment in section 6.4. We performed individual tensor completion of corrupted video data $\mathcal{V}_1$ using the overlapped trace norm (OTN) and scaled latent trace norm (SLTN). We performed coupled completion using the coupled norm $\|\mathcal{V}_1, \mathcal{V}_2\|_{(O,O,O),(O,O,O)}^{((1,2),1)}$ by extending the coupled norms in Wimalawarne et al. (2018). In Figure 6, we compare these baseline methods with the proposed norms $\|\mathcal{V}_1, \mathcal{V}_2\|_{([2],O),([2],O)}^{((1,2),1)}$ and $\|\mathcal{V}_1, \mathcal{V}_2\|_{([2],S),([2],S)}^{((1,2),1)}$. We observed that the baseline methods performed poorly compared to proposed methods.

Acknowledgments

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 European 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.

Berclaz, J., Fleuret, F., Turetken, E., & Fua, P. (2011). Multiple object tracking using K-shortest paths optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*, 1806–1819.

Bouchard, G., Yin, D., & Guo, S. (2013). Convex collective matrix factorization. In *Proceedings of the International Conference on Artificial Intelligence and Statistics* (pp. 144–152). JMLR.org.

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

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. New York: Cambridge University Press.

Candès, E. J., & Recht, B. (2009). Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, *9*(6), 717.

El-Yaniv, R., & Pechyony, D. (2007). Transductive Rademacher complexity and its applications. In N. H. Bghouty & C. Gentile (Eds.), *Learning theory*, Vol. 4539, 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.

Fazel, M. (2002). *Matrix rank minimization with applications*. PhD diss., Stanford University.

Jeon, B., Jeon, I., Sael, L., & Kang, U. (2016). Scout: Scalable coupled matrix-tensor factorization—algorithm and discoveries. In *Proceedings of the 32nd IEEE International Conference on Data Engineering* (pp. 811–822). Piscataway, NJ: IEEE.

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 29th AAAI Conference on Artificial Intelligence*. Palo Alto, CA: AAAI.

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. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*, 2114–2121.

Meir, R., & Zhang, T. (2003). Generalization error bounds for Bayesian mixture algorithms. *Journal of Machine Learning Research*, *4*, 839–860.

Mu, 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*. JMLR.org.

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

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

Tomioka, R., & Suzuki, T. (2013). Convex tensor decomposition via structured Schatten norm regularization. In C. J. C. Burgess, L. Bottou, M. Welling, Z. S. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems, 26* (pp. 1331–1339). Red Hook, NY: Curran.

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., Yamada, M., & Mamitsuka, H. (2018). Convex coupled matrix and tensor completion. *Neural Computation*, *30*(11), 3095–3127.

Yuan, M., & Zhang, C.-H. (2016). On tensor completion via nuclear norm minimization. *Foundations of Computational Mathematics*, *16*(4), 1031–1068.

Zhou, T., Qian, H., Shen, Z., Zhang, C., & Xu, C. (2017). Tensor completion with side information: A riemannian manifold approach. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence* (pp. 3539–3545). Palo Alto, CA: AAAI.