Beyond matrices: High-dimensional tensor methods meets data sciences

Miaoyan Wang, UW-Madison

My research is in the intersection of statistics, machine learning, and optimization, with a focus on high-dimensional tensor data science. Rapid developments in modern technologies have made multiway data readily available in daily lives. Methods built on tensors provide powerful tools to capture complex structures that lower-order (i.e. matrices) methods fail to exploit. My research on higher-order tensors combines mathematical foundations from multilinear numerical analysis with scalable optimizations from computer science to create new application horizons in biomedical data sciences.

Challenges. High dimensional tensor methods are making enormous impacts on science and society. High-dimensional statistics focuses on datasets in which the number of parameters is magnitude larger than the number of observations. Data sets of this type present a variety of new challenges, since classical theory and methodology can break down in surprising and unexpected ways: decisions arising from classical statistical premises are sensitive to model misspecification; greedy algorithms are yielding biased solutions; and concerns on prediction-interpretability balance continue to mount. Theses challenges can be addressed only by an appeal to a genuine combination of multidisciplinary approaches.

Research summary. My past and current research can be divided into two main directions. The first series concern mathematical foundations of machine learning methods, i.e., *the statistical and algorithmic trade-off* in high-dimensional tensor problems including structured estimation, regression, and classification; this work has so far won me various honors at top-notch venues (best paper awards, and young researcher career awards). The second line concerns modeling challenges for multimodal biomedical data integration. It has been my favorite part of research working with diverse collaborators, using ideas from different fields in unexpected ways, and seeing that my work has been useful and well received across these communities.

(1) **Mathematical foundations of machine learning.** Most of my recents focus on key notions of complexity to characterize the difficulty of estimation and prediction tasks from various perspectives, including sample complexity, oracle complexity, and computational (runtime) complexity. I have characterizes trade-offs among these notions in several contexts, establish complexity bounds for empirical risk minimization problems, bring complexity bounds more closely into line with practical performance for several important optimization algorithms, and improve complexity of sampling from distributions. These results are important, not only because they improve on the state-of-the art, but because they introduce new machinery that beats the prior arts, and thus potentially opens up new inquiry for many open questions in this field.

(2) **Applications to biomedical data sciences.** Another major line of my work focuses on modeling challenges arise in high-dimensional genomics data integration. I have been actively contributing two consortium efforts: Genotype-Tissue Expression (GTEx) studies and Human Connectome Project (HCP). Methodological innovations I develop include new estimators for spectral properties of matrices, randomized procedures for sketching and optimization, as well as algorithms for decision-making in sequential settings.

Future vision. Looking forward, I will leverage my interdisciplinary background to create arising from science and engineering. I have a unique combination of training background in mathematics (2006-2010), statistics (2011-2015), genetics (2015-2018), and computer science (2015-2018). My research, currently and over the next few years, will focus on developing a suite of statistical learning theory, efficient algorithms, and data-driven solutions for high-dimensional data problems. At a higher level, I strive to creating mathematical abstraction of practical systems, understanding system characteristic, and guiding design from upstream to operation.

The PI will integrate the research and eduction goals by establishing a unique research-training experience, thereby broadening participation in STEM fields. Comprehensive plans for education and outreach will draw on PI's past success and leverage institutional resources through the four-university TRIPODS and Data Science Hub that the PI is affiliated with. The PI will organize and host summer research positions that engage a diverse student body at all levels of seniority. New courses on *fairness in data science* will be created in both traditional and online formats. The course focuses on critical thinking skills and development of equitable access for underrepresented groups. The PI is committed to training students the interdisciplinary skills for them to make independent, insightful contributions to the larger society. Collaborations with domain science researchers in academia and non-profit labs will continue through the established channels. The research-training agenda is well integrated into the education, which as a whole, creates an inclusive environment for every student to succeed in their own individual careers.

High-dimensional Tensor Learning: The Good, the Bad, and the Pragmatic

Miaoyan Wang, UW-Madison

1 Research Goals and Significance

My proposed research is in the intersection of statistics, machine learning, and optimization, with a focus on tensor data analysis. Rapid developments in modern technologies have made multiway data readily available in daily lives. Tensor provides a generalized data structure in many learning procedures. Methods built on tensors provide powerful tools to capture complex structures that lower-order methods fail to exploit. The ambitious goal in my proposal is to create new horizons in tensor methods and applications across the disciplines of mathematics, computer science, and biomedical sciences.

Below I give two examples of biomedical tensor data arising from my current collaborations. I will propose the core fundamental challenges based on the questions raised in the applied endeavors.

Multi-tissue, multi-individual gene expression. The recent completion of Genotype-Tissue Expression (GTEx, Figure 1a) project has provided an unprecedented opportunity to investigate transcriptome diversity and complexity. A typical multi-tissue experiment collects gene expression profiles from different individuals in a number of tissues. The study results in a huge compendium of tensor data consisting of millions of expression measurements from over 20K genes across 544 individuals and 53 human tissues. My previous work [1] reveals that the variation in the expression levels is attributable to genes, tissues, individuals, and interactions thereof. Understanding the multiway patterns of whole-genome transcriptome variation is crucial to unravel unravel genomic basis for tissue functions and personalized medicines.

Multimodal neuroimaging data analysis. The Human Connectome Project (HCP, Figure 1b) consists of massive datasets representing the anatomical and functional connectivities across brain regions from over 1,200 individuals. Multimodal brain connectivity networks are constructed based on various measurements, including functional magnetic resonance imaging, electroencephalography, and diffusion tensor imaging. My previous work [2, 10] provides strong evidence for inter- and intra-modal correlation in the brain network data. Efficient multiway methods are essential for further investigating the structural and functional connectivities in human brains.
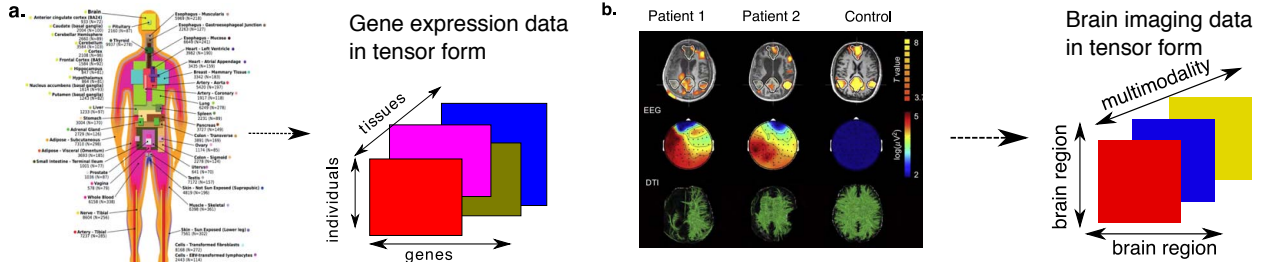

Figure 1: (a) GTEx collects gene expression profiles of over 20,000 genes from 544 individuals across 53 human tissues. (b) HCP collects multimodal imaging data from over 1,200 individuals. Figures are reproduced based on [20, 42].

In the above two examples and many other applications, scientists are interested in identifying interpretable latent structure within the high-dimensional tensor data. The research question goes beyond the traditional multivariate analysis: we are interested in the distribution over multi-indexed entries in order to capture the multilinear relations among variables. Recent years have seen an upsurge of interest in tensor data analysis. However, the empirical success has uncovered a myriad of new and pressing challenges. Multidisciplinary efforts are necessary to address these challenges.

- **Old premises fail in new settings.** High-dimensional tensor models and multilinear algorithms can yield unexpected behaviors that challenge classical statistical premises. Notable examples include *double descent*

(i.e., no bias-variance trade-off in overparameterized tensor models [19, 43]) and *benign non-convexity* (no spurious local optima in tensor factorization [28]). Classical statistical premised need be revolutionized to address the potential and limitations of new settings.

● **Multiple complexity measures interact.** Complexity is the fundamental tool in modern data science to characterize the difficulty of learning tasks. Statistics use *model complexity* to measures the number of samples needed for accurate estimation. Computer sciences use *computational complexity* to measure the number of arithmetic operations required in optimization algorithms. The fundamental challenges of tensor problems lie in the computational-statistical tradeoffs, in which model complexity hinges on and interacts with computational complexity. Current mechanisms for optimizing these tradeoffs are limited.

● **One paradigm may not fit all.** Off-the-shelf tensor data analytics tools are increasingly challenged by important domain applications [37]. Parametric model facilitates statistical inference but often lacks the robustness; nonparametric model boosts the prediction but hinders the interpretation; classical asymptotic theory holds only for exponential time algorithms but is brittle with computational constraints. In these practical situations, we need to carve out the regimes for which a learning approach succeeds and/or fails. Challenges associated with ever-growing tensor applications await breakthroughs in understanding.

The proposed project is to develop a suite of statistical learning theory, efficient algorithms, and data-driven solutions for high-dimensional tensor problems. The developed tools will allow domain scientists to examine complex interactions among tensor entries and between multiple tensors, thereby providing solutions to questions that cannot be addressed by traditional analysis. The detailed plan of my research agenda is organized into three core themes:

Aim 1. Parametric tensor models with statistical and computational optimality. Block tensors arise frequently in neuroimaging, recommendation system, topic modeling, and sensor network localization. Despite the popularity, little is known about the fundamental limits in tensor block models. The block tensor serves as a meta structure to most—if not all—common models including the low-rankness [45], latent space models [2], and isotonic tensors [38]. We plan to use tensor block models as a lens to explore statistical and computational trade-off in parametric structured tensor problems. We will fully characterize, for the first time, the critical signal-to-noise threshold in block tensor models, revealing the intrinsic distinctions between (vector) clustering, (matrix) biclustering, and (tensor) higher-order clustering. The results generalize existing low-rank methods and provide solutions that were previously impossible. See Section 2.

Aim 2. Beyond low-rankness: nonparametric estimation and completion for high-rank tensors. Parametric tensor models (e.g., low-rank models, block models) aim to explain data with a finite number of parameters; these methods are useful when the sample outsizes the parameters. Nonparametric models, on the other hand, use infinite number of parameters to allow growing model complexity as sample increases. In this vein, we aim to develop nonparametric tensor models to effectively address both low- and high-rank structures. We will develop a new complexity framework to capture nonlinear high-rank signals, while encompassing existing models—including CP models [33], Tucker models [23], single index models [27]—as special cases. The approach leads to a substantially richer class of tensor models in many learning problems, including regression, completion, multi-task learning, and compressed sensing. See Section 3.

Aim 3. Predictive tensor models with auxiliary information. Predictive learning models are useful in daily decision making tasks, i.e., predicting diseases based on multimodal neuroimaging, classifying documents based on their contents, and deciding target customers based on shopping behaviors in recommendation system. These problems are often formulated as predicting the response variable Y_{new} from explanatory variable X_{new}, where the sample available is in the form of pairs $(X_i, Y_i)_{i=1,\dots,n}$. We plan to combine tensor structure models with predictive tasks, by jointly estimating the latent structure in tensors as well as the relationship between the tensor data and auxiliary features. The outlined direction builds new links between machine learning methodology and scientific discovery, and also produces new areas in which machine

learning theory and applications complement each other. See Section 4.

PI's qualification and prior supports. The PI is a young faculty member in the Department of Statistics at UW-Madison. The PI is also affiliated with Institute for Foundations of Data Science (IFDS), an NSF-funded TRIPOD Phase II collaboration initiative between four universities. The PI has an excellent track record and a unique combination of training background in mathematics (2006-2010), statistics (2010-2015), genetics (2015-2018), and computer science (2015-2018). The PI's current work is supported by NSF MDS-1915978 "spectral methods for high dimensional tensor data" (Sep 2019 - Aug 2022, $179,349 in total).

The PI has built tremendous momentum since the tenure track position. The PI is prolific in publishing papers in top methodological venues (NeurIPS, ICML, etc [3, 8, 10, 11, 13]), applied venues (PNAS, AOAS, etc [1, 4, 7, 12, 15]), and theory venues (JMLR, AOS, etc [2, 5, 6, 9, 14]). Most of these publications are exclusively with PI's students; this has demonstrated the independence and leadership of PI in research and education. The past NSF grant has supported the PI and the students to deliver over 40 talks at departmental seminars in the US, Canada, and Europe, as well as invited talks at international conferences including JMS, AMS, ENAR, IMS, and SIAM. The PI's research group has developed 15 open-source software packages in R, Matlab, and C. All packages are regularly maintained on CRAN and Github.

As a young female faculty member, the PI has demonstrated effective integration of research and education. The PI is leading a diverse research group, and has served as co-advisor/committee for students in Statistics, Industrial Engineering, and Civil Engineering. The PI has won **two Best Student Paper Awards (with PI as the advisor)** from American Statistical Association in 2021, the Madison Teaching and Learning Excellence Fellow, and multiple prestigious young researcher awards in statistics, machine learning, and genetics. A highlight of the PI's achievement is to engage undergraduate students from underrepresented population into cutting-edge research; these projects have resulted in publications at NeurIPS (the best venue in machine learning). The PI's research/eduction integration has attached media coverage and been featured in the online article *"Women in STEM: 5 Thoughtful Ways to Recruit and Retain Them"* [16].

The agenda in this proposal has broader impacts on multiple research fields and education. The PI's work has a transformational nature involving connections across diverse disciplines, and the PI strives to create a welcoming venue for diverse trainees, thereby broadening participation in STEM. The proposal lays out an ambitious plan for a full spectrum of parametric, non-parametric, high-rank, high-dimensional tensors with multimodal data integration. Education and research will be integrated in the form of developing new courses, creating summer research-training positions, and engaging underrepresented students in outreach. Theory and application will be combined with the emphasize on data-intensive problems in social network science and biomedical science. More detailed plan will be described in Section 5.

2 AIM 1. Parametric Tensor Models with Statistical and Computational Optimality

This section focuses on unsupervised clustering methods for high-dimensional tensors, where the learner has little or no label information during training. The modeling is notably difficult due to limited information. The proposed research provides tools for optimizing trade-off between statistical and computational efficiency in parametric tensor models.

We focus on tensors of order 3 or greater, known as higher-order tensors. A tensor $\Theta \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is a higher-order generalization of matrix that obeys the mode-k multiplication rule [33] denoted by $\times_k$: $\Theta \mapsto \Theta \times_1 \boldsymbol{M}_1 \times_2 \cdots \times_K \boldsymbol{M}_K$, for matrices $\boldsymbol{M}_k$ of consistent dimensions. Higher-order tensors are not simply matrices with more indices; rather, they are data objects whose structures remain covariant under multilinear transformations. Key challenges of multilinear nonconvesity await breakthroughs in understanding.

2.1 Motivation

A central theme in modern data analysis is to find a low-dimensional representation to better understand, compress, and convey the key phenomena buried in noisy observations. Dimension reduction is crucial for large-scale tensor data analysis. The most popular model in tensor dimension reduction is the multilinear low-rank model. An order-K $(d_1, \ldots, d_K)$-dimensional tensor Θ is called Tucker low-rank if it can be represented by some low-dimensional core tensor $\mathcal{S}$ under a change of basis

$$\Theta = \mathcal{S} \times_1 \boldsymbol{M}_1 \times_2 \cdots \times_K \boldsymbol{M}_K$$

for some transformation matrices $\boldsymbol{M}_k \in \mathbb{R}^{d_k \times r_k}$ with $r_k \ll d_k$ along each of the modes (see Figure 2c).

Low-rank formulation has been actively studied in various contexts, such as tensor regression [11, 46], tensor completion [10, 26], tensor PCA [5, 48], and generalized tensor learning [31]. A special but more fundamental low-rank model, *multiway tensor block model*, however has yet to be investigated. In comparison with general low-rankness, the discrete block structure provides better interpretation, because the transformation matrix $\boldsymbol{M}_k$ encodes the clustering membership along the mode k of the tensor. Figure 2a shows an order-3 tensor with a block structure, in which each of the modes is partitioned into several clusters. The goal is to identify the block structure, as well as to recover the true signal tensor, from a noisy observation. The tensor block model and higher-order clustering arise commonly in practical applications. For example,

● *Multi-tissue Gene Expression Analysis.* Gene expression profiles such as scRNA-seq and microarrays are collected from multiple individuals across numbers of tissues [1, 35]. Genes involved in the same biological function exhibit similar expressions within a group of tissues and individuals, while these expression values vary from group to group. Similarly, tissues and individuals exhibit gene-specific clustering patterns. Identifying complex interactions among these three entities is of great interest to biomedical research.
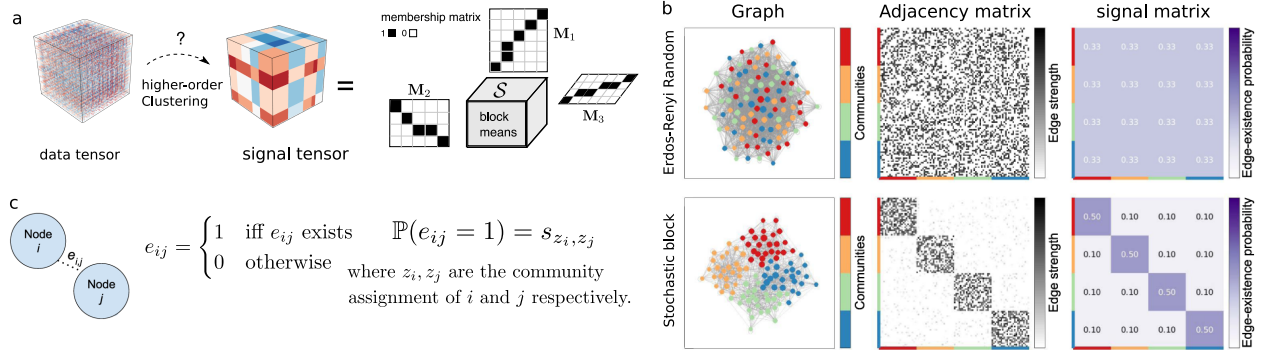

Figure 2: (a) Tensor block model generalizes usual matrix block model and is useful for higher-order clustering [9]. (b) Examples of common social connectivity networks [24]. (c) Representation of connectivity edges in the usual stochastic (matrix) block model.

● *Multilayer Network Analysis.* A multi-layer network consists of multiple undirected graphs (or adjacency matrices), where each graph represents the connection among the same set of vertices (Fig 2b-c). The data structure is naturally organized as an order-3 tensor with the first two modes being vertices and the third mode being the contexts under which the graph is observed. Multilayer networks arise commonly in longitudinal study of network [34] and multi-relational data [8, 11]. While the community structure in single-layer network has been widely analyzed in the literature, little work has studied the heterogeneous pattern across multiple layers. The connectivity within/between communities may change correspondingly as the context changes, and it is important to simultaneously obtain the clusters across different contexts.

● *Online Click-through Prediction.* In e-commerce, predicting click-through for user-item pairs in a time-specific way plays an important role in online recommendation system [41]. The click-through data can be

organized as an order-3 tensor, where each entry indexed by (users, items, time) represents the presence or absence of user-item action across 24 hours in a day. The users and items often exhibit clustering structures due to user features and item attributes, respectively. The shopping behavior also varies in time, and this heterogeneity depends on the specific groups of users and items. For instance, salarymen tend to have higher activities than non-office workers in the period of after-work hours.

2.2 Multiway Tensor Block Models

Methods built on tensors provide generalized tools to capture complex data structure. For cleaner illustration, we consider balanced setting with equal dimension and cluster size in each of the tensor modes. Let $\mathcal{Y} \in \mathbb{R}^{d \times \cdots \times d}$ be an order-K data tensor of interest. The main assumption of tensor block model is that the observed data tensor $\mathcal{Y}$ is a noisy realization of a signal tensor that exhibits a checkerbox structure (see Figure 2a). Specifically, suppose there are r clusters on each of the K modes, and we represent the clustering along the model k by a vector $\boldsymbol{z}_k = (z_{k,1}, \ldots, z_{k,d})^T \in [r]^d$, for all $k \in [K]$. Under the block structure, the entries $\mathcal{Y}(i_1, \ldots, i_d)$ are realizations from the model

$$\mathcal{Y}(i_1, \ldots, i_K) = \mathcal{S}(z_{1,i_1}, \ldots, z_{K,i_K}) + \mathcal{E}(i_1, \ldots, i_K), \quad \text{for all } (i_1, \ldots, i_K) \in [d]^K, \tag{1}$$

where $\mathcal{S} \in \mathbb{R}^{r \times \cdots \times r}$ is a low-dimensional core tensor, and $\mathcal{E}$ is a noise tensor consisting of i.i.d. zero-mean Gaussian entries with variance σ^2. Model (1) can be equivalently written in tensor-matrix product:

$$\mathcal{Y} = \mathcal{S} \times_1 \boldsymbol{M}_1 \times \cdots \times_K \boldsymbol{M}_K + \mathcal{E}, \tag{2}$$

where $\boldsymbol{M}_k \in \{0,1\}^{d \times r}$ is the membership matrix associated with $\boldsymbol{z}_k$ such that $\boldsymbol{M}_k(i,j) = 1$ if and only if $z_{k,i} = j$. That is, $\boldsymbol{M}_k$ has one copy of 1 and $(r-1)$ copies of 0's in each of the rows. We use forms (1) and (2) exchangeably for notational convenience.

The learning goal is to estimate multiway connection signal $\mathcal{S}$ and clustering assignments $\boldsymbol{M}_k$ from a noisy observation $\mathcal{Y}$. We focus on two inference tasks for tensor block model:

Open problem 1.
- (Clustering) How to recover the membership matrices $\boldsymbol{M}_k$, or equivalently the label vectors $\boldsymbol{z}_k$?
- (Denoising) How to accurately estimate the underlying signal tensor $\mathcal{S}$?

2.3 Proposed Agenda: Statistical and Computational Phase Transition

My initial investigation reveals a surprising but simple characterization for the models (1) and (2).

Definition 2.1 (A key quantity for hardness of the problem)**.** The statistical and computational hardness of tensor block model is completely determined by the following signal-to-noise ratio (SNR),

$$\text{SNR} = \Delta_{\min}^2 / \sigma^2, \quad \text{where } \Delta_{\min}^2 := \min_{k \in [K]} \min_{j_1 \neq j_2} \|\text{Unfold}_k(\mathcal{S})_{j_1:} - \text{Unfold}_k(\mathcal{S})_{j_2:}\|_F^2,$$

where $\text{Unfold}_k(\mathcal{S})_{j:}$ denotes the j-th row of the matrix obtained by unfolding the tensor $\mathcal{S}$ along mode k.

We aim to investigate the intriguing interplay between the statistical optimality and computational efficiency. Figure 3a summarizes our preliminary findings on phase transition with respect to different SNR regions in tensor block model. In the strong SNR region A, a higher-order Lloyd algorithm developed by my group achieves exact clustering *in polynomial time*. We find that the estimation error bound of the target tensor is *free of tensor dimension*. This feature is appealing in modern large-scale tensor analysis. In the weak SNR region B, we discover the information-theoretical impossibility of higher-order clustering. In the modest SNR region C, the problem exhibits a gap between computational and statistical limits. Remarkably,

the statistical-computational gap emerges only for higher-order tensors with $K \geq 3$, revealing the intrinsic distinctions between (matrix) biclustering and (tensor) higher-order clustering.

Theorem 2.1. [WZ'19 [3], HLWZ on-going work] Consider the higher-order clustering on an order-K dimensional-d tensor under the tensor block model.

- When $\mathrm{SNR} \gg d^{-K/2}$, the proposed higher-older Lloyd algorithm performs exact clustering;
- When $\mathrm{SNR} \ll d^{-(K-1)}$, no algorithm can achieve exact clustering;
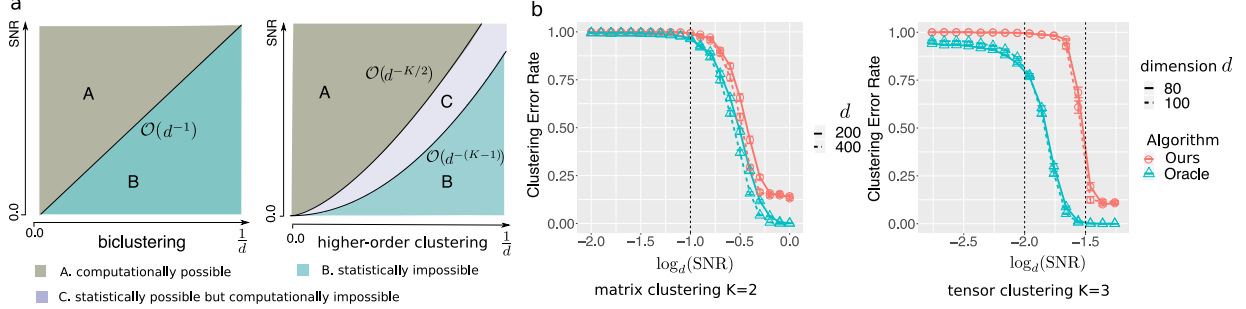- When $d^{-(K-1)} \ll \mathrm{SNR} \ll d^{-K/2}$, no polynomial-time algorithm exists for exact clustering.

Figure 3: (a) Three SNR regimes of statistical and computational behaviors for the order-K dimensional-d tensor block problems. A statistical-computational gap arises only for tensors of order 3 or greater. (b) Simulation confirmation of the discovered gap.

To our knowledge, we are among the first to explore the statistical and computational limits for tensor block model. We conjecture that the threshold $d^{K/2}$ is a necessary SNR condition for polynomial-time label recovery. Figure 3b shows the empirical results based on our simulations. In the matrix case, the polynomial estimate (ours) and statistical optimal rate (oracle) undergo similar SNR phase transitions. In contrast, the order-3 tensor clustering reveals a striking gap between polynomial estimate and oracle estimate. Similar puzzles about statistical-computational gaps have attracted enormous interests in modern, non-convex problems [36, 39]. We aim to provide rigorous evidence for the following conjecture:

> **Open problem 2.** (Conjecture) There exists no polynomial-time algorithm that exactly recovers the block labels in tensor block model when $\mathrm{SNR} = \mathcal{O}(d^{-K/2-\varepsilon})$, for any $\varepsilon > 0$.

2.4 Preliminary Results and Assessing Mechanisms

Classical statistical theory often focuses on maximum likelihood estimation (MLE). Unfortunately, MLE requires exponential time complexity and is practically less useful for tensor problems. We are currently investigating an efficient polynomial-time algorithm based on the recent advances in non-convex optimization. We propose to combine spectral initialization and alternating optimization for locating local optimum with desired statistical accuracy (Figure 4a). Our preliminary analysis shows that the proposed Algorithm 1 achieves *exact recovery* of block labels under relaxed assumptions than existing algorithms.

Algorithm 1 Multiway clustering based on tensor block model

Input: Data tensor $\mathcal{Y} \in \mathbb{R}^{d \times \cdots \times d}$, clustering size $(r, \ldots, r)$.
Output: Block mean tensor $\hat{\mathcal{S}} \in \mathbb{R}^{r \times \cdots \times r}$, and the membership matrices $\hat{\boldsymbol{M}}_k \in \{0,1\}^{r \times d}$.
  1: Initialize the marginal clustering by higher-order spectral clustering.
  2: **for** $t = 1, 2, \ldots$ **do**
  3:     Update the core tensor $\hat{\mathcal{S}}$ using the sample averages within each multi-way block.
  4:     Update each membership matrix $\hat{\boldsymbol{M}}_k$ by a small nearest neighbor search over r discrete points.
  5: **end for**

My preliminary results reveal a two-component error in the clustering error bound.

Theorem 2.2 (Statistical and computational accuracy). The output from Algorithm 1 satisfies

$$\text{Error at step } (t+1) \lesssim \underbrace{\sigma^2 \exp\left(-\frac{d^{K-1}}{r^{K-1}}\text{SNR}\right)}_{\text{statistical error}} + \underbrace{\rho^t \text{ Error at step } 0}_{\text{computational error}}.,$$

where $\rho \in (0,1)$ is a contradiction parameter independent of problem dimensions, and $t = 1,2,\ldots$ denotes the iteration number in the algorithm. This bound reveals the interesting interplay between the computational and statistical errors, and the result immediate suggests a practical tradeoff between these two aspects. One of my main research goals is to understanding the gap between the algorithm property and statistical optimality, which will in turn offer a useful guide to algorithm designs.

Our experience indicates that parametric tensor models often exhibit the *benign nonconvexity* as shown in Theorem 2.2 and Figure 4a. That is, a well-designed local solution achieves similar statistical accuracy as the global MLE, albeit a highly non-convex optimization algorithm. Examples include tensor SVD [2], tensor completion [10], and tensor regression [31]. We plan to explore the optimization landscape for a broader range of problems, including exponential-family tensors, orthogonal decomposable tensors, methods-of-moment tensors, deep tensor neural networks, and tensor-based reinforcement learning.

Open problem 3. Can we characterize the full class of specially-structured tensors for which efficient, provable algorithms exist?

My group will assess our proposed research by empirical performance on real data. Figure 4b shows the application of our method to flight route data from OpenFlights database. The values in the depicted matrix represent the connectivity of airports based on our clustering results. We find that US airlines tend to show highly concentrated connections among US, Europe, and Asia airports. In contract, mixture airlines tend to have dense but low connections. This reflects the nature of this cluster consisting of many regional airports around the world. The results demonstrate the applicability and potential of our method.
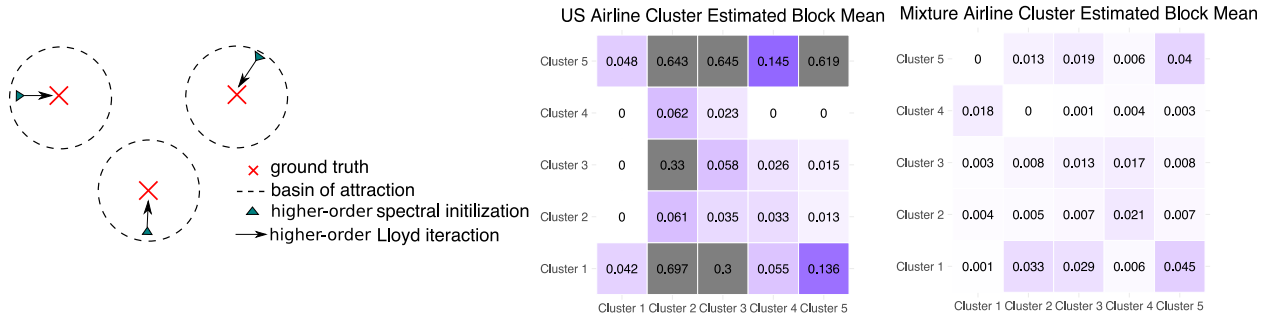

Figure 4: (a) Optimization landscape of tensor block model. (b). Application of higher-order clustering to flight routes data.

Significance and transformative concepts:
- Establish statistical learning foundations for unsupervised tensor clustering and dimension reduction.
- Design robust and efficient algorithms for large-scale tensor data analysis.
- Create data-driven solutions and pipelines for tensor data pattern recognition.

3 AIM 2. Beyond Low-rankness: Nonparametric Estimation for High-rank Tensors

Parametric tensor methods often rely on two assumptions: (i) a known relationship between observed data and latent low-rank representation, and (ii) a global low-rank structure across all tensor entries.

This section develops a new nonparametric approach to tensor estimation by relaxing these two assumptions. The framework leads to a substantially larger class of signal tensor models than previously possible.

3.1 Motivation

While low-rank methods have shown great success in theory, tensors in applications often violate the parametric form in the model. We provide two examples to illustrate the limitation of classical models.

The first example reveals the sensitivity of tensor rank to order-preserving transformations (Figure 5a). Let $\mathcal{Z} \in \mathbb{R}^{30 \times 30 \times 30}$ be an order-3 tensor with tensor rank$(\mathcal{Z}) = 3$. Suppose a monotonic transformation $f(z) = (1 + \exp(-cz))^{-1}$ is applied to $\mathcal{Z}$ entrywise, and we let the signal Θ be the tensor after transformation. Figure 5a plots the numerical rank of Θ versus c. As we see, the rank increases rapidly with c, rendering traditional low-rank tensor methods ineffective in the presence of mild order-preserving nonlinearities. In many applications, the tensor of interest often undergoes unknown transformation prior to measurements. The sensitivity to transformation makes the parametric model less desirable in practice.
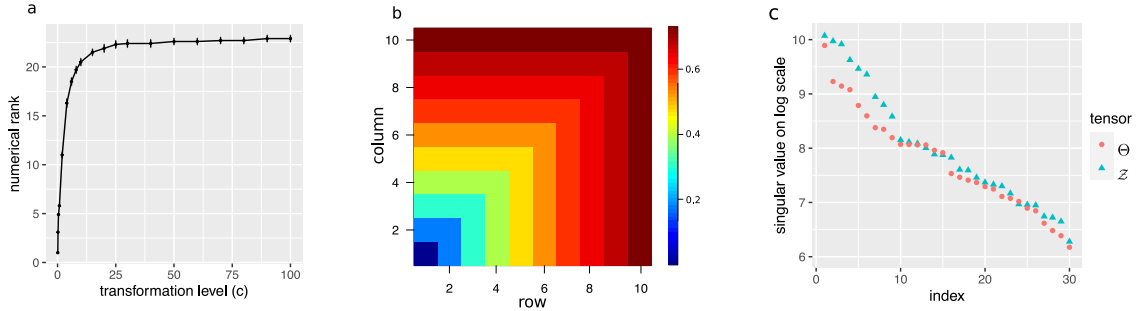

Figure 5: High-rank tensor examples. (a) Numerical rank of tensor $g(\mathcal{Z})$ versus c in the transformation. (b) Heatmap of a full-rank matrix $\Theta \in \mathbb{R}^{d \times d}$ with (i,j)-th entry equal to $\log(1 + \log(i,j))$. (c) Top $d = 30$ tensor singular values in the second example.

The second example demonstrates the inadequacy of classical low-rankness in representing special structures (Figure 5b-c). Here we consider the signal tensor of the form $\Theta = \log(1 + \mathcal{Z})$, where $\mathcal{Z} \in \mathbb{R}^{d \times d \times d}$ is an order-3 tensor with entries $\mathcal{Z}(i,j,k) = d^{-1} \max(i,j,k)$ for $i,j,k \in \{1,\ldots,d\}$. The matrix analogy of Θ was studied in graphon analysis [21]; see Figure 5b for an illustration. In this case, neither Θ nor $\mathcal{Z}$ is low-rank; in fact, Figure 5c shows that the rank is no smaller than the dimension d. Again, classical low-rank models fail to address this type of tensor structure.

These examples reveal the inadequacy of the conventional low-rank model in capturing important yet complex tensor structure. The observation has motivated us to develop a flexible class of nonparametric tensor models for estimating nonlinear, local, and possibly high-rank effects. Our goal is to accurately estimate Θ from the incomplete, noisy observation of $\mathcal{Y}$. In particular, we focus on the following two problems:

Open problem 4.
- How to estimate Θ under a wide range of structures, including both low-rankness and high-rankness?
- How many observed tensor entries do we need to consistently estimate the signal Θ?

3.2 Proposed Agenda: Nonparametric Sign Representable Tensor Models

We plan to develop new tools to address aforementioned challenges. In the earlier two examples, the high-rankness in the signal Θ makes the estimation challenging. The key crux is to design a new notion of nonparametric complexity that incorporates both low and high-rank tensors. Let us examine the sign of the π-shifted signal tensor $\text{sgn}(\Theta - \pi)$ for a given $\pi \in \mathbb{R}$. Here the operation $\text{sgn}(\cdot)$ converts a real-valued tensor

8

to an $\{-1,1\}$-valued tensor in an entrywise fashion. It turns out that, the sign tensors have the same sign patterns as low-rank tensors. Indeed, the signal tensor in the first example has the same sign pattern as a rank-4 tensor, since $\mathrm{sgn}(\Theta - \pi) = \mathrm{sgn}(\mathcal{Z} - f^{-1}(\pi))$. The signal tensor in the second example has the same sign pattern as a rank-2 tensor, since $\mathrm{sgn}(\Theta(i,j,k) - \pi) = \mathrm{sgn}(\max(i,j,k) - d(e^\pi - 1))$.

The above observation suggests a general framework to estimate both low- and high-rank signal tensors. Figure 6 illustrates the main idea of our method. Suppose we aim to estimate a signal tensor $\Theta \in [-1,1]^{d \times \cdots \times d}$ from a data tensor observation $\mathcal{Y} = \Theta + \mathcal{E}$ contaminated by a noise $\mathcal{E}$ with i.i.d. sub-Gaussian entries. We develop a nonparametric estimate $\hat{\Theta}$ by averaging over the series of de-noised sign tensors

$$\hat{\Theta} = \frac{1}{2H+1} \sum_{\pi \in \mathcal{H}} \mathrm{sgn}(\hat{\mathcal{Z}}_\pi), \quad \text{where } \hat{\mathcal{Z}}_\pi = \underset{\text{low rank tensor } \mathcal{Z}}{\arg\min} \; \text{Weighted-Loss}(\mathrm{sgn}(\mathcal{Z}), \mathrm{sgn}(\mathcal{Y} - \pi)). \quad (3)$$

Here $\mathrm{sgn}(\hat{\mathcal{Z}}_\pi)$ is the de-noised sign tensor estimated at a series of $\pi \in \mathcal{H} = \{-1, \ldots, -1/H, 0, 1/H, \ldots, 1\}$, $H \in \mathbb{N}_+$ is a pre-specified precision level, and Weighted-Loss$(\cdot, \cdot)$ is a classification objective that assesses the total sign mismatches, with entrywise weights, between two tensor arguments.
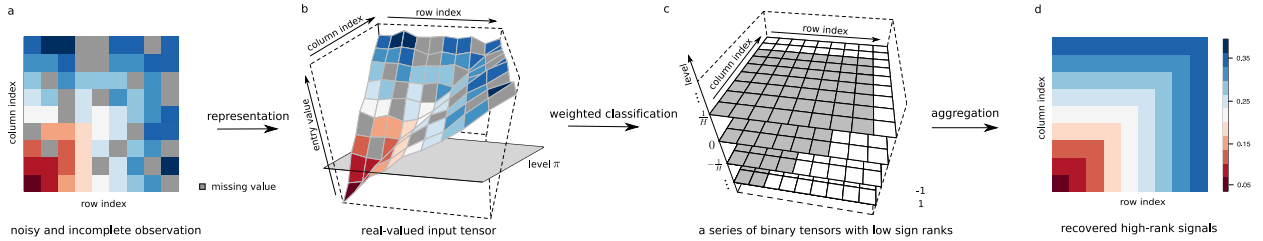

Figure 6: Illustration of our nonparametric methods based on based on Example 3.4 in Section 3.3. (a): a noisy, incomplete tensor input. (b)-(c): Estimation of sign tensor series $\mathrm{sgn}(\Theta - \pi)$ for $\pi \in \{-1, \ldots, -1/H, 0, 1/H, \ldots, 1\}$. (d): recovered signal $\hat{\Theta}$.

Our approach is built on the nonparametric representation of signal tensors. Surprisingly, a careful aggregation of dichotomized data not only preserves all information in the original signals, but also brings the accuracy and flexibility over classical low-rank models. The sign representation is guaranteed to recover both low- and high-rank signals. In addition, a total of $H = \mathrm{poly}(d)$ base algorithms suffice to recover Θ under the considered model. The method therefore enjoys both statistical and computational efficiency.

3.3 Preliminary Results

Our future work will formulate the above intuition into a general nonparametric tensor framework. We plan to investigate a rich nonparametric family based on what we coin as the "sign-rank" of a tensor.

Definition 3.1 (Sign representable tensors)**.** The sign-rank of a tensor Θ is defined by the minimal rank among all tensors that have the same sign pattern as Θ; i.e., $\mathrm{srank}(\Theta) = \min\{\mathrm{rank}(\Theta') \colon \mathrm{sgn}(\Theta') = \mathrm{sgn}(\Theta)\}$. We use $\mathscr{P}_{\mathrm{sgn}}(r) = \{\Theta \colon \max_{\pi \in [-1,1]} \mathrm{srank}(\Theta - \pi) \leq r\}$ to denote the r-sign representable tensor family.
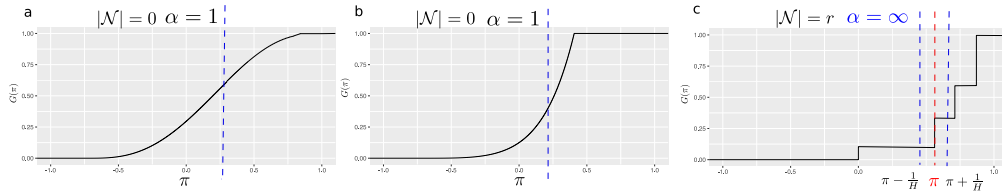

Figure 7: (a)-(b) Functions $G(\pi)$ with local Hölder smoothness index $\alpha = 1$; (c) $G(\pi)$ with $\alpha = \infty$ at most π (in blue), except for a total number of $|\mathcal{N}| = r$ jump points (in red). Here $|\mathcal{N}|$ denotes the number of jump points in G.

We characterize the statistical hardness of the estimation problem using $G(\pi) = \mathbb{P}_{\omega \in \Pi}(\Theta(\omega) \leq \pi)$, the cumulative distribution function (CDF) of tensor entries under sampling distribution Π over index $\omega \in [d]^K$.

See Figure 7 for three examples of $G(\pi)$ with various smoothness behaviors.

Theorem 3.1 (L<u>W</u>, on-going work). Suppose $\Theta \in \mathscr{P}_{\mathrm{sgn}}(r)$ is an α-smooth tensor with bounded number of jump points. Let $\Omega \subset [d]^K$ denote the index set of observed entries, and assume the entries ω in Ω are i.i.d. draws with replacement from the full index set using sampling distribution Π. The global optimizer in (3) based on our method (Figure 6) obeys the following accuracy guarantees.

(a) (Tensor denoising) For all precision levels $H \in \mathbb{N}_+$, with high probability,

$$
\mathbb{E}_{\omega \sim \Pi} |\hat{\Theta} - \Theta| \lesssim \underbrace{\left(\frac{dr}{|\Omega|} \right)^{\frac{\alpha}{\alpha+2}}}_{\text{Error inherited from sign estimation}} + \underbrace{\frac{1}{H}}_{\text{Bias}} + \underbrace{\frac{Hdr}{|\Omega|}}_{\text{Variance}} \lesssim \left(\frac{dr}{|\Omega|} \right)^{\min\left(\frac{\alpha}{\alpha+2}, \frac{1}{2} \right)},
$$

where the last inequality is obtained by setting $H \asymp |\Omega|/dr$.

(b) (Tensor completion). Suppose $\alpha \neq 0$. Then, with high probability, the minimal sample requirement for consistent tensor completion is $\tilde{\mathcal{O}}(dr)$, where $\tilde{\mathcal{O}}(\cdot)$ hides the logarithmic factors.

Our preliminary investigation shows that the r-sign representable tensor family is a general model that incorporates most existing tensor models including low-rank tensors, single index models, logistic tensor models, and structured tensors with repeating entries. Table 1 shows the outperformance of our estimation accuracy compared to previous methods.

Example 3.1 (Tensor block model). Tensor block model [3, 22] assumes a checkerboard structure among tensor entries under marginal index permutation. The signal tensor Θ takes at most r distinct values, where r is the total number of multiway blocks. Our model incorporates tensor block model in that $\Theta \in \mathscr{P}_{\mathrm{sgn}}(r)$.

Example 3.2 (Generalized linear model). Logistic tensor model [2] assumes a binary tensor $\mathcal{Y}$ with mean $\Theta = \mathrm{logit}(\mathcal{Z})$, where $\mathcal{Z}$ is a latent low-rank tensor. The mean tensor Θ itself if often high-rank (see Figure 5a); however we find that Θ has low complexity based on our sign-rank notion. Same results hold for general exponential-family models with a (known) link function [32].

Example 3.3 (Single index model). Single index model is a flexible semiparametric model proposed in economics [40] and high-dimensional statistics [18, 27]. The single index model assumes the existence of a (unknown) monotonic function $g\colon \mathbb{R} \to \mathbb{R}$ such that $g(\Theta)$ has rank r. We see that Θ belongs to the sign representable family; i.e., $\Theta \in \mathscr{P}_{\mathrm{sgn}}(r+1)$.

Example 3.4 (Structured tensors with repeating entries). Here we revisit the model introduced in Figure 5b. We find that $\Theta \in \mathscr{P}_{\mathrm{sgn}}(2)$, because the sign tensor $\mathrm{sgn}(\Theta - \pi)$ with an arbitrary $\pi \in (0, \log 2)$ is a block tensor with at most two blocks (see Figure 6c).

Model	Our rate (power of d)	Previous results
Tensor block model	$-(K-1)/2$	$\alpha = \infty$; agrees with the minimax rate [3].
Generalized linear model	$-(K-1)/3$	$\alpha = 1$; close to parametric rate [10].
Single index model	$-(K-1)/3$	$\alpha = 1$; conjecture to be optimal; our rate in matrix case $d^{-1/3}$ improves previous result $d^{-1/4}$ [27].
Nonparametric tensor model	$-(K-1)\min(\frac{\alpha}{\alpha+2} \wedge \frac{1}{2})$	faster rate as smoothness index α increases.

Table 1: Comparison of statistical convergence rate between our nonparametric tensor models and previous methods.

The current theoretical guarantees we obtain are for the global optimizer. The proposed agenda will focus on characterization of local optimizers, or relatedly, the computational-statistical trade-off for nonparametric tensor estimation. Our extensive experience on tensor related work has shown that tensors sought in applications often possess special structures. These properties provide solutions to address global optimality issues that are impossible in general cases. In a wide range of non-convex problems, all saddle points are strict

and all local minima are global minima. My group is actively exploring the interplay between optimization error and statistical accuracy in nonparametric tensor problems. The proposed approach will revolutionize the tensor problems including, but not limited to, tensor SVD, tensor completion, tensor decomposition, and tensor regression.

> **Open problem 5.**
> - What is the minimax optimal statistical rate for nonparametric tensor estimation?
> - Can we develop polynomial-time algorithms to achieve the optimal statistical rate?

3.4 Assessing Mechanisms

We plan to assess the success of our approach using two tensor datasets, the MRN-114 human brain connectivity data [44], and NIPS data [30]. The brain dataset records the structural connectivity among 68 brain regions for 114 individuals along with their Intelligence Quotient (IQ) scores. The NIPS dataset consists of word occurrence counts in papers published from 1987 to 2003. The resulting dataset is an order-3 tensor with entry representing the log counts of words by authors across years.

Table 2 and Figure 8 show the preliminary data analysis on the two data applications. Table 2 compares the prediction accuracy of different methods. Reported prediction errors are averaged over five runs of cross-validation, with 20% entries for testing and 80% for training. Our method substantially outperforms the low-rank CP method for every configuration under consideration. Figure 8a plots the top brain connections related to IQ scores are mostly inter-hemisphere edges, consistent with recent research on brain connectivity [44]. Figure 8b illustrates the results from NIPS data, where we plot the entries in $\hat{\Theta}$ corresponding to top authors and most frequent words. The identified pattern agrees with active topics in the NIPS publication. Among the top words are *neural* (marginal mean = 1.95), *learning* (1.48), and *network* (1.21), whereas top authors are *T. Sejnowski* (1.18), *B. Scholkopf* (1.17), *M. Jordan* (1.11), and *G. Hinton* (1.06). The detected pattern and achieved accuracy demonstrate the applicability of our method.

MRN-114 brain connectivity dataset			
Method	$r = 6$	$r = 9$	$r = 12$
Nonparametric (Ours)	**0.14**(0.001)	**0.12**(0.001)	**0.12**(0.001)
CPT	0.23(0.006)	0.22(0.004)	0.21(0.006)
NIPS word occurrence dataset			
Method	$r = 6$	$r = 9$	$r = 12$
Nonparametric (Ours)	**0.16**(0.002)	**0.15**(0.001)	**0.14**(0.001)
CPT	0.20(0.007)	0.19(0.007)	0.17(0.007)

Table 2: Prediction comparison between our method ($H = 20$) and low-rank CP method (CPT) in the two data applications. Standard errors are in parenthesis.
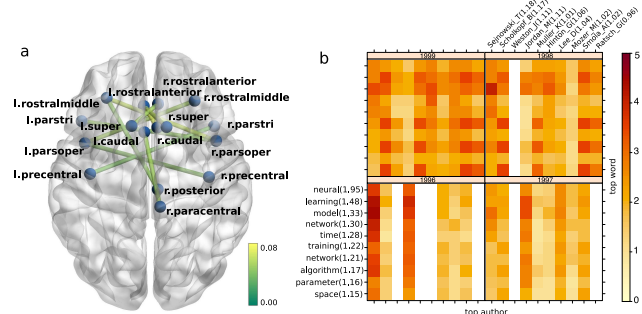

Figure 8: (a) top IQ-associated edges in the brain connectivity data. (b) top (authors, words, year) triplets in the NIPS data.

> **Significance and transformative concepts:**
> - Develop efficient nonparametric prediction tools for tensor classification, tensor regression, and deep tensor neural network, in the presence of computational constraints.
> - Improve efficiency in scientific discovery using powerful multimodal brain data connectivity analyses.
> - Build data-driven prototypes that integrate tensor data learning into the data-to-decision process.

4 AIM 3: Predictive Tensor Models with Auxiliary Information

Tensor datasets are often collected together with additional auxiliary features in modern scientific and engineering studies. In neuroimaging analysis, brain connectivity networks are collected from a sample of

individuals, accompanied by individual characteristics such as age, gender, and diseases status (see Figure 9a). In social network analysis, connectivity strengths between individual pairs are collected, with auxiliary information such as people's demographic information and friendship types available. In both examples, scientists are interested in learning the relationship between tensor data and auxiliary features. These seemingly different scenarios pose a common yet challenging problem for predictive tensor data modeling.
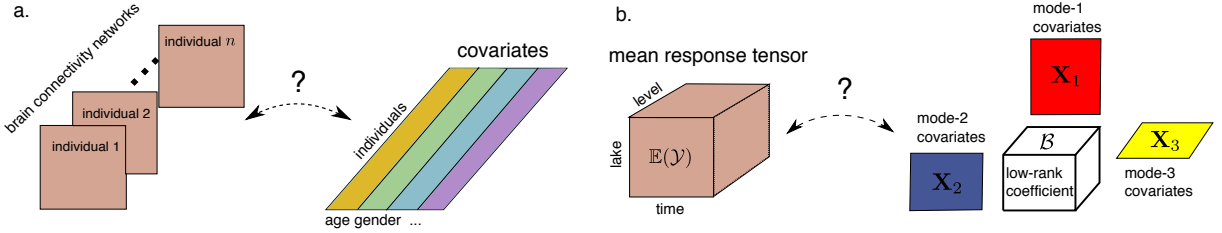

Figure 9: Examples of predictive tensor models with auxiliary information. (a) Network regression with auxiliary information. (b) Spatial-temporal growth data represented as an order-3 tensor $\mathcal{Y}$, where feature matrices $\boldsymbol{X}_k$ are on each of the three modes.

4.1 Three Motivating Examples

We give three concrete examples to highlight the applicability of predictive tensor models.

Example 4.1 (Dyadic data with node attributes). Dyadic dataset consists of measurements on pairs of objects. Common examples include graphs and networks. Let $\mathcal{G} = (V, E)$ denote a graph, where $V = [d]$ is the node set of the graph, and $E \subset V \times V$ is the edge set. Suppose that we also observe feature $\boldsymbol{x}_i \in \mathbb{R}^p$ associated to each node $i \in V$. A probabilistic model on the graph $\mathcal{G} = (V, E)$ can be described by matrix regression. The edge connects the two vertices i and j independently of other pairs, and the probability of connection is modeled as

$$\mathbb{P}((i, j) \in E | \boldsymbol{x}_i, \boldsymbol{x}_j) = \text{logistic}(\langle \boldsymbol{B}, \ \boldsymbol{x}_i^T \boldsymbol{x}_j \rangle),$$

where $\boldsymbol{B} \in \mathbb{R}^{p \times p}$ is an unknown parameter matrix of interest. The dyadic data is a predictive matrix model that uses matrix coefficient $\boldsymbol{B}$ for structure learning and edge prediction.

Example 4.2 (Network regression model). Network response model [47] is recently developed for neuroimanig analysis. The goal is to study the relationship between brain network connectivity patterns and individual's features. Suppose we have a sample of n observations, $\{(\boldsymbol{Y}_i, \boldsymbol{x}_i): i = 1, \ldots, n\}$, where for each individual $i \in [n]$, $\boldsymbol{Y}_i \in \{0, 1\}^{d \times d}$ is the symmetric adjacency matrix whose entries indicate presences/absences of connectivities between d brain nodes, and $\boldsymbol{x}_i \in \mathbb{R}^p$ is the individual's feature such as age, gender, cognition score, etc. The network regression models takes the form

$$\mathbb{E}(\boldsymbol{Y}_i | \boldsymbol{x}_i) = \text{logistic}(\mathcal{B} \times_3 \boldsymbol{x}_i), \quad \text{or more generally} \quad \mathbb{E}(\boldsymbol{Y}_i | \boldsymbol{x}_i) = f(\mathcal{B} \times_3 \boldsymbol{x}_i),$$

where $\mathcal{B} \in \mathbb{R}^{d \times d \times n}$ is the unknown parameter tensor of interest, and $f: \mathbb{R} \to \mathbb{R}$ is a possibly unknown function. The network regression (Figure 9b) is a predictive tensor model that uses tensor coefficient $\mathcal{B}$ for structure learning and connectivity prediction.

Example 4.3 (Spatio-temporal growth model). The growth curve model [25] was originally proposed as a bilinear model for matrix data, and we adopt its higher-order extension here. Let $\mathcal{Y} = [\![y_{ijk}]\!] \in \mathbb{R}^{d \times m \times n}$ denote the pH measurements of d lakes at m levels of depth and for n time points. Assume that the expected pH trend is polynomial in depth and in time. Then, the multilinear growth curve model is represented as

$$\mathbb{E}(\mathcal{Y} | \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3) = \mathcal{B} \times_1 \boldsymbol{X}_1 \times_2 \boldsymbol{X}_2 \times_3 \boldsymbol{X}_3,$$

where $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3$ are the design matrices for lake types, spatial effects, and temporal effects, respectively, $\mathcal{B} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ is the unknown parameter tensor of interest. The spatial-temporal model (Figure 9a) belongs to the predictive tensor model, with the tensor $\mathcal{B}$ serving the role of joint structure learning and response prediction.

Significance and transformative concepts:
- Develop powerful predictive tensor models for joint structure estimation and function learning.
- Address tensor data mining challenges with interpretable inference and uncertainty quantification.
- Foster interdisciplinary research that integrates methodological developments and domain sciences.

4.2 Proposed Agenda: Tensor Neural Network Models

The proposed project focuses on developing a general framework for tensor learning and prediction with auxiliary information. The aforementioned examples represent three different forms of predictive tensor regression models, i.e., whether tensor is treated as predictors (Example 4.1), as responses (Example 4.2), or both (Example 4.3). Here we outline the research direction for the first setting; the generalization to the other two are similar.

We propose to formulate the predictive task as learning the regression function between tensor predictor $\mathcal{X} \in \mathbb{R}^{d_1 \cdots \times d_K}$ and the scalar response $Y \in \mathbb{R}$. Based on the i.i.d. training set $(\mathcal{X}_i, Y_i)_{i=1,\ldots,n}$, we aim to estimate the conditional mean of Y as a function of $\mathcal{X}$ based on the tensor neural network model

$$Y|\mathcal{X} = f(\mathcal{X}) + \varepsilon, \quad \text{where} \quad f(\mathcal{X}) = \sum_{h=1}^{H} v_h \sigma(\langle \mathcal{B}_h, \mathcal{X} \rangle - \pi_h), \tag{4}$$

where ε is a zero-mean sub-Gaussian noise, $\sigma \colon \mathbb{R} \to \mathbb{R}$ is a pre-specified activation function, $\mathcal{X} \mapsto \langle \mathcal{B}_h, \mathcal{X} \rangle$ is the trace effect with an unknown parameter tensor $\mathcal{B}_h \in \mathbb{R}^{d_1 \times \cdots \times d_K}$, $v_h, \pi_h \in \mathbb{R}$ are unknown scalar parameters, and $h = 1, \ldots, H$ is the index in function basis representation. The activation function is often nonlinear so that f is also nonlinear in parameter $\mathcal{B}_h$. Figure 9 illustrates our model with activation function $\sigma(z) = \mathrm{sgn}(z)$. Other common activation functions are Rectified Linear Unit (ReLU) $\sigma(z) = z_+$ and logistic $\sigma(z) = (1 + \exp(-z))^{-1}$. The model (4) extends the multivariate two-layer neural network to tensor data, so we refer it to as "tensor neural network" (see Figure 9).
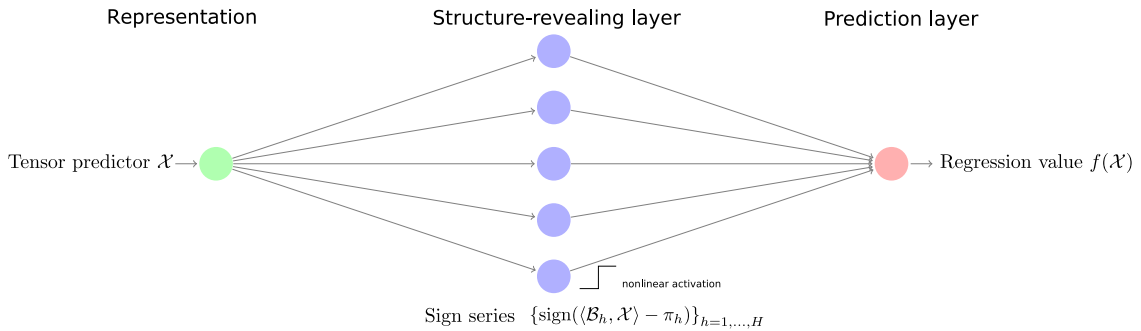

Figure 10: Proposed tensor neural network model for predictive tensor learning tasks.

Our proposed tensor neutral network (Figure 9) blends the power of neural network and structure learning in order to take the best out of both worlds. Learning parameter tensors $\mathcal{B}_h$ in (4) is challenging because of the extremely high dimensionality in the tensor space and functional space. The PI's previous extensive experience on tensor related work has shown that tensors sought in applications often possess special structures, such as (nearly) low-rankness, sparsity, non-negativity, and smoothness. We will leverage the formalisms of *intrinsic dimension* to impose suitable structure models on $\mathcal{B}_h$. The structure-revealing model

13

effectively mitigates the curse of high dimensionality. We will further develop multi-layer tensor neural network that incorporates practical constraints in domain applications, such as incomplete observation with missing labels, corrupted distributions, and computation with time and memory constraints.

> **Breaking previous limits:**
> • Do higher-order tensors require a completely new modeling paradigm compared to traditional data?
> • How to develop robust, adaptive estimates that achieve both statical and computational efficiency?
> Solving these questions will revolutionize the tensor learning research by bridging theory and practice.

5 Integration of Education and Research

The project will create deliverables in three aspects: **research**, **education**, and **social good**. Figure 11 summarizes the planned milestones and deliverables; details will be described in the next paragraph. My plan consists of a range of activities aimed at increasing understanding, awareness, and participation in data science, integrated with the research and teaching agenda in this proposal.

	2022	2023	2024	2025	2026
Research	aim 1. parametric models		aim 2. nonparametric models		aim 3. tensor data integration
		non-convex optimization with accuracy gaurantees			neural network algorithm
		application to imaging	application to social network sciences		application to genomics
Education		develop new course on data sciences			create online courses
	online tutorial	diversity and inclusion in education			students' career development
		host undergraduates for summer research			presentation at IFDS, Data Hub
Social Impact		build system for multimodal data analysis			build data-driven prototypes
		develop cloud-based collaboration tools			release free software
		disseminate research outcomes in academic conferences and public lectures			

Figure 11: Planned deliverables and milestones in my proposal.

5.1 Education and Diversity

I am one of the few female faculty members in my home department. As a young academic, I have been committed to promoting diversity, equity, and inclusion throughout all facets of my career – in teaching, education, and research. I am currently starting my fourth year in the tenure track, and I have successfully mentored a number of female undergraduate students – two of which are embarking on their PhD studies at top US universities this fall. Female students are traditionally underrepresented in my filed, and I have been striving to motivate more women in my research. My experience of recruiting underrepresented students has attracted social media coverage [16]. With the support of CAREER research grant, I will extend this effort to a larger student body by creating unique research-training experience to more diverse undergraduates.

I plan to create a new course on *fairness in data science* in both traditional and online formats. This course will introduce students to address bias and fairness when deploying data science system, and will be closely tied with social science and scientific domains. The pedagogical tools will draw on my rich teaching experience and recent success as the Madison Teaching and Learning Excellence (MTLE) Fellow [17]. I will apply active learning techniques in my class. Statistics, as a discipline, transforms "data" into "information". However data are not just numbers; they are numbers with social, economical, and cultural contexts. My new course will promote critical thinking skills and develop equitable access for underrepresented groups. The personnel constitution in our education environment is becoming increasingly diverse, reflecting the variety witnessed in our broader society. Diversity can be measured across many dimensions – gender, ethnicity, origin, age, etc. Nevertheless, diversity is not limited to these visible physical differences; diversity

also consists of difference in each student's personality (introvert vs. extravert), learning skills, social status, emotional development, and every aspect that distinguishes one individual from another. I believe that every student is unique in their own way, and I will create an open, inclusive education environment that prepares students to become better professionals in the larger, dynamic society.

5.2 Nurturing the Next Generation of Statisticians

The educational plan to establish a unique research-training experience and broaden participation in STEM will be well integrated with the research agenda. The outreach activities are targeted at several different populations: PhD students, undergraduate students, minority students, visiting international students, and AP students at Madison West High School. I have and will continue to leverage my professional affiliations in Women in Probability, Women in Machine Learning, and Institute for Foundations of Data Science (IFDS), to engage a diverse pool of students into STEM fields.

I will establish my group as a unique center for tightly integrated training in two key components: disciplinary expertise and collaborative skills. The multi-disciplinary nature of my publication record, the current spread of projects in my group, and the details of the proposed research make this goal very realistic. Beyond research goals, my mentoring will provide skills, knowledge, and experience to prepare every student to achieve their highest potential. Specific agenda includes:

- Interaction with the PI:
  - The orientation meeting will involve in-depth conversations between the PI and the students to establish the common goals. Key topics include the career development, amount of independence the students require, work habits, research productivity expectation, and data/code ethnicity.
  - Have regular one-on-one meeting with the students to review progress, provide feedback, exchange new techniques, help with publication and fellowship application.
- Individual career guidance:
  - Provide supports to students on personal branding techniques including resume writing, interview preparation, faculty-job searching, and career path outside the academia.
  - Train the students into the new generation of statisticians that have strong skills in quantitative reasoning and problem solving to address large-scale data problems now readily available.
- Outreach activity:
  - Partner with the Data Science Hub and the Division of Diversity, Equality & Education Achievement (DDEEA) program at UW-Madison to broaden participation in STEM.
  - Provide students the collaborative skills to make independent, insightful contributions to the society.
  - Sponsor students for at least two conferences or meetings each year (travel cost included in the budget), with the goal to build up students' visibility and emotional ownership for their career development.

5.3 Broader Impacts

The mission of this proposal is to foster an interdisciplinary research team in a way that benefits fellow scientists, academia, and larger society. The proposed research will engage a diverse pool of underrepresented students into STEM, and establish an inclusive environment where every student can success in their own future careers. The PI will organize a series of tutorials on tensor data analytics, with the theme to encourage interdisciplinary collaborations. The PI's ongoing collaborations include research groups at University of Chicago, University of California-Berkeley, and Chan Zuckerberg Biohub. Further collaboration will continue through the NSF-funded TRIPODS (+X) Institute for which the PI is affiliated faculty. Research outputs will be disseminated through publications and software packages with detailed documentation. The PI's group have released over 15 open-source packages in R, Matlab, and C++ at Github thus far. These software packages will provide reproducible research that benefits the larger society.

References

[1] **Miaoyan Wang**, Jonathan Fischer, Yun S Song, Three-way clustering of multi-tissue multi-individual gene expression data using semi-nonnegative tensor decomposition, The Annals of Applied Statistics 13 (2019), No. 2, 1103-1127.

[2] **Miaoyan Wang** and Lexin Li, Learning from binary multiway data: Probabilistic tensor decomposition and its statistical optimality, Journal of Machine Learning Research 154 (2021), 1-38.

[3] **Miaoyan Wang** and Yuchen Zeng, Multiway clustering via tensor block models, Advances in Neural Information Processing Systems 33 (2019), 713-723.

[4] **Miaoyan Wang**, Fabrice Roux, Claudia Bartoli, Carine Huard-Chauveau, Christopher Meyer, Hana Lee, Dominique Roby, Mary Sara McPeek, and Joy Bergelson. Two-way mixed-effects methods for joint association analyses using both host and pathogen genomes, Proceedings of the National Academy of Sciences (direct submission) 115 (2018) No. 24, E5440-E5449.
Top 5% attention score among research articles tracked by Altmetric; Top 11% among all papers published in PNAS. Platform talk (top 13%) at the 2nd Probabilistic Modeling in Genomics, Cold Spring Harbor Laboratory,

[5] **Miaoyan Wang** and Yun S. Song. Tensor decomposition via two-mode higher-order SVD (HOSVD), Journal of Machine Learning Research W&CP 54 (AISTATS track 2017), 614-622.

[6] **Miaoyan Wang**, Khanh Dao Duc, Jonathan Fischer, and Yun S. Song. Operator Norm inequalities between tensor unfoldings on the partition lattice, Linear Algebra and its Applications 520 (2017), 44-66.

[7] **Miaoyan Wang**, Johanna Jakobsdottir, Albert V Smith, Mary Sara McPeek. G-STRATEGY: Optimal selection of individuals for sequencing in genetic association studies, Editor's Pick in Genetic Epidemiology 40 (2016), No. 6, 446-60.
2014 Charles J. Epstein Trainee Award semifinalist (27 predoctoral recipients out of ∼550 candidates) from American Society of Human Genetics (ASHG); Platform talk (top 8%) in 2014 ASHG meeting.
2013 IGES Williams Award – finalist (3 predoctoral recipients out of 156 candidates) from International Genetic Epidemiology Society (IGES); Platform talk in 2013 IGES meeting.

[8] Jiaxin Hu, Chanwoo Lee and **Miaoyan Wang**. Supervised Tensor Decomposition with interactive side information (2021). arXiv:1910.09499.
2021 Best Student Paper Award from the Statistical Computing and Graphics Section of American Statistical Association (ASA).

[9] Rungang Han, Yuetian Luo, **Miaoyan Wang**, and Anru R Zhang, Exact clustering in tensor block model: Statistical optimality and computational limit (2021). arXiv:2012.09996.
2021 Best Student Paper Award from the Statistical Learning and Data Science Section of the American Statistical Association (ASA).

[10] Chanwoo Lee and **Miaoyan Wang**, Tensor denoising and completion based on ordinal observations, International Conference on Machine Learning (2020), 5778-5788.

[11] Jiaxin Hu, Chanwoo Lee, and **Miaoyan Wang**. Learning multiple networks via supervised tensor decomposition, Neural Information Processing Systems 34 (2020), Workshop on Machine Learning and the Physical Sciences.

[12] Duo Jiang and **Miaoyan Wang**. Recent developments in statistical methods for GWAS and high-throughput sequencing studies of complex traits, Biostatistics and Epidemiology 2 (2018), No. 1, 132-159.

[13] Chanwoo Lee and **Miaoyan Wang**. Beyond the signs: nonparametric tensor completion via sign series, Under review by Neural Information Processing Systems 35 (2021). arXiv:2102.00384.

[14] Chanwoo Lee, Lexin Li, Helen Zhang, and **Miaoyan Wang**. Nonparametric trace regression in high dimensions via sign series representation, Under review by Annals of Statistics (2021).

[15] B. W. Engelmann, Y. Kim, **Miaoyan Wang**, B. Peters, R. S. Rock, and P. D. Nash. The development and application of a quantitative peptide microarray platform to protein interaction domain specificity space, Molecular and Cellular Proteomics 13 (2014), No. 12, 3647-62.

[16] **Miaoyan Wang** and Julie Warner, Women in STEM: 5 Thoughtful Ways to Recruit and Retain Them, Course Hero faculty club feature article (2020). https://www.coursehero.com/faculty-club/classroom-tips/miaoyan-wang/.

[17] **Miaoyan Wang** and et al. UW-Madison Teaching and Learning Excellence Fellow (2020). https://mtle.wisc.edu/about-us/meet-the-fellows/omicron-cohort/.

[18] Fadoua Balabdaoui, Cécile Durot, and Hanna Jankowski, Least squares estimation in the monotone single index model, Bernoulli 25 (2019), No. 4B, 3276-3310.

[19] Mikhail Belkin, Daniel Hsu, Siyuan Ma, and Soumik Mandal, Reconciling modern machine-learning practice and the classical bias–variance trade-off, Proceedings of the National Academy of Sciences 116 (2019), No. 32, 15849-15854.

[20] Marie-Aurélie Bruno, D Fernández-Espejo, Remy Lehembre, Luaba Tshibanda, Audrey Vanhaudenhuyse, Olivia Gosseries, Emilie Lommers, Martino Napolitani, Quentin Noirhomme, Mélanie Boly, et al., Multimodal neuroimaging in patients with disorders of consciousness showing functional hemispherectomy, Progress in Brain Research 193 (2011), 323-333.

[21] Stanley Chan and Edoardo Airoldi, A consistent histogram estimator for exchangeable graph models, International Conference on Machine Learning (2014), 208-216.

[22] Eric C Chi, Brian J Gaines, Will Wei Sun, Hua Zhou, and Jian Yang, Provable convex co-clustering of tensors, Journal of Machine Learning Research 21 (2020), No. 214, 1-58.

[23] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle, A multilinear singular value decomposition, SIAM Journal on Matrix Analysis and Applications 21 (2000), No. 4, 1253-1278.

[24] Joshua Faskowitz, Xiaoran Yan, Xi-Nian Zuo, and Olaf Sporns, Weighted stochastic block models of the human connectome across the life span, Scientific Reports 8 (2018), No. 1, 1-16.

[25] Gabriel, K. Ruben. Generalized bilinear regression, Biometrika 85 (1998), No. 3, 689-700.

[26] Silvia Gandy, Benjamin Recht, and Isao Yamada, Tensor completion and low-rank tensor recovery via convex optimization, Inverse Problems 27 (2011), No. 2, 025010.

[27] Ravi Sastry Ganti, Laura Balzano, and Rebecca Willett, Matrix completion under monotonic single index models, Advances in Neural Information Processing Systems (2015), 1873-1881.

[28] Rong Ge and Tengyu Ma, On the optimization landscape of tensor decompositions, Mathematical Programming (2020), 1-47.

[29] Navid Ghadermarzy, Yaniv Plan, and Ozgur Yilmaz, Learning tensors from partial binary measurements, IEEE Transactions on Signal Processing 67 (2018), No. 1, 29-40.

[30] Amir Globerson, Gal Chechik, Fernando Pereira, and Naftali Tishby, Euclidean embedding of co-occurrence data, Journal of Machine Learning Research 8 (2007), 2265-2295.

[31] Rungang Han, Rebecca Willett, and Anru Zhang, An optimal statistical and computational framework for generalized tensor estimation, Annals of Statistics, In press (2020).

[32] David Hong, Tamara G Kolda, and Jed A Duersch, Generalized canonical polyadic tensor decomposition, SIAM Review 62 (2020), No. 1, 133-163.

[33] Tamara G Kolda and Brett W Bader, Tensor decompositions and applications, SIAM review 51 (2009), No. 3, 455-500.

[34] Jing Lei, Kehui Chen, and Kevin Lynch, Consistent community detection in multilayer network data, Biometrika, to appear (2019).

[35] Marta Melé, Pedro G Ferreira, Ferran Reverter, David S DeLuca, Jean Monlong, Michael Sammeth, Taylor R Young, Jakob M Goldmann, Dmitri D Pervouchine, Timothy J Sullivan, The human transcriptome across tissues and individuals, Science 348 (2015), No. 6235, 660-665.

[36] Zongming Ma and Yihong Wu, Computational barriers in minimax submatrix detection, The Annals of Statistics 43 (2015), No. 3, 1089-1116.

[37] W James Murdoch, Chandan Singh, Karl Kumbier, Reza Abbasi-Asl, and Bin Yu, Definitions, methods, and applications in interpretable machine learning, Proceedings of the National Academy of Sciences 116 (2019), No. 44, 22071-22080.

[38] Ashwin Pananjady and Richard J Samworth, Isotonic regression with unknown permutations: Statistics, computation, and adaptation (2020), arXiv:2009.02609.

[39] Emile Richard and Andrea Montanari, A statistical model for tensor PCA, Advances in Neural Information Processing Systems (2014), 2897-2905.

[40] Peter M Robinson, Root-n-consistent semiparametric regression, Econometrica: Journal of the Econometric Society (1988), 931-954.

[41] Will Wei Sun, Junwei Lu, Han Liu, and Guang Cheng, Provable sparse tensor decomposition, Journal of Royal Statistical Association Society, Series B 79 (2017), No. 3, 899-916.

[42] Nicholas J Timpson, Celia MT Greenwood, Nicole Soranzo, Daniel J Lawson, and J Brent Richards, Genetic architecture: the shape of the genetic contribution to human traits and disease, Nature Reviews Genetics 19 (2018), No. 2, 110.

[43] Xiang Wang, Chenwei Wu, Jason D Lee, Tengyu Ma, and Rong Ge, Beyond lazy training for overparameterized tensor decomposition, Advances in Neural Information Processing systems (2020) 21934-21944.

[44] Lu Wang, Daniele Durante, Rex E Jung, and David B Dunson, Bayesian network-response regression, Bioinformatics 33 (2017), No. 12, 1859-1866.

[45] Jean-Gabriel Young, Guillaume St-Onge, Patrick Desrosiers, and Louis J Dubé, Universality of the stochastic block model, Physical Review E 98 (2018), No. 3, 032309.

[46] Hua Zhou, Lexin Li, and Hongtu Zhu, Tensor regression with applications in neuroimaging data analysis, Journal of the American Statistical Association 108 (2013), No. 502, 540-552.

[47] Jingfei Zhang, Will Wei Sun, and Lexin Li, Network response regression for modeling population of networks with covariates (2018), arXiv:1810.03192 .

[48] Anru Zhang and Dong Xia, Tensor SVD: Statistical and computational limits, IEEE Transactions on Information Theory 64 (2018), No. 11, 7311-7338.