Tensor denoising and completion based on ordinal observations

Anonymous Authors[1]

Abstract

1. Introduction

Multidimensional array datasets, a.k.a. a tensor, appear in a huge variety of applications including recommendation systems (Kutty et al., 2012; Adomavicius & Tuzhilin, 2011; Sun et al., 2017), social networks (Sun et al., 2009; Nickel et al., 2011), genomics (Wang & Li, 2018), neuroimaging (EEG, fMRI) (Miwakeichi et al., 2004) and signalprocessing (Sidiropoulos et al., 2000; Cichocki et al., 2015). Instead of unfolding those data tensors into matrices where many analysis methods have been proposed, we suggest preserving multi-modal tensor structure in this paper. Studying tensor data while respecting the structure allows us to examine complex interactions among tensor entries. Thereby we can expect to provide extra more interpretation that cannot be addressed by traditional matrix analysis. Furthermore, It has been shown that the tensor preserving analysis improves performance (Zare et al., 2018; Wang & Li, 2018). With those reasons, there is a growing need to develop dimension reduction method without losing tensor structure. In the line of the attempts, a number of tensor decomposition methods have been proposed in many applications. CANDECOMP/PARAFAC (CP) decomposition was first introduced (Hitchcock, 1928) and made use of in psychometrics (Harshman et al., 1970) and in linguistics (Smilde et al., 2005). The Tucker decomposition was proposed in psychometrics (Tucker & Tucker, 1964; Tucker, 1966).

Classical tensor completion with those decompositions has treated the entries of data as real-valued. In many cases, However, we encounter data sets of which the entries are not real-valued but discrete or quantized i.e. binary-valued or ordinal-valued. For example, many survey data sets take the integer values. To be specific, the data set in the Netflix problem has the three modes of 'user','movie' and 'date of grade'. The entries of the data are the ratings from the users which take integer value from 1 to 5. Another example is the case where the data sets are quantized in real application. In signal processing, the data sets are frequently rounded or truncated so that only integer values are available. In network analysis, entries of adjacency matrix can be labeled as from 1 to 3 taking 3 when pairs of vertexes have strong connection and giving 1 when two vertexes have the weak connection according to a given threshold. If we add one more mode on adjacency matrix such as 'context' or 'individual', the data turns into tensor data with 3 integer values.

We expect that performance improvement can be achieved when the observations are treated as discrete ordinal value not a continuous value. In matrix case, there has been many achievements to complete matrix for discrete cases: Models for the case of binary or 1-bit were introduced and studied (Davenport et al., 2014; Bhaskar & Javanmard, 2015). Furthermore, Bhaskar (2016) suggested matrix completion method for general ordinal observations. In tensor case, however, only binary tensor has gained an attention and achieved performance improvement using binary tensor decomposition methods (Hore et al., 2016; Wang & Li, 2018; Hong et al., 2018; Hu et al., 2018). Accordingly, a general method for the data which has more than 2 ordered label is needed.

We organized this paper as follows. We specify our notations that we will use over this paper in the next section. In section 3, we discuss detailed assumptions and descriptions about our probabilistic model. Also, we suggest the estimation method for the latent parameters and related algorithm. In section 4, we provide the statistical properties of the upper, lower bounds and the phase-transition. We then provide the numerical experiments. Our model is applied to real-world data to check validity and performance in section 6. Finally, we wrap up the paper with a discussion.

2. Preliminaries

Let $\mathcal{Y} \in \mathbb{R}^{d_1\times\cdots\times d_K}$ denote an order-K $(d_1,\ldots,d_K)$-dimensional tensor. We use y_ω to denote the tensor entry indexed by ω, where $\omega \in [d_1]\times\cdots\times[d_K]$. The Frobenius norm of $\mathcal{Y}$ is defined as $\|\mathcal{Y}\|_F = \sum_\omega y_\omega^2$ and the infinity norm of $\mathcal{Y}$ is defined as $\|\mathcal{Y}\|_\infty = \max_\omega |y_\omega|$. We use $\mathcal{Y}_{(k)}$ to denote the unfolded matrix of size d_k-by-

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

$\prod_{i \neq k} d_k$, obtained by reshaping the tensor along the mode-k. The Tucker rank of $\mathcal{Y}$ is defined as a length-K vector $\boldsymbol{r} = (r_1, \ldots, r_K)$, where r_k is the rank of matrix $\mathcal{Y}_{(k)}$ for all $k \in [L]$. We say that an event A occurs "with very high probability" if $\mathbb{P}(A)$ tends to 1 faster than any polynomial of tensor dimension $d_{\min} = \min\{d_1, \ldots, d_K\}$ as $d_{\min} \to \infty$.

We use lower-case letters (e.g., a, b, c) for scalars/vectors, upper-case boldface letters (e.g., $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$) for matrices, and calligraphy letters (e.g., $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$) for tensors of order three or greater. For ease of notation, we allow the basic arithmetic operators (e.g., $\leq, +, -$) to be applied to pairs of vectors in an element-wise manner. We use the shorthand $[n]$ to denote the n-set $\{1, \ldots, n\}$ for $n \in N_+$.

3. Model

3.1. Low-rank ordinal tensor model

For the K-mode ordinal tensor $\mathcal{Y} = [\![y_\omega]\!] \in [L]^{d_1 \times \cdots \times d_K}$, we assume that its entries are realizations of independent multinomial random variables, such that

$$\mathbb{P}(y_\omega \leq \ell) = f(\theta_\omega + b_\ell), \quad \omega \in [d_1] \times \cdots \times [d_K]. \quad (1)$$

In this model, a twice differentiable function $f : \mathbb{R} \to [0, 1]$ is known and a cumulative distribution function and $\boldsymbol{b} = (b_1, \cdots, b_{L-1})$ is a set of unknown scalars satisfying $b_1 < \cdots < b_{L-1}$. We refer to $\boldsymbol{b}$ as the cut-off points and f the link function. The tensor $\Theta = [\![\theta_\omega]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is a hidden parameter which we are interested in. We assume that the entries of the parameter tensor Θ are continuous valued and Θ admits a rank $\boldsymbol{r} = (r_1, \cdots, r_K)$ Tucker decomposition,

$$\Theta = \mathcal{C} \times_1 \boldsymbol{M}_1 \cdots \times_K \boldsymbol{M}_N, \quad (2)$$

where $\mathcal{C} \in \mathbb{R}^{r_1 \times \cdots \times r_K}$ is a full rank core tensor and $\boldsymbol{M}_k \in \mathbb{R}^{d_k \times r_k}$, for $n \in [L]$ is a factor matrix with orthogonal columns. This low rank tensor structure allows us to reduce dimension and complete tensors from available entries.

The ordinal tensor model (1) has the equivalent representation as a latent model with L-level quantization. (Davenport et al., 2014; Lan et al., 2014; Bhaskar & Javanmard, 2015; Cai & Zhou, 2013) where the entry of $\mathcal{Y} = [\![y_\omega]\!]$ is a quantized value such that,

$$y_\omega = \mathcal{Q}(\theta_\omega + \epsilon_\omega), \quad \omega \in [d_1] \times \cdots [d_K], \quad (3)$$

where $\mathcal{E} = [\![\epsilon_\omega]\!]$ is a noise tensor i.i.d. entries from cumulative distribution function $\mathcal{F}_\epsilon$ such that $f(\theta) = \mathbb{P}(\epsilon \geq -\theta)$ and $\mathcal{Q} : \mathbb{R} \to [L]$ is a quantizer having the following rule.

$$\mathcal{Q}(x) = \ell, \quad \text{if } b_{\ell-1} < x \leq b_\ell, \quad \text{for all } \ell \in [L],$$

That is, the entries of observed tensor $\mathcal{Y}$ fall into category ℓ when the associated entries of the latent tensor $\Theta + \mathcal{E}$ fall into the ℓ-th interval of values. From this point of view, we can see the latent parameter Θ as the signal prior to contamination and quantization. We can diversify our model by the choices of f, or equivalently the distribution of $\mathcal{E}$. The followings are two common choices of f.

Example 1 (Logistic link/Logistic noise)**.** The logistic model is represented by (1) with $f(\theta) = \Phi_{log}(\theta/\sigma)$ where $\Phi_{log}(x/\sigma) = (1 + e^{-x/\sigma})$. Equivalently, the noise ϵ_ω in (3) follows i.i.d. logistic distribution with the scale parameter σ.

Example 2 (Probit link/Gaussian noise)**.** The probit model is represented by (1) with $f(\theta) = \Phi_{norm}(\theta/\sigma)$ where Φ_{norm} is the cumulative distribution function of the standard Gaussian. Equivalently, the noise ϵ_ω in (3) follows i.i.d.$N(0, \sigma^2)$.

3.2. Rank-constrained likelihood-based estimation

Our goal is to estimate unknown parameter tensor Θ and cut-off points $\boldsymbol{b}$ from observed tensor $\mathcal{Y}$ using a constrained likelihood approach. With a little abuse of notation, we use Ω to denote either the full index set $\Omega = [d_1] \times \cdots \times [d_K]$ or a random subset induced from the subsampling distribution. The log-likelihood function for (1) is

$$\mathcal{L}_{\mathcal{Y},\Omega}(\Theta, \boldsymbol{b}) = \sum_{\omega \in \Omega} \Big[\sum_{\ell \in [L]} \log(g_\ell(\theta_\omega)) \mathbb{1}_{\{y_\omega = \ell\}} \Big], \quad (4)$$

where $g_\ell(x) = \mathbb{P}(y_\omega = \ell) = f(\theta_\omega + b_\ell) - f(\theta_\omega + b_{\ell-1})$ from (1). We define $b_0 = -\infty, b_L = \infty$ so that $f(b_0) = 0, f(b_L) = 1$. The cut-off points $\boldsymbol{b}$ is implicitly contained in the function g_ℓ. Considering the Tucker structure in (2), we have the following constrained optimization problem.

$$\begin{aligned} &\max_{(\Theta, \boldsymbol{b}) \in \mathcal{D} \times \mathcal{B}} \mathcal{L}_{\mathcal{Y},\Omega}(\Theta, \boldsymbol{b}), \text{ where} \\ &\mathcal{D} = \{\Theta \in \mathbb{R}^{d_1 \times \cdots \times d_K} : \text{rank}(\Theta) \leq \boldsymbol{r} \text{ and } \|\Theta\|_\infty \leq \alpha\} \\ &\mathcal{B} = \{\boldsymbol{b} \in \mathbb{R}^{L-1} : b_1 < \cdots < b_{L-1}\}, \end{aligned} \quad (5)$$

for a given rank $\boldsymbol{r} \in {\mathbb{N}_+}^K$ and a bound $\alpha \in \mathbb{R}_+$. The search space $\mathcal{D}$ has two constraints on unknown parameter Θ. The first constraint ensures that the unknown parameter Θ admits the Tucker decomposition with rank $\boldsymbol{r}$. The second constraint makes the entries of Θ bounded by a constant α. This bound condition is a technical assumption to help to recover Θ in the noiseless case. Similar conditions has been imposed in many literatures for the matrix case (Davenport et al., 2014; Bhaskar & Javanmard, 2015; Cai & Zhou, 2013; Bhaskar, 2016). The search space $\mathcal{B}$ makes sure that the probability function g_ℓ in (4) is strictly positive.

3.3. Optimization

In this section, we describe the algorithm to seek the optimizer of (5). The objective function $\mathcal{L}_{\mathcal{Y},\Omega}(\Theta, \boldsymbol{b})$ is concave in $(\Theta, \boldsymbol{b})$ whenever $f(x)$ is log-concave (McCullagh, 1980; Burridge, 1981). However, the feasible set $\mathcal{D}$ is not a convex set, which makes the optimization (5) a non-convex problem. One approach to handle this problem is utilizing the Tucker decomposition and converting optimization into a block-wise convex problem. From (4) and (2), we have $K+2$ blocks of variables in the objective function, one for the cut-points vector $\boldsymbol{b}$, one for the core tensor $\mathcal{C}$ and K for the factor matrices $\boldsymbol{M}_K$'s. We can change the optimization problem to simple convex problem if any $K+1$ out of the $K+2$ blocks being fixed. Therefore, we can alternately update one block at a time while other blocks being fixed. The algorithm 1 gives the full description.

Algorithm 1 Ordinal tensor decomposition

Input: Ordinal tensor $\mathcal{Y} \in [L]^{d_1 \times \cdots \times d_K}$,
Rank $\boldsymbol{r} \in {\mathbb{N}_+}^{K-1}$,
Entry-wise bound $\alpha \in \mathbb{R}_+$.
Output: $(\hat{\Theta}, \hat{\boldsymbol{b}}) = \arg\max_{(\Theta, \boldsymbol{b}) \in \mathcal{D} \times \mathcal{B}} \mathcal{L}_{\mathcal{Y},\Omega}(\Theta, \boldsymbol{b})$.
Initialize Core tensor $\mathcal{C}^{(0)}$,
Factor matrices $\{\boldsymbol{M}_1^{(0)}, \cdots, \boldsymbol{M}_K^{(0)}\}$,
Cut-off points $\boldsymbol{b}^{(0)}$.
for t = 1,2,$\cdots$, **do**
  for k = 1,2,$\cdots$,K **do**
    Update $\boldsymbol{M}_k$ while fixing other blocks:
    $\boldsymbol{M}_k^{(t+1)} \leftarrow \arg\max_{\boldsymbol{M}_k \in \mathbb{R}^{d_k \times r_k}} \mathcal{L}_{\mathcal{Y},\Omega}(\boldsymbol{M}_k)$
    s.t. $\|\Theta^{(t+1)}\|_\infty \leq \alpha$, where $\Theta^{(t+1)}$ is the parameter tensor based on the current block estimates.
  end for
  Update $\mathcal{C}$ while fixing other blocks:
  $\mathcal{C}^{(t+1)} \leftarrow \arg\max_{\mathcal{C} \in \mathbb{R}^{r_1 \times \cdots \times r_k}} \mathcal{L}_{\mathcal{Y},\Omega}(\mathcal{C})$ s.t. $\|\Theta^{(t+1)}\|_\infty \leq \alpha$
  Update Θ based on the current block estimates:
  $\Theta^{(t+1)} \leftarrow \mathcal{C}^{(t+1)} \times_1 \boldsymbol{M}_1^{(t+1)} \cdots \times_K \boldsymbol{M}_K^{(t+1)}$
  Update $\boldsymbol{b}$ while fixing $\Theta^{(t+1)}$:
  $\boldsymbol{b}^{(t+1)} \leftarrow \arg\max_{\boldsymbol{b} \in \mathbb{R}^{L-1}} \mathcal{L}_{\mathcal{Y},\Omega}(\boldsymbol{b})$
end for
return $\Theta, \boldsymbol{b}$

3.4. Rank selection

Algorithm 1 takes the rank $\boldsymbol{r}$ as an input variable. In practice, the rank $\boldsymbol{r}$ is hardly known. Estimating an appropriate rank $\boldsymbol{r}$ from a given tensor is an important issue. We suggest to use Bayesian Information Criterion(BIC) to choose the rank.

$$\begin{aligned}\hat{\boldsymbol{r}} &= \arg\min_{\boldsymbol{r} \in \mathbb{N}_+^K} BIC(\boldsymbol{r}) \\ &= \arg\min_{\boldsymbol{r} \in \mathbb{N}_+^K} \Big[-2\mathcal{L}_{\mathcal{Y},\Omega}(\hat{\Theta}(\boldsymbol{r}), \hat{\boldsymbol{b}}(\boldsymbol{r})) + p_e(\boldsymbol{r}) \log\Big(\prod_{k \in [K]} d_k\Big)\Big],\end{aligned}$$

where $\hat{\Theta}(\boldsymbol{r}), \hat{\boldsymbol{b}}(\boldsymbol{r})$ is a maximum likelihood estimate given the rank $\boldsymbol{r}$, and $p_e() \stackrel{\text{def}}{=} \sum_{\mathrm{k}} (\mathrm{d_k} - \mathrm{r_k})\mathrm{r_k} = \prod_{\mathrm{k}} \mathrm{r_k}$ is the effective number of free parameters in the model. We select the rank $\hat{\boldsymbol{r}}$ that minimizes BIC value through the grid search method.

4. Real-world Data Applications

In this section, we apply our ordinal tensor decomposition method to two real-world datasets of ordinal tensors. In the first application, we use our model to analyze an ordinal tensor consisting of structural connectivity patterns among 68 brain regions for 136 individuals from Human Connectome Project (HCP) (Geddes, 2016). In the second application, we perform tensor completion from the data with missing values. The data tensor records the ratings from scale 1 to 5 of 42 users to 139 songs on 26 contexts (Baltrunas et al., 2011).

4.1. Human Connectome Project (HCP)

The human connectome project (HCP) is a $68 \times 68 \times 136$ tensor where the first two modes have 68 indices representing brain regions and the last mode has 136 indices meaning individuals. All the individual images were preprocessed following a standard pipeline (Zhang et al., 2018), and the brain was parcellated to 68 regions of interest following the Desikan atlas (Desikan et al., 2006). The tensor entries consist of $\{1, 2, 3\}$, the strength of fiber connections between 68 brain regions for each of the 136 individuals. We apply our ordinal tensor decomposition method with a logistic link function to the HCP data. The BIC result suggests $\boldsymbol{r} = (23, 23, 8)$ with $\mathcal{L}_{\mathcal{Y}}(\hat{\Theta}, \hat{\boldsymbol{b}}) = -216645.8$.

We check the performance of our model and compare with other methods: the continuous tensor decomposition method, the 1 bit-tensor completion method with probit link function and the 1 bit-tensor completion method with logistic link function. We use 5 folded cross validation method for the comparison. Specifically, we randomly split the tensor entries into 5 similar sized pieces and alternately use each piece of entries as a test data using other 80% entries as a training data. The entries in the test data are encoded as missing and then predicted based on each method from the training data. In our model, predict y_ω as $\hat{y}_\omega^{(\text{mode})} = \arg\max_\ell (g_\ell(\hat{\theta}_\omega, \hat{\boldsymbol{b}}))$. Table 4.1 shows RMSE, MAE and error(false prediction rate), averaged over 5 test set results. We can check that our ordinal tensor decomposition model outperforms the other methods in all criteria.

4.2. InCarMusic recommendation system

InCarMusic is a mobile application that offers music recommendation to passengers of cars based on contexts (Baltrunas et al. 2011). To be specific, the music recommenda-

METHOD	MSE	MAD	MCR
ORIDNAL TENSOR DECOMPOSITION	0.1503	0.1502	0.1502
CONTI TENSOR DECOMPOSITION	0.1603	0.1600	0.1598
1BIT-COMPLETION (PROBIT)	0.3658	0.3632	0.3619
1BIT-COMPLETION (LOGISTIC)	0.4514	0.4215	0.4066

Table 1. Results of comparisons among 4 methods on the HCP data predicting the test data. Four methods are the ordinal tensor decomposition algorithm, the continuous tensor decomposition algorithm and the 1 bit tensor completion method with probit link function and logistic link function. Each method is evaluated by RMSE, MAE, error(false prediction rate).

tions can be different according to the mood of the driver or the traffic condition in this system. Our goal is to perform the tensor completion to the $42 \times 139 \times 26$ ordinal tensor and thereby we can offer context-specific music recommendation to users. The tensor entries consist of ordinal observations on the scale 1 to 5, the ratings of 42 users to 139 songs on 26 contexts and are encoded as NA for missing values. The number of missing values is 148,904 and the number of available values is 2,884. We suggest three estimators $\hat{y}_\omega$ based on $(\hat{\theta}_\omega, \hat{\boldsymbol{b}})$:

- (Mean) $\hat{y}_\omega^{(\text{Mean})} = \sum_\ell \ell g_\ell(\hat{\theta}_\omega, \hat{\boldsymbol{b}})$;
- (Median) $\hat{y}_\omega^{(\text{Median})} = \min\{\ell \in [L] \colon f_\ell(\hat{\theta}_\omega, \hat{\boldsymbol{b}}) \geq 0.5\}$;
- (Mode) $\hat{y}_\omega^{(\text{Mode})} = \arg\max_\ell g_\ell(\hat{\theta}_\omega, \hat{\boldsymbol{b}})$

Note that, under the ordinal tensor model (1), the estimator $\hat{y}_\omega^{(\text{Mean})}$ minimizes $\mathbb{E}_{\hat{\theta}_\omega, \hat{\boldsymbol{b}}}(y_\omega - y)^2$, the estimator $\hat{y}_\omega^{(\text{Median})}$ minimizes $\mathbb{E}_{\hat{\theta}_\omega, \hat{\boldsymbol{b}}}|y_\omega - y|$ and the estimator $\hat{y}_\omega^{(\text{Mode})}$ minimizes $\mathbb{E}_{\hat{\theta}_\omega, \hat{\boldsymbol{b}}}\mathbb{1}_{\{y_\omega = y\}}$. In contrast, these three estimators degenerate to a single estimator under the continuous-valued tensor decomposition model. We assess the accuracy using three metrics: Mean squared error (MSE), Mean absolute deviation (MAD), and Misclassification rate (MCR).

Results are averaged from 5 fold cross validation. add s.e. to the above table.

Critera	Our method			Continuous method
	$\hat{y}_\omega^{(\text{Mean})}$	$\hat{y}_\omega^{(\text{Median})}$	$\hat{y}_\omega^{(\text{Mode})}$	(?)
MSE	2.01	2.14	3.64	8.18
MAD	1.23	1.13	1.26	2.36
MCR	0.80	0.75	0.54	0.96

Table 2. Comparison of tensor completion performance on InCarMusic dataset.

We also attempted to run matrix ordinal mehod (Bhaskar, 2016) to the InCarMusic dataset. Unfortunately, the $\boldsymbol{b}$ cannot be estimated from their method.

References

Adomavicius, G. and Tuzhilin, A. Context-aware recommender systems. In *Recommender systems handbook*, pp. 217–253. Springer, 2011.

Baltrunas, L., Kaminskas, M., Ludwig, B., Moling, O., Ricci, F., Aydin, A., Lüke, K.-H., and Schwaiger, R. Incarmusic: Context-aware music recommendations in a car. In *International Conference on Electronic Commerce and Web Technologies*, pp. 89–100. Springer, 2011.

Bhaskar, S. A. Probabilistic low-rank matrix completion from quantized measurements. *The Journal of Machine Learning Research*, 17(1):2131–2164, 2016.

Bhaskar, S. A. and Javanmard, A. 1-bit matrix completion under exact low-rank constraint. In *2015 49th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6. IEEE, 2015.

Burridge, J. A note on maximum likelihood estimation for regression models using grouped data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 43 (1):41–45, 1981.

Cai, T. and Zhou, W.-X. A max-norm constrained minimization approach to 1-bit matrix completion. *The Journal of Machine Learning Research*, 14(1):3619–3647, 2013.

Cichocki, A., Mandic, D., De Lathauwer, L., Zhou, G., Zhao, Q., Caiafa, C., and Phan, H. A. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE signal processing magazine*, 32(2):145–163, 2015.

Davenport, M. A., Plan, Y., Van Den Berg, E., and Wootters, M. 1-bit matrix completion. *Information and Inference: A Journal of the IMA*, 3(3):189–223, 2014.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980, 2006.

Geddes, L. Human brain mapped in unprecedented detail. *Nature International weekly journal of science*, 2016.

Harshman, R. A. et al. Foundations of the parafac procedure: Models and conditions for an" explanatory" multimodal factor analysis. 1970.

Hitchcock, F. L. Multiple invariants and generalized rank of a p-way matrix or tensor. *Journal of Mathematics and Physics*, 7(1-4):39–79, 1928.

Hong, D., Kolda, T. G., and Duersch, J. A. Generalized canonical polyadic tensor decomposition. *arXiv preprint arXiv:1808.07452*, 2018.

Hore, V., Viñuela, A., Buil, A., Knight, J., McCarthy, M. I., Small, K., and Marchini, J. Tensor decomposition for multiple-tissue gene expression experiments. *Nature genetics*, 48(9):1094, 2016.

Hu, Q., Li, G., Wang, P., Zhang, Y., and Cheng, J. Training binary weight networks via semi-binary decomposition. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 637–653, 2018.

Kutty, S., Chen, L., and Nayak, R. A people-to-people recommendation system using tensor space models. In *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, pp. 187–192, 2012.

Lan, A. S., Studer, C., and Baraniuk, R. G. Matrix recovery from quantized and corrupted measurements. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4973–4977. IEEE, 2014.

McCullagh, P. Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):109–127, 1980.

Miwakeichi, F., Martınez-Montes, E., Valdés-Sosa, P. A., Nishiyama, N., Mizuhara, H., and Yamaguchi, Y. Decomposing eeg data into space–time–frequency components using parallel factor analysis. *NeuroImage*, 22(3):1035–1045, 2004.

Nickel, M., Tresp, V., and Kriegel, H.-P. A three-way model for collective learning on multi-relational data. In *Icml*, volume 11, pp. 809–816, 2011.

Sidiropoulos, N. D., Bro, R., and Giannakis, G. B. Parallel factor analysis in sensor array processing. *IEEE transactions on Signal Processing*, 48(8):2377–2388, 2000.

Smilde, A., Bro, R., and Geladi, P. *Multi-way analysis: applications in the chemical sciences*. John Wiley & Sons, 2005.

Sun, J., Papadimitriou, S., Lin, C.-Y., Cao, N., Liu, S., and Qian, W. Multivis: Content-based social network exploration through multi-way visual analysis. In *Proceedings of the 2009 SIAM International Conference on Data Mining*, pp. 1064–1075. SIAM, 2009.

Sun, W. W., Lu, J., Liu, H., and Cheng, G. Provable sparse tensor decomposition. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):899–916, 2017.

Tucker, L. R. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.

Tucker, L. R. and Tucker, L. The extension of factor analysis to three-dimensional matrices. 1964.

Wang, M. and Li, L. Learning from binary multiway data: Probabilistic tensor decomposition and its statistical optimality. *arXiv preprint arXiv:1811.05076*, 2018.

Zare, A., Ozdemir, A., Iwen, M. A., and Aviyente, S. Extension of pca to higher order data structures: An introduction to tensors, tensor decompositions, and tensor pca. *Proceedings of the IEEE*, 106(8):1341–1358, 2018.

Zhang, Z., Descoteaux, M., Zhang, J., Girard, G., Chamberland, M., Dunson, D., Srivastava, A., and Zhu, H. Mapping population-based structural connectomes. *NeuroImage*, 172:130–145, 2018.