Beyond the Signs: Nonparametric Tensor Completion via Sign Series

Chanwoo Lee[1] and Miaoyan Wang[2]

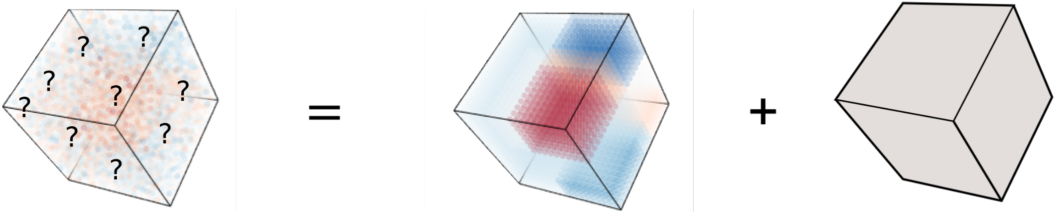
Department of Statistics, University of Wisconsin - Madison

NeurIPS 2021

[1]chanwoo.lee@wisc.edu [2]miaoyan.wang@wisc.edu

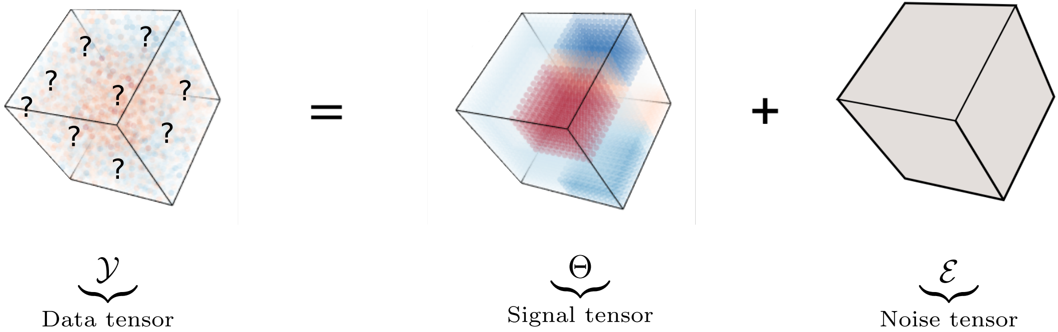
Preprint Software

Main problems: the signal plus noise model



$$\underbrace{\mathcal{Y}}_{\text{Data tensor}} = \underbrace{\Theta}_{\text{Signal tensor}} + \underbrace{\mathcal{E}}_{\text{Noise tensor}}$$

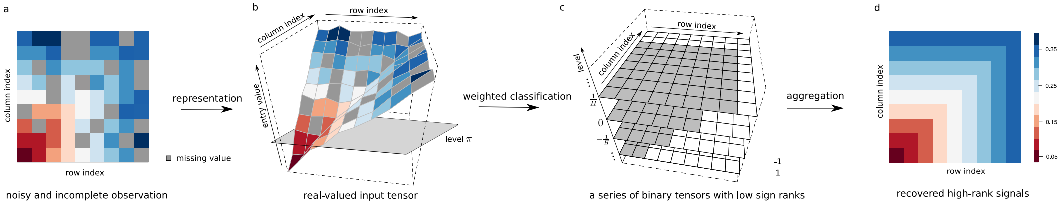
Main problems: the signal plus noise model



We focus on the two problems

1. Signal tensor estimation: How to estimate the signal tensor Θ?
2. Complexity of tensor completion: How many observed tensor entries do we need?

Our contribution

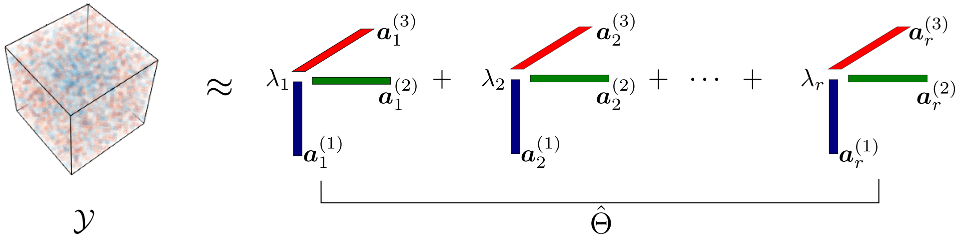


Special case with full observation:

Model	Our rate* (power of d)	Previous results
Tensor block model	$-(K-1)/2$	$\alpha=\infty$; minimax rate in Wang and Zeng (2019)
Single index model	$-(K-1)/3$	$\alpha=1$; conjecture on the optimality; matrix rate $d^{-1/3}$ improves $\mathcal{O}(d^{-1/4})$ by Ganti et al. (2017)
Generalized linear model	$-(K-1)/3$	$\alpha=1$; close to parametric rate in Lee and Wang (2020)
α-smooth $\mathscr{P}_{\text{sgn}}(r)$	$-(K-1)\min(\frac{\alpha}{\alpha+2}\wedge\frac{1}{2})$	faster rate as α increases; extended matrix case in Lee et al. (2021)

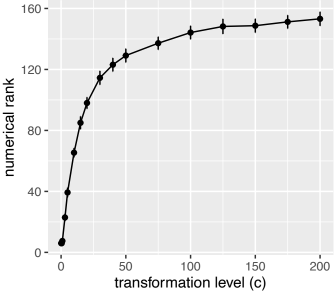
Inadequacies of low-rank models

- Low-rank models (Anandkumar et al., 2014; Montanari and Sun, 2018; Cai et al., 2019).



Inadequacies of low-rank models

- Sensitivity to order-preserving transformation

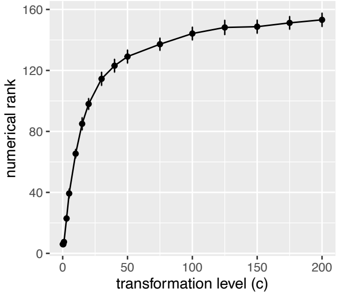


$$\Theta = \frac{1}{1 + \exp(-c\,(\mathcal{Z}))}, \quad \text{where}$$

$$\mathcal{Z} = \boldsymbol{a}^{\otimes 3} + \boldsymbol{b}^{\otimes 3} + \boldsymbol{c}^{\otimes 3}.$$

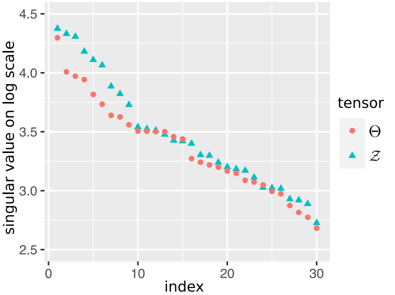
Inadequacies of low-rank models

- Sensitivity to order-preserving transformation

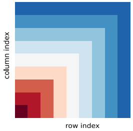


$$\Theta = \frac{1}{1+\exp(-c\,(\mathcal{Z}))}, \quad \text{where}$$
$$\mathcal{Z} = \boldsymbol{a}^{\otimes 3} + \boldsymbol{b}^{\otimes 3} + \boldsymbol{c}^{\otimes 3}.$$

- Inadequacy for special structures.



$$\Theta = \log(1+\mathcal{Z}), \quad \text{where}$$
$$\mathcal{Z}(i,j,k) = \frac{1}{d}\max(i,j,k).$$



Why sign matters?

For a bounded tensor $\Theta \in [-1,1]^{d_1\times\cdots\times d_K}$,

$$\Theta \approx \frac{1}{|\mathcal{H}|}\sum_{\pi\in\mathcal{H}} \text{sgn}(\Theta-\pi), \quad \text{where } \mathcal{H} = \left\{-1\ldots,-\frac{1}{H},0,\frac{1}{H},\ldots,1\right\}.$$

- Sign tensors are invariant to order-preserving transformation.
- More flexible signal tensors are allowed by using sign tensor series representation.
- In noisy case, we estimate $\text{sgn}(\Theta-\pi)$ from the tensor data $\text{sgn}(\mathcal{Y}-\pi)$.

Sign rank

- Key idea: we use a local notion of low-rankness to allow a richer family of signal tensors.
- Two tensors are sign equivalent denoted $\Theta \simeq \Theta'$ if $\text{sgn}(\Theta) = \text{sgn}(\Theta')$, where

$$[\text{sgn}(\Theta)]_\omega := \begin{cases} 1 & \text{if } \Theta_\omega \geq 0, \\ -1 & \text{otherwise.} \end{cases}$$

.

- Sign rank is defined as

$$\text{srank}(\Theta) = \min\{\text{rank}(\Theta') \colon \Theta' \simeq \Theta, \Theta' \in \mathbb{R}^{d_1 \times \cdots \times d_K}\}.$$

Sign rank

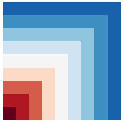
- Key idea: we use a local notion of low-rankness to allow a richer family of signal tensors.
- Two tensors are sign equivalent denoted $\Theta \simeq \Theta'$ if $\text{sgn}(\Theta) = \text{sgn}(\Theta')$, where

$$[\text{sgn}(\Theta)]_\omega := \begin{cases} 1 & \text{if } \Theta_\omega \geq 0, \\ -1 & \text{otherwise.} \end{cases}$$

.

- Sign rank is defined as

$$\text{srank}(\Theta) = \min\{\text{rank}(\Theta') \colon \Theta' \simeq \Theta, \Theta' \in \mathbb{R}^{d_1 \times \cdots \times d_K}\}.$$

$\Theta =$  , $\text{sgn}(\Theta) =$  $\Longrightarrow$ $\text{rank}(\Theta) = d$, $\text{srank}(\Theta) = 2$

Sign representable tensors

Sign representable tensors

A tensor Θ is called r-sign representable if the tensor $(\Theta - \pi)$ has sign rank bounded by r for all $\pi \in [-1, 1]$.

- Most existing structure tensors belong to sign representable family:
  - Low-rank CP tensors, Tucker tensors, stochastic block models.
  - High-rank tensors from GLM, single index models,
  - Tensors with repeating patterns, e.g. $\Theta(i_1, \ldots, i_K) = \log(1 + \max(i_1, \ldots, i_K))$ is 2-sign representable.

Sign representable tensors

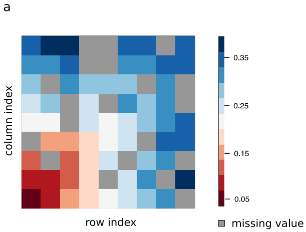
Sign representable tensors

A tensor Θ is called r-sign representable if the tensor $(\Theta - \pi)$ has sign rank bounded by r for all $\pi \in [-1, 1]$.

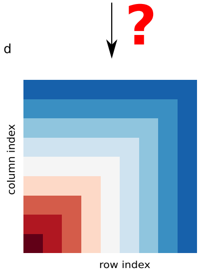
- Most existing structure tensors belong to sign representable family:
  - Low-rank CP tensors, Tucker tensors, stochastic block models.
  - High-rank tensors from GLM, single index models,
  - Tensors with repeating patterns, e.g. $\Theta(i_1, \ldots, i_K) = \log(1 + \max(i_1, \ldots, i_K))$ is 2-sign representable.
- Instead of the classical low-rank assumption, we propose the sign representable tensor family

$$\Theta \in \mathscr{P}_{\text{sgn}}(r) := \{\Theta\colon \text{srank}(\Theta - \pi) \leq r \text{ for all } \pi \in [-1, 1]\}.$$

Our solution: sign signal helps!

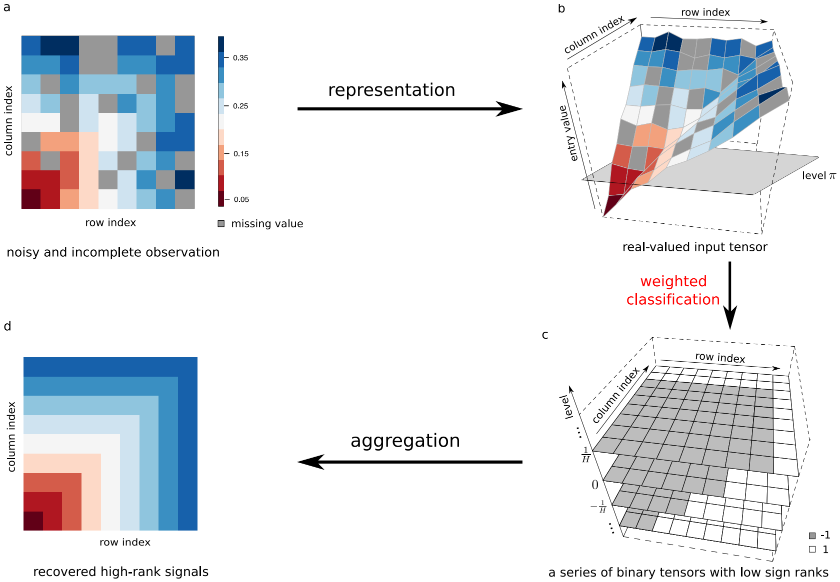


noisy and incomplete observation

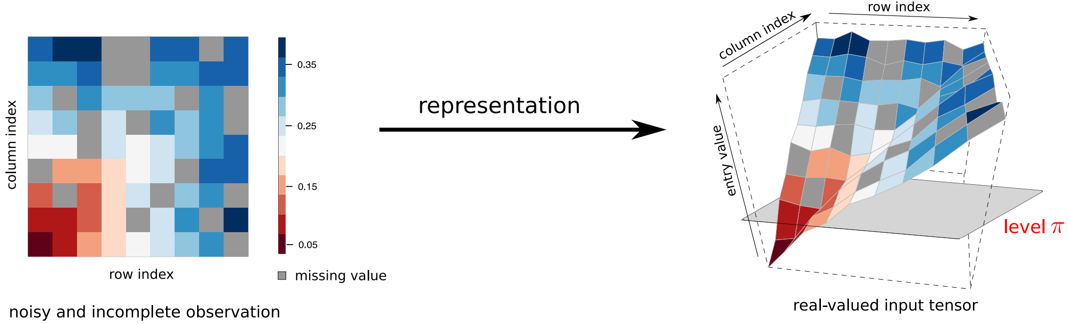


recovered high-rank signals

Our solution: sign signal helps!



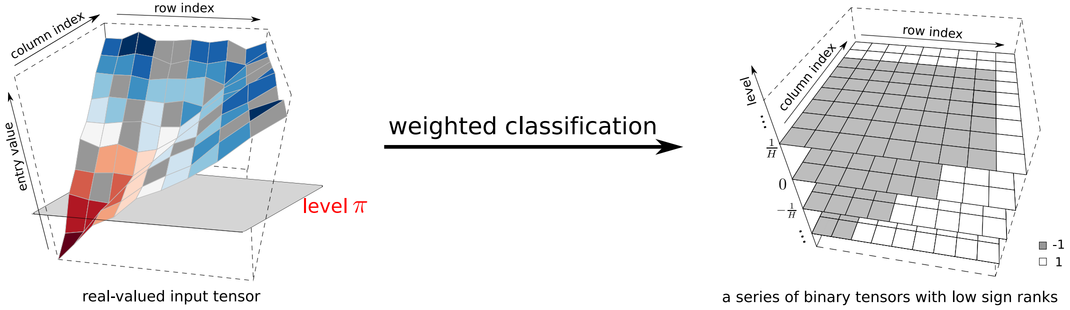
Step 1: representation



- We observe a noisy incomplete tensor $\mathcal{Y}_\Omega \in [-1,1]^{d_1\times\cdots\times d_K}$ with observed index set $\Omega \subset [d_1]\times\cdots\times[d_K]$.
- We dichotomize the data into a series of sign tensors:

$$\{\text{sgn}(\mathcal{Y}_\Omega - \pi)\}_{\pi\in\mathcal{H}}, \quad \text{where } \mathcal{H} = \left\{-1,\ldots,-\frac{1}{H},0,\frac{1}{H},\ldots,1\right\}.$$

Step 2: weighted classification



- We estimate $\text{sgn}(\Theta - \pi)$ through $\text{sgn}(\mathcal{Y}_\Omega - \pi)$ via weighted classification.
- Objective function of weighted classification is

$$L(\mathcal{Z}, \mathcal{Y}_\Omega - \pi) = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \underbrace{|\mathcal{Y}(\omega) - \pi|}_{\text{weight}} \times \underbrace{|\text{sgn}(\mathcal{Z}(\omega)) - \text{sgn}(\mathcal{Y}(\omega) - \pi)|}_{\text{classification loss}}$$

Step 2: weighed classification

real-valued input tensor

a series of binary tensors with low sign ranks

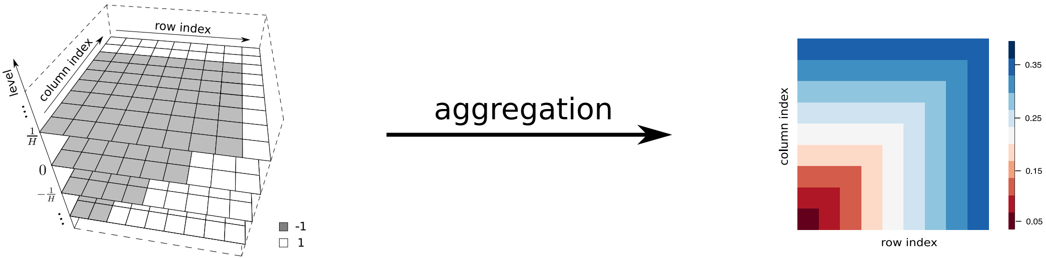
- If $\Theta \in \mathscr{P}_{\text{sgn}}(r)$ is α-smooth ($\alpha > 0$), we have a unique optimizer such that

$$\text{sgn}(\Theta - \pi) = \underset{\mathcal{Z}:\ \text{rank}(\mathcal{Z}) \le r}{\arg\min} \ \mathbb{E}_{\mathcal{Y}_\Omega} L(\mathcal{Z}, \mathcal{Y}_\Omega - \pi).$$

- We obtain a series of optimizers $\{\hat{\mathcal{Z}}_\pi\}_{\pi \in \mathcal{H}}$ as

$$\hat{\mathcal{Z}}_\pi = \underset{\mathcal{Z}:\ \text{rank}(\mathcal{Z}) \le r}{\arg\min} \ L(\mathcal{Z}, \mathcal{Y}_\Omega - \pi).$$

Step 3: aggregation



a series of binary tensors with low sign ranks

recovered high-rank signals

- From a series of optimizers $\{\hat{\mathcal{Z}}_\pi\}_{\pi\in\mathcal{H}}$ in the weighted classification, we obtain the tensor estimate

$$\hat{\Theta} = \frac{1}{2H+1}\sum_{\pi\in\mathcal{H}} \text{sgn}\hat{\mathcal{Z}}_\pi.$$

Identification for sign tensor estimation

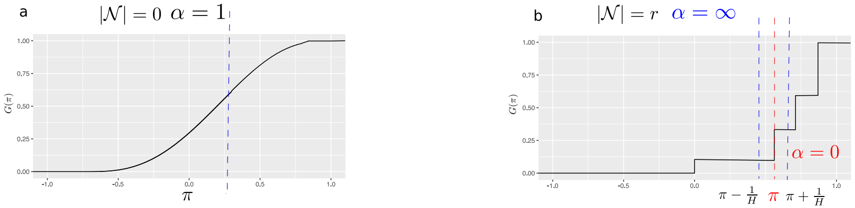
- We quantify difficulty of the problem using CDF $G(\pi) = \mathbb{P}_{\omega \in \Pi}[\Theta(\omega) \leq \pi]$.

α-smoothness

- Partition $[-1,1] = \mathcal{N} \cup \mathcal{N}^c$, where $\mathcal{N}^c$ consists of levels whose pseudo density (histogram with bin size $\Delta s = d^{-K}$) is uniformly bounded, and $\mathcal{N}$ otherwise.
- $G(\pi)$ is globally α-smooth in that for all $\pi \in \mathcal{N}^c$,

$$\sup_{\Delta s \leq t < \rho(\pi, \mathcal{N})} \frac{G(\pi + t) - G(\pi - t)}{t^{\alpha}} \leq c,$$

for two constants $\alpha, c > 0$, where $\rho(\pi, \mathcal{N}) = \min_{\pi' \in \mathcal{N}} |\pi - \pi'| + \Delta s$.



Estimation error

For two tensor Θ_1, Θ_2, define $\text{MAE}(\Theta_1, \Theta_2) = \mathbb{E}_{\omega \in \Pi} |\Theta_1(\omega) - \Theta_2(\omega)|$.

Estimation error (L. and Wang 2021)

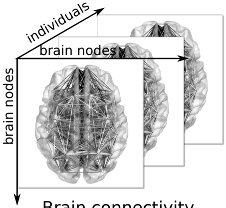
Suppose $\Theta \in \mathscr{P}_{\text{sgn}}(r)$ is α-smooth with bounded $|\mathcal{N}|$, and $d_1 = \cdots = d_K = d$.

1. (Sign tensor estimation) For all $\pi \in \mathcal{N}^c$, with high probability,
$$\text{MAE}(\text{sgn}\hat{\mathcal{Z}}_\pi, \text{sgn}(\Theta - \pi)) \lesssim^* \left(\frac{dr}{|\Omega|}\right)^{\frac{\alpha}{\alpha+2}}.$$
2. (Tensor estimation)
$$\text{MAE}(\hat{\Theta}, \Theta) \lesssim^* \underbrace{\left(\frac{dr}{|\Omega|}\right)^{\frac{\alpha}{\alpha+2}}}_{\text{Error inherited from sign estimation}} + \underbrace{\frac{1}{H}}_{\text{Bias}} + \underbrace{\frac{Hdr}{|\Omega|}}_{\text{Variance}} \asymp^{**} \left(\frac{dr}{|\Omega|}\right)^{\min\left(\frac{\alpha}{\alpha+2}, \frac{1}{2}\right)}.$$

*log term suppressed, **$H \asymp (|\Omega|/dr)^{1/2}$

- Tensor estimation is generally no better than sign tensor estimation.
- See paper for general case that allows unbounded $|\mathcal{N}|$ and sub-Gaussian noise.

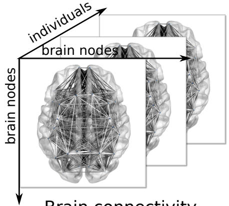
Data application: Brain connectivity



Brain connectivity

- The human brain connectivity dataset consists of 68 brain regions for 114 individuals with their IQ scores.
- Data tensor $\mathcal{Y} \in \{0,1\}^{68\times 68\times 114}$.

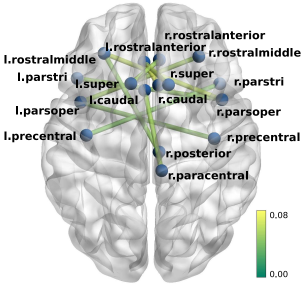
Data application: Brain connectivity



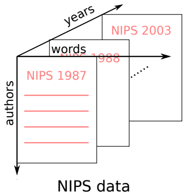
Brain connectivity

- The human brain connectivity dataset consists of 68 brain regions for 114 individuals with their IQ scores.
- Data tensor $\mathcal{Y} \in \{0,1\}^{68\times 68\times 114}$.

- We examine the estimated signal tensor $\hat{\Theta}$.
- Top 10 brain edges based on regression analysis show inter-hemisphere connections.

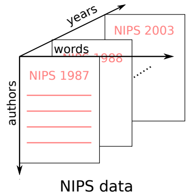


Data application: NIPS



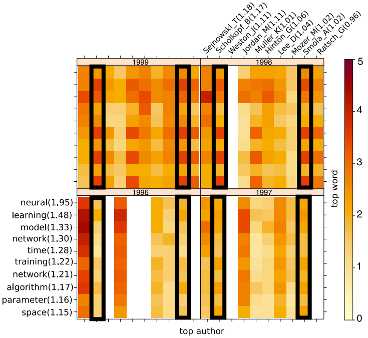
- The NIPS dataset consists of word occurrence counts in papers published from 1987 to 2003.
- Data tensor $\mathcal{Y} \in \mathbb{R}^{100 \times 200 \times 17}$.

Data application: NIPS



- The NIPS dataset consists of word occurrence counts in papers published from 1987 to 2003.
- Data tensor $\mathcal{Y} \in \mathbb{R}^{100 \times 200 \times 17}$.

- We examine the estimated signal tensor $\hat{\Theta}$.
- Most frequent words are consistent with the active topics
- Strong heterogeneity among word occurrences across authors and years.
- Similar word patterns (B. Schölkopf and A. Smola).



Data application: Brain connectivity + NIPS

MRN-114 brain connectivity dataset					
Method	$r = 3$	$r = 6$	$r = 9$	$r = 12$	$r = 15$
NonparaT (Ours)	**0.18**(0.001)	**0.14**(0.001)	**0.12**(0.001)	**0.12**(0.001)	**0.11**(0.001)
Low-rank CPT	0.26(0.006)	0.23(0.006)	0.22(0.004)	0.21(0.006)	0.20(0.008)
NIPS word occurrence dataset					
Method	$r = 3$	$r = 6$	$r = 9$	$r = 12$	$r = 15$
NonparaT (Ours)	**0.18**(0.002)	**0.16**(0.002)	**0.15**(0.001)	**0.14**(0.001)	**0.13**(0.001)
Low-rank CPT	0.22(0.004)	0.20(0.007)	0.19(0.007)	0.17(0.007)	0.17(0.007)
Naive imputation (Baseline)	0.32(.001)				

Table: MAE comparison in the brain data and NIPS data on 5-folded cross-validation

- Our method outperforms the low-rank CP method in applications.

Summary

- We have developed a completion method that address both low- and high-rankness based on sign series representation.
- Estimation error rates and sample complexities are established.
- Our approach has good interpretation and prediction performance in both simulations and data applications.
- Preprint: https://arxiv.org/abs/2102.00384
- Software: https://cran.r-project.org/web/packages/TensorComplete/index.html

Thank you!

References I

Anandkumar, A., Ge, R., Hsu, D., Kakade, S. M., and Telgarsky, M. (2014). Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research*, 15(1):2773–2832.

Cai, C., Li, G., Poor, H. V., and Chen, Y. (2019). Nonconvex low-rank tensor completion from noisy data. In *Advances in Neural Information Processing Systems*, pages 1863–1874.

Ganti, R., Rao, N., Balzano, L., Willett, R., and Nowak, R. (2017). On learning high dimensional structured single index models. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pages 1898–1904.

Lee, C., Li, L., Zhang, H. H., and Wang, M. (2021). Nonparametric trace regression in high dimensions via sign series representation. *arXiv preprint arXiv:2105.01783*.

Lee, C. and Wang, M. (2020). Tensor denoising and completion based on ordinal observations. In *International Conference on Machine Learning*, pages 5778–5788.

References II

Montanari, A. and Sun, N. (2018). Spectral algorithms for tensor completion. *Communications on Pure and Applied Mathematics*, 71(11):2381–2425.

Wang, M. and Zeng, Y. (2019). Multiway clustering via tensor block models. In *Advances in Neural Information Processing Systems*, pages 713–723.