1. Hi, this is Chanwoo Lee, a phd student in UW Madison. This paper is a joint work with my advisor Miaoyan Wang. I am happy to present our work in NeurIPS 2021. Today, I am going to talk about nonparametric tensor completion method via sign series.

2. Tensor-valued data arises in many applications. The first figure shows the human brain connectivity dataset which shows pairwise connections among 68 brain nodes for 114 individuals. Stacking adjacency matrix for each individual gives us 3-order tensor dataset.

   A figure in the middle illustrates NIPS data that consists of word occurrence counts across authors, words, and years from 1987 to 2003 published in NIPS. Therefore, The resulting dataset is an order-3 tensor with entry representing the counts of words by authors across years.

   The last example is a RGB image dataset which consists of pixel values across three channels: height, width and color. As you can see from these examples, tensor datasets appear in many contexts and it is important to develop great tools to analyze tensor data.

3. Let me introduce main problem we are going to solve today. We consider the signal plus noise model where the observed data tensor is from signal and noise tensors with possibly missing entries. We focus on the two problems. First signal tensor estimation: How we can estimate the signal tensor from noisy observation. Second, complexity of tensor completion: How many observed tensor entries do we need to successfully estimate the signal tensor.

4. Low rank models are one of the most popular methods to solve these problems. This model assumes that the signal tensor is represented by sum of rank one tensors. While these methods have shown great success in theory, tensors in applications often violate the low-rankness. explain the figure

5. First it's sensitive to order preserving transformation. Let's assume that the signal tensor is from logistic function of rank-3 latent tensor. This figure shows that numerical rank of the signal tensor is an increasing function of transformation level c. So rank can be very high depending on transformation level.

   The second example demonstrates the inadequacy of classical low-rankness in representing special structures. Let's assume that the latent tensor has a special structure like this equation and signal tensor is obtained from log transformation. In this case neither signal tensor nor latent tensor is low-rank. The right figure shows that rank of both tensors are no smaller than its dimension. explain the figure. The figure plots the top d singular values for a d-dimensiona tensor. … We find all singular values are all above the zero.

   In the above and many other examples, the signal tensors of interest have high rank and low rank approaches miss these important structure. How do we overcome this?

6. We use sign concept to overcome the limit of low rank model. Here is an intuition for our method. For a bounded signal tensor, we can always approximate the signal tensor by the average of sign tensor series with given level set as in the equation.

   Since sign tensors are invariant to order preserving transformation, more flexible signal tensors are allowed compared to low rank model. Therefore, we shift our goal to estimate sign $(\Theta - \pi)$ out of the sign$(y - \pi)$ in the noisy case.

7. The key idea of the estimation is that we use a local notion of low-rankness to have richer family of signal tensors. Let me introduce several definitions for better understanding our method. Two tensors are called sign equivalent if sign of two tensors are the same. Sign rank of tensor is defined by the minimum rank of tensor among all sign equivalent tensors. For example, when signal tensor has special structure as in the figure, then, sign of the tensor has this block structure. So, sign rank of tensor is 2 while the original signal tensor has full rank.

8. Now we introduce sign representable tensors. We call tensor r sign representable if the tensor Theta - pi is sign rank bounded by r for all pi within the interval from -1 to 1. We verify that many existing structured tensor belong to ours sign representable family including low-rank tensors, high rank tensors from glm ,single index model, and tensors with repeating patterns. Therefore, we propose to use the sign represetable family instead of the classical low rank assumption.

9. Now we are ready! Let's go back to our main problem. Our goal is to estimate possibly high rank signals from noisy and incomplete observation. How do we do that?

10. We have three major steps for the recovery. Representation, weighted classification, and aggregation. I will explain each step in a next few slides.

11. The first step is representation. We observed a noisy and incomplete tensor Y with observed index set Omega. In this step, we dichotomize the data Y into sign$(Y - \pi)$ for given π and obtain a series of sign tensors based on a series of level πs.

12. The second step is weighted classification. In this step, we estimate $\text{sign}(\Theta - \pi)$ from $\text{sign}(y - \pi)$ based on weighted classification. The weighted classification consists of weight and 0-1 loss. This weight penalizes more to misclassification when the magnitude of $y - \pi$ is large.

13. We show that if signal tensor is r-sign representable and α-smooth, sign $(\Theta - \pi)$ is an unique optimizer of the risk of weighted classification among all r sign rank tensors. I will introduce alpha smoothness in the next slide. Therefore, we obtained a series of optimizers that minimize empirical risk of weighted classification loss at each level.

14. alpha smoothness measures the difficulty of the classification problem based on CDF of tensor Theta. First we partition the interval into classification hard and easy regions according to whether pseudo density is uniformly bounded or not. Then, we define alpha smoothness if the CDF function is well controlled by polynomial α-degree function.

    The first figure shows CDF where the set of hard classification region is empty and alpha is one. The second figure illustrates CDF function where the number of hard classification points is r while alpha is infinity.

    Based on this smoothness assumption, we are able to find unique optimizer of weighted classification.

15. The last step is aggregation. We obtain the signal tensor estimate by averaging a series of optimizers from weighted classification.

16. Now we provide theoretical guarantees for our estimation. Let's define mean absolute error of two tensors by expected distance between two tensor entries. Suppose that signal tensor is r sign representable and alpha smooth with equal dimension, then mean absolute error of estimated sign tensor is bounded by polynomial decays.

    explain the figure. The left figure is ... the right figure is ...
    Do not ask your audience to solve the puzzle by themselves. Put yourself on audience's shoes.
    Use this principle to check other parts of the talk.

    For tensor estimation, the estimation error consists of three source of errors: Error inherited from sign estimation, bias from sign series representation, and variance. With optimal choice of resolution parameter H, the estimation error again has the polynomial decays.

    We find that tensor estimation is generally no better than sign tensor estimation. This makes sense because estimating sign is usually easier than estimating the magnitude.

    We also constructed more general case such as unbounded hard classification region and sub-Gaussian case.

17. We compare our theoretical results with other papers. We achieve the minimax rate provided in wang and zeng's paper for tensor block model. Our results improve from d to the power minus a fourth by ganti et al to minus a third for single index model. For Generalized linear model, our error rate is close to parametric rate but not better. This is expected because we did not use the link function information. We constructed the our own error rate for smooth r sign representbale tensor as follows.

18. We show the performance of our method based on real data application. First, we analyze the brain connectivity data which consists of 68 brain regions for 114 individuals with their IQ scores. Data tensor is 68 by 68 by 114 binary tensor.

    The figure shows the top 10 brain edges based on regression analysis of denoised tensor from our method against normalized IQ scores. We find that the top connections are mostly inter-hemisphere edges, which is consistent with recent research on brain connectivity

19. Next we examine NIPS dataset which consists of word occurence counts in paper published from 1987 to 2003. We pick top 100 active authors and top 200 frequent words so the data tensor is 100 by 200 by 17 array.

    We examine the estimated signal and found that the most frequent words are consistent with the active topics and there is strong hetrogenity among word occurrences across authors and years. Right figure shows the similar word patterns of two authors whose name are scholkoft and smolar. I found that these two authors are actually co authors for many papers. The detected pattern demonstrate the applicability of our method.

20. Last, we performed cross validation on both brain and nips dataset to check the prediction accuracy and compare with low rank method. As we can see from the table, our method outperforms the low-rank method in real-world datasets.

21. Let me summarize. We developed a completion method that addresses both low and high rankness based on sign series representation. We established estimation error rate and sample complexities. Finally, we show applicability and prediction performance of our method based on both simulations and data applications. Here is an available R-package for our method. Thank you for listening and feel free to contact us if you have any comments.