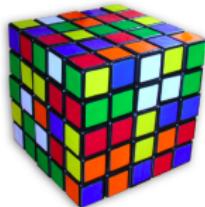


Beyond matrices: nonparametric tensor completion via sign series

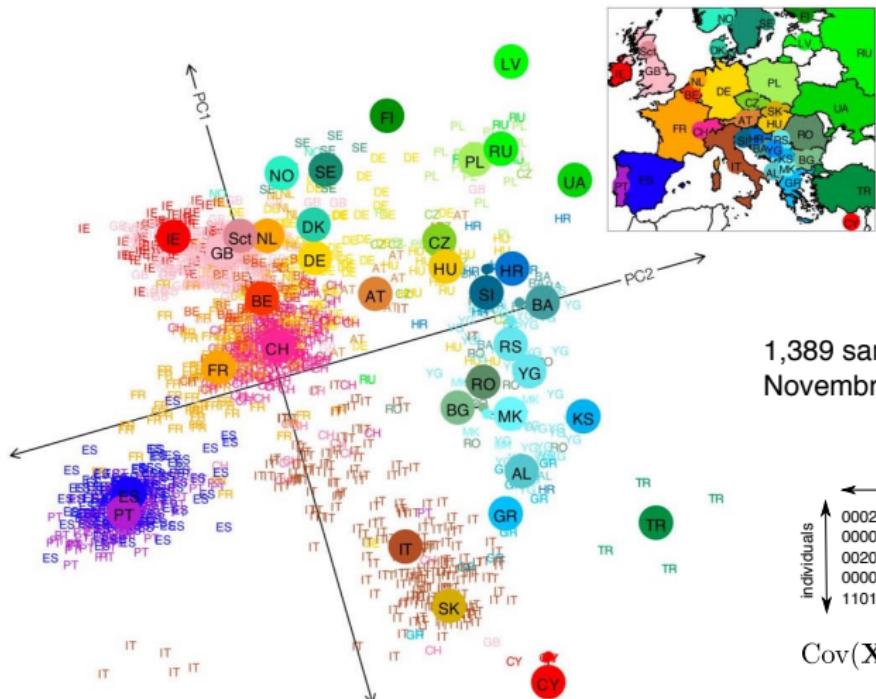
Miaoyan Wang

Department of Statistics, UW-Madison

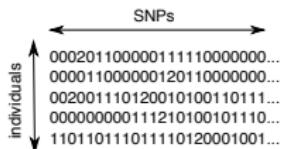
Joint work with Chanwoo Lee (3rd-year PhD student)



A successful story: PCA of Europeans

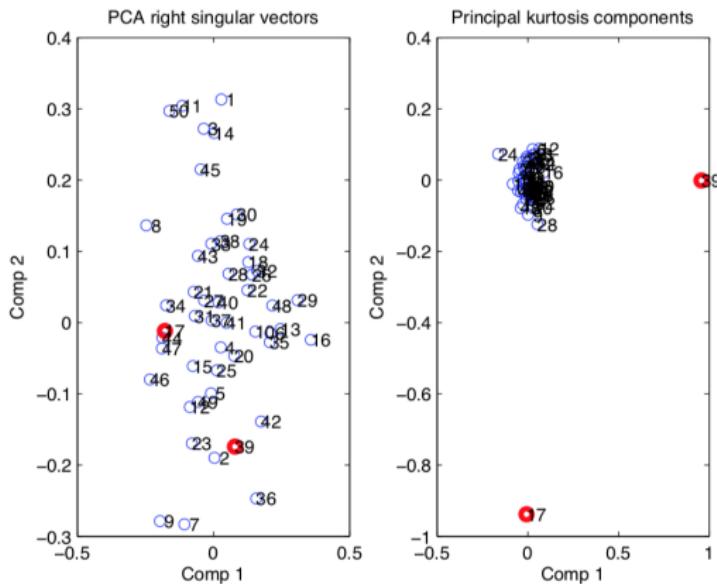


1,389 samples, ~ 200k SNPs
Novembre et al. (2008)



$$\text{Cov}(\mathbf{X}) = \sum_{i=1}^r \lambda_i \mathbf{u}_i \mathbf{u}_i^T$$

Matrix methods are powerful, however...



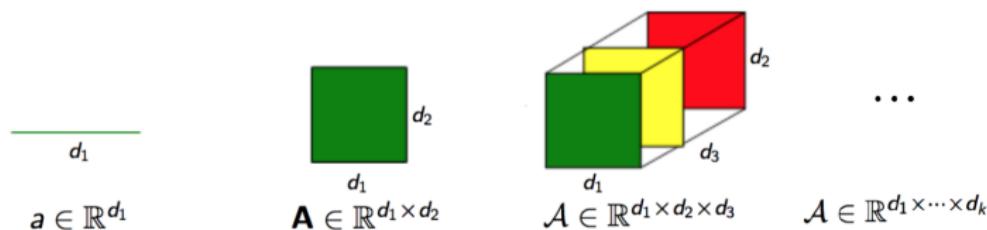
All Gaussian except points 17 and 39.

left: matrix PCA; right: principal components of kurtosis.

Figure credit: Jason Morton and Lek-Heng Lim (2009/2015).

What is a tensor?

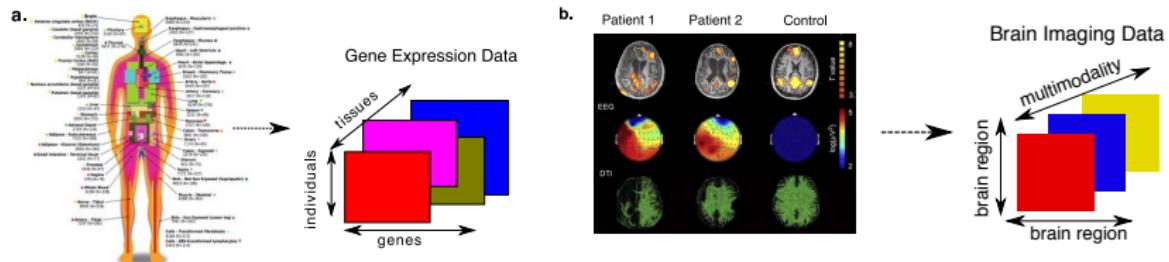
- ▶ Tensors are generalizations of vectors and matrices:



- ▶ An order- k tensor $\mathcal{A} = [[a_{i_1 \dots i_k}]] \in \mathbb{R}^{d_1 \times \dots \times d_k}$ is a hypermatrix with dimensions (d_1, \dots, d_k) and entries $a_{i_1 \dots i_k} \in \mathbb{R}$.
- ▶ This talk will focus on tensor of order 3 or greater, also known as **higher-order tensors**.

Tensors in genomics

- ▶ Many datasets come naturally in a multiway form.
- ▶ Multi-tissue, multi-individual gene expression measures could be organized as an order-3 tensor $\mathcal{A} = [[a_{git}]] \in \mathbb{R}^{n_G \times n_I \times n_T}$.



Tensors in statistical modeling

“Tensors are the new matrices” that tie together a wide range of areas:

- ▶ Longitudinal social network data $\{\mathbf{Y}_t : t = 1, \dots, n\}$
- ▶ Spatio-temporal transcriptome data
- ▶ Joint probability table of a set of variables $\mathbb{P}(X_1, X_2, X_3)$
- ▶ Higher-order moments in topic models
- ▶ Markov models for the phylogenetic tree $K_{1,3}$

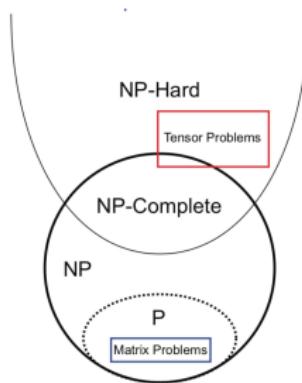
M. Yuan et al 2017, P. Hoff 2015, Montanari & Richard 2014

Anandkumar et al 2014, Mossel et al 2004, P. McCullagh 1987

Talk outline

Prohibitive Computational Complexity

Most higher-order tensor problems are NP-hard [Hillar & Lim, 2013].



Fortunately, the tensors sought in statistical and machine learning applications are often **specially structured**:

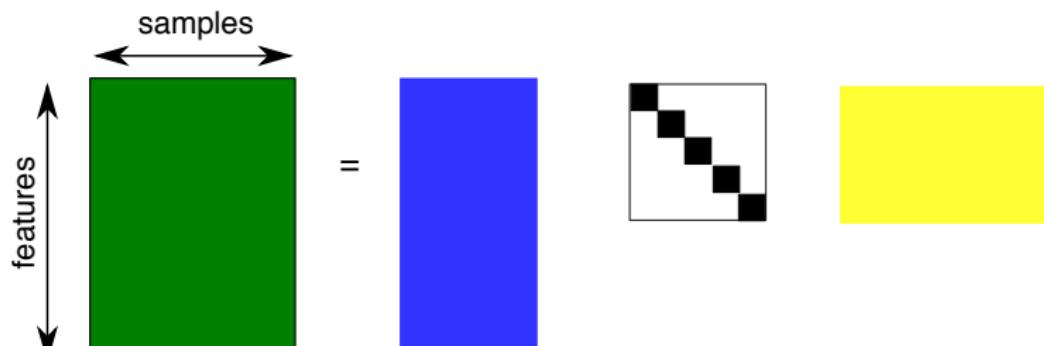
- ▶ Low-rankness
- ▶ Sparsity
- ▶ Non-negativity
- ▶ ...

This talk is based on

Beyond the Signs: Nonparametric Tensor Completion via Sign Series. Lee and W.

- ▶ arXiv:

Review: Matrix SVD for biclustering



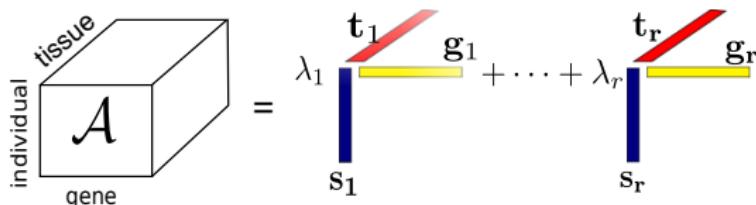
$$\begin{aligned}\mathbf{X} &= \mathbf{U} \quad \Lambda \quad \mathbf{V}^T \\ &= \sum_{i=1}^r \lambda_i \mathbf{u}_i \mathbf{v}_i^T\end{aligned}$$

- ▶ Columns of \mathbf{U} describe patterns across samples
- ▶ Columns of \mathbf{V}^T describe patterns across genes

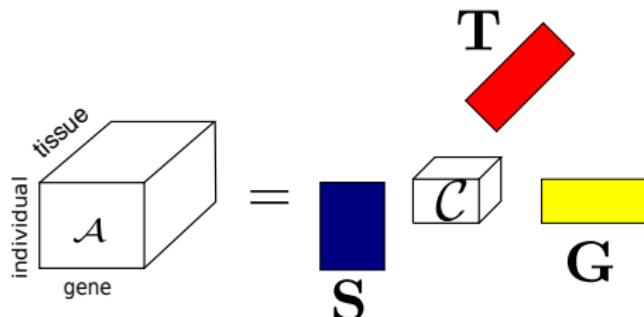
Y Kluger et al, Genome Research (2003). 13(4): 703-71
Data Science Specialization (COURSERA) by Brian Caffo and Jeff Leek

Various notions of low-rankness

- Canonical polyadic (CP) low-rankness: $\mathcal{A} = \sum_{r=1}^R \lambda_r \mathbf{s}_r \otimes \mathbf{g}_r \otimes \mathbf{t}_r.$

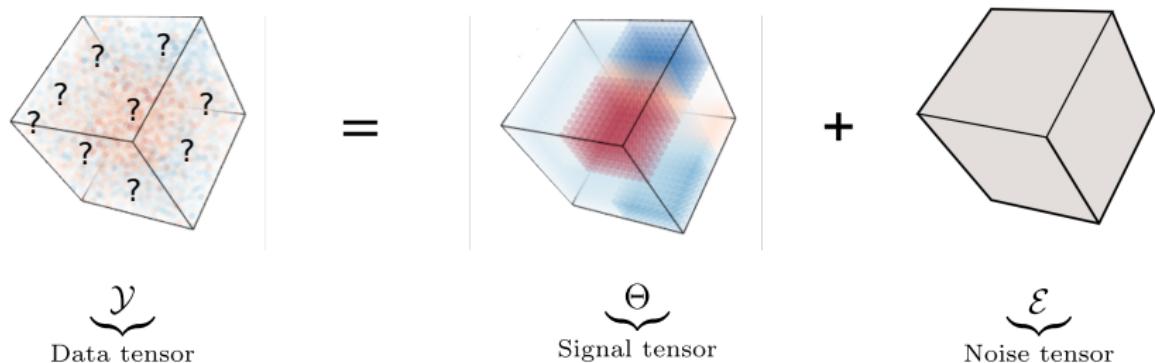


- Tucker low-rankness: $\mathcal{A} = \mathcal{C} \times_1 \mathbf{S} \times_2 \mathbf{G} \times_3 \mathbf{T}.$



- Others: tensor train [Oseledet '11], tensor block model [W. & Zeng '19; Han, Luo, W. et al '20], etc.

Setup: signal plus noise model

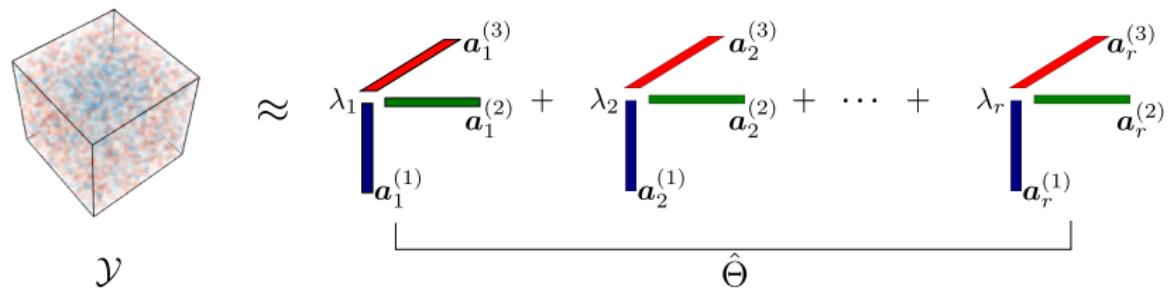


We focus on two problems:

- ▶ Nonparametric tensor estimation: How to estimate the signal tensor Θ under **a wide range of structures?**
- ▶ Tensor completion: How many **observed tensor entries** do we need in order for consistent recovery?

Inadequacies of low-rank models

- ▶ Classical low-rank models (Jain & Oh '14; Motanari & Sun '18):

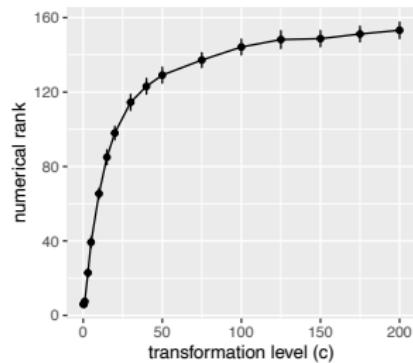


- ▶ Two limitations of classical low-rank models:

1. Sensitivity to order-preserving transformation.
2. Inadequacy for special structures.

Inadequacies of low-rank models

- ▶ Tensor rank is sensitive to order-preserving transformation.

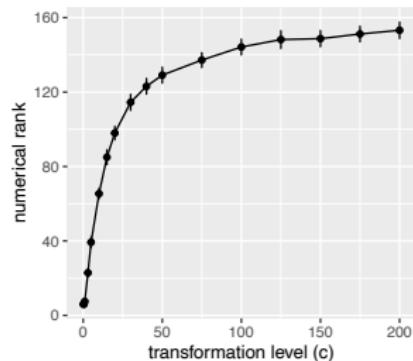


$$\Theta = \frac{1}{1 + \exp(-c(\mathcal{Z}))}, \quad \text{where}$$
$$\mathcal{Z} = \mathbf{a}^{\otimes 3} + \mathbf{b}^{\otimes 3} + \mathbf{c}^{\otimes 3}$$

⇒ Θ is high-rank but \mathcal{Z} is low-rank.

Inadequacies of low-rank models

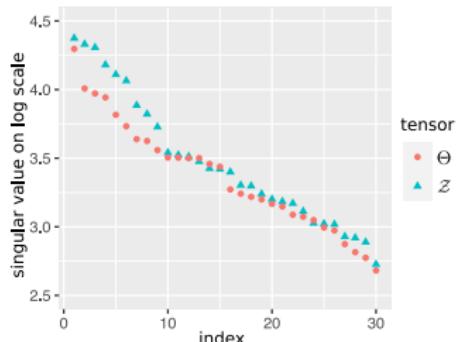
- ▶ Tensor rank is sensitive to order-preserving transformation.



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

⇒ Θ is high-rank but \mathcal{Z} is low-rank.

- ▶ Low-rank model fails to address several important structures.



$$\Theta = \log(1 + \mathcal{Z}), \quad \text{where}$$

$$\mathcal{Z} = [\mathcal{Z}(i, j, k)] = \frac{1}{d} \max(i, j, k).$$

⇒ Both Θ and \mathcal{Z} are full rank.

Related work and our contributions

- ▶ **Low-rank tensor estimation** (Anandkumar et al. '14; Montanari & Sun '18; Cai et al. '19) focuses on the regime of fixed rank r growing dimension d . ⇒ low-rank assumption is often violated in practice.
- ▶ **High-rank matrix estimation** was studied under nonlinear models (Ganti et al. '15) and subspace clustering (Ongie et al. '17; Fan & Udell '19). ⇒ tensors are more challenging because tensor rank may exceed dimension.

Related work and our contributions

- ▶ **Low-rank tensor estimation** (Anandkumar et al. '14; Montanari & Sun '18; Cai et al. '19) focuses on the regime of fixed rank r growing dimension d . \Rightarrow low-rank assumption is often violated in practice.
- ▶ **High-rank matrix estimation** was studied under nonlinear models (Ganti et al. '15) and subspace clustering (Ongie et al. '17; Fan & Udell '19). \Rightarrow tensors are more challenging because tensor rank may exceed dimension.

Our contributions

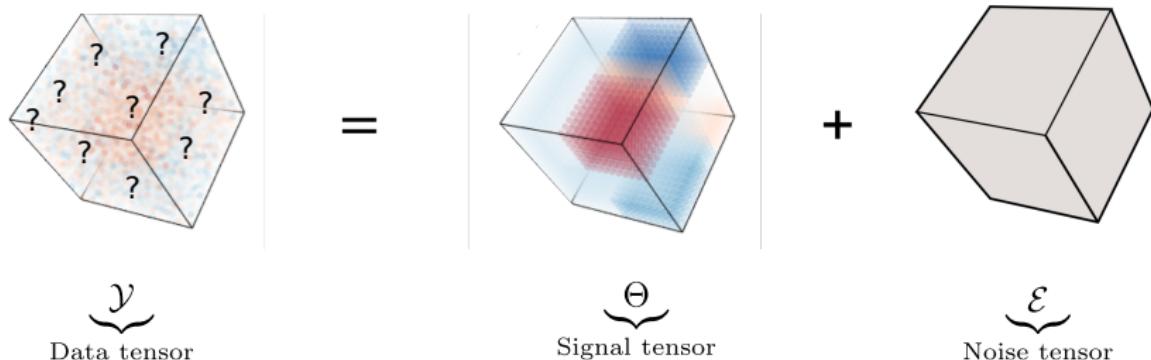
- ▶ We develop a new model called sign representable tensors to fill the gap between parametric (low-rank) and nonparametric (high-rank) tensors.
- ▶ Our tensor estimate is provably reductable to a series of classifications. \Rightarrow computationally efficient via a divide-and-conquer algorithm.

Why sign matters?

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

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

- ▶ We do not observe Θ ; instead, we observe a noisy incomplete version \mathcal{Y} .
- ▶ How to estimate the signal tensor Θ given data tensor \mathcal{Y} ?



Sign rank

- ▶ Key ideas: we use a local (nonparametric) notion of “low-rankness” to allow a broad family of signal tensors.
- ▶ Two tensors are sign equivalent, denoted $\Theta \simeq \Theta'$, if $\text{sgn}(\Theta) = \text{sgn}(\Theta')$.
- ▶ Define the **sign rank** by

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

$$\Theta = \begin{array}{c} \text{A 3D tensor with } 3 \text{ layers, each } 2 \times 2 \\ \text{The first layer has a red square at position } (1,1) \end{array}, \quad \text{sgn}(\Theta) = \begin{array}{c} \text{A 2D matrix with } 2 \text{ columns and } 2 \text{ rows} \\ \text{The top-left entry is dark red, all other entries are blue} \end{array} \implies \begin{array}{l} \text{rank}(\Theta) = \color{red}d \\ \text{srank}(\Theta) = \color{red}2 \end{array}$$

- ▶ For any strictly monotonic function $g: \mathbb{R} \rightarrow \mathbb{R}$ with $g(0) = 0$,

$$\text{srank}(\Theta) \leq \text{rank}(g(\Theta)).$$

Sign representable tensors

Sign representable tensors

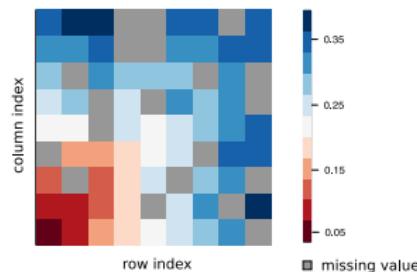
A tensor Θ is called **r-sign representable** if the tensor $(\Theta - \pi)$ has sign rank bounded by r for all $\pi \in [-1, 1]$.

- ▶ Most existing structured tensors belong to sign representable family:
 - ▶ **Low-rank** CP tensors, Tucker tensors, stochastic tensor block models.
 - ▶ **High-rank** tensors from GLM, single index models.
 - ▶ Earlier example $\Theta(i_1, \dots, i_K) = \log(1 + \max(i_1, \dots, i_K))$ is 2-sign representable \Rightarrow properties extends to general max/min hypergraphon models.
- ▶ We propose the signal tensor family

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

Our solution: sign signal helps!

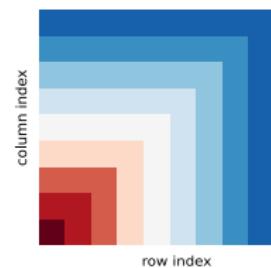
a



noisy and incomplete observation

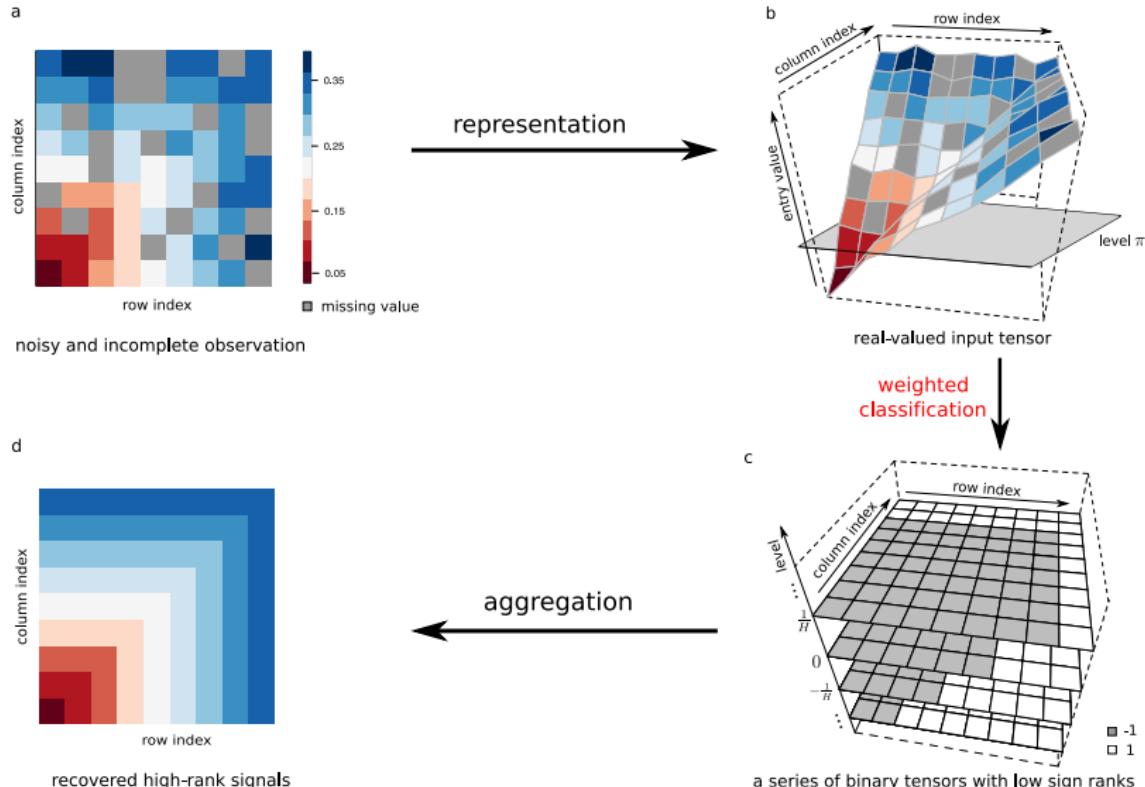


d



recovered high-rank signals

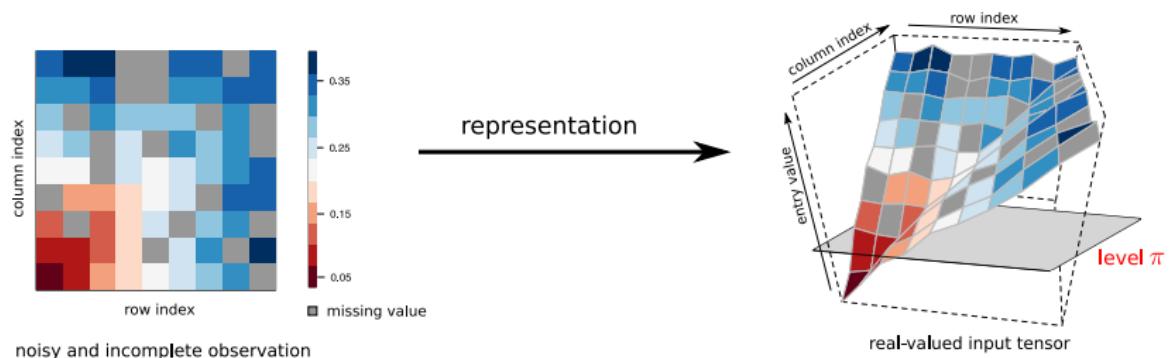
Our solution: sign signal helps!



Sign representation

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

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



Sign estimation via weighted classification

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

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



Identification for sign tensor estimation

α -smoothness of signal tensor

For fixed π , Θ is α -smooth if there exist $\alpha = \alpha(\pi) > 0, c = c(\pi) > 0$, s.t.

$$\sup_{0 \leq t < \rho(\pi, \mathcal{N})} \frac{\mathbb{P}_{\omega \sim \Pi}[|\Theta(\omega) - \pi| \leq t]}{t^\alpha} \leq c,$$

where $\rho(\pi, \mathcal{N}) = \min_{\pi' \in \mathcal{N}} |\pi - \pi'|$ and $\mathcal{N} = \{\pi : \mathbb{P}(\Theta(\omega) = \pi) \neq 0\}$. If α and c are global constants for all* π 's, we call Θ is α -globally smooth.

* except for a finite number of π 's.

Identification for sign tensor estimation

α -smoothness of signal tensor

For fixed π , Θ is α -smooth if there exist $\alpha = \alpha(\pi) > 0, c = c(\pi) > 0$, s.t.

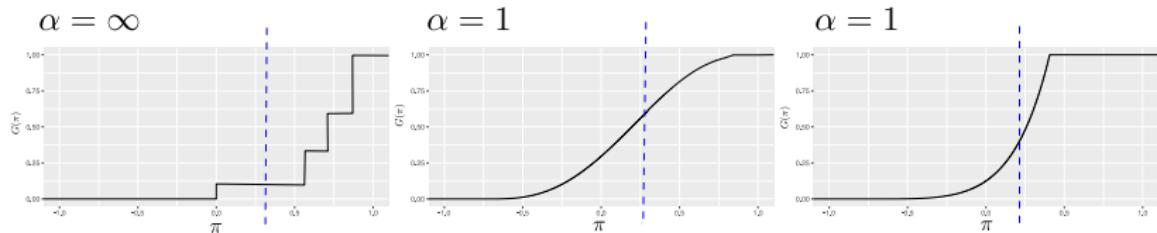
$$\sup_{0 \leq t < \rho(\pi, \mathcal{N})} \frac{\mathbb{P}_{\omega \sim \Pi}[|\Theta(\omega) - \pi| \leq t]}{t^\alpha} \leq c,$$

where $\rho(\pi, \mathcal{N}) = \min_{\pi' \in \mathcal{N}} |\pi - \pi'|$ and $\mathcal{N} = \{\pi : \mathbb{P}(\Theta(\omega) = \pi) \neq 0\}$. If α and c are global constants for all* π 's, we call Θ is α -globally smooth.

* except for a finite number of π 's.

Intuition: **sign recovery** is harder at levels where point mass concentrates.

Rate depends on the behavior of CDF function $G(\pi) = \mathbb{P}_{\omega \sim \Pi}[\Theta(\omega) \leq \pi]$.



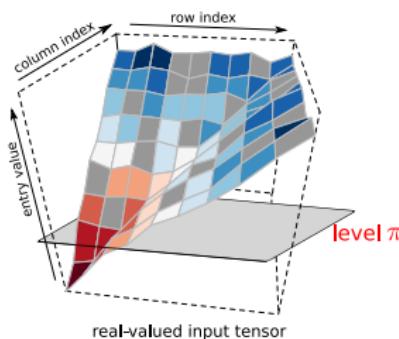
Identification for sign tensor estimation

- If Θ is α -smooth ($\alpha > 0$), we have **a unique optimizer** such that

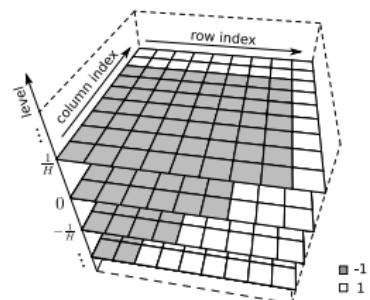
$$\text{sgn}(\Theta - \pi) = \arg \min_{\mathcal{Z}: \text{rank}(\mathcal{Z}) \leq r} \mathbb{E}_{\mathcal{Y}_\Omega} L(\mathcal{Z}, \mathcal{Y}_\Omega - \pi).$$

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

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



weighted classification



a series of binary tensors with low sign ranks

* Uniqueness up to sign equivalence, meaning the optimizer $\Theta_{\text{opt}} \simeq \text{sgn}(\Theta - \pi)$.

Sign tensor estimation error

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

Sign tensor estimation for fixed π (Lee and W. 2021)

Suppose $\Theta \in \mathcal{P}_{\text{sgn}}(r)$ and $\Theta(\omega)$ is α -smooth for fixed π . Let $d_{\max} = \max_{k \in [K]} d_k$. Then, with very high probability over \mathcal{Y}_Ω ,

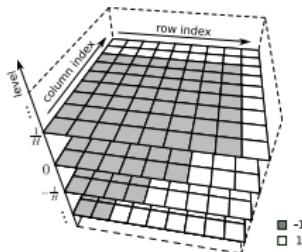
$$\text{MAE}(\text{sgn} \hat{\mathcal{Z}}_\pi, \text{sgn}(\Theta - \pi)) \lesssim \left(\frac{d_{\max} r}{|\Omega|} \right)^{\frac{\alpha}{\alpha+2}}.$$

- ▶ Sign estimation error shows a polynomial decay with $|\Omega|$.
- ▶ Best rate attains at $\alpha = \infty$ for stochastic tensor block models.

From sign to signal estimation

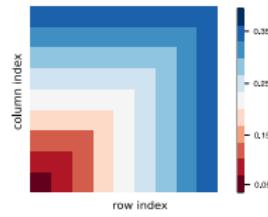
- ▶ Aggregation of sign tensors from weighted classification yields our signal tensor estimate:

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



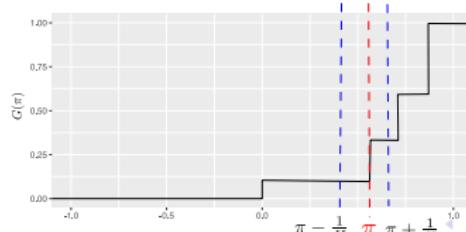
a series of binary tensors with low sign ranks

aggregation →



recovered high-rank signals

- ▶ Signal tensor estimation is robust to a few off-target classifications.



Signal tensor estimation error

Tensor estimation error (Lee and W. 2021)

Suppose $\Theta \in \mathcal{P}_{\text{sgn}}(r)$ and $\Theta(\omega)$ is α -globally smooth. Then, with very high probability over \mathcal{Y}_Ω ,

$$\text{MAE}(\hat{\Theta}, \Theta) \lesssim \underbrace{\left(\frac{d_{\max} r}{|\Omega|} \right)^{\frac{\alpha}{\alpha+2}}}_{\text{error inherited from sign estimation}} + \underbrace{\frac{1}{H}}_{\text{Bias}} + \underbrace{\frac{H d_{\max} r}{|\Omega|}}_{\text{Variance}}.$$

In particular, setting $H \asymp \left(\frac{|\Omega|}{d_{\max} r} \right)^{1/2}$ yields the error bound

$$\text{MAE}(\hat{\Theta}, \Theta) \lesssim \left(\frac{d_{\max} r}{|\Omega|} \right)^{\frac{\alpha}{\alpha+2} \vee \frac{1}{2}}.$$

- ▶ Sample requirement for tensor completion:

$$\text{MAE}(\hat{\Theta}, \Theta) \rightarrow 0, \text{ as } \frac{|\Omega|}{d_{\max} r} \rightarrow \infty.$$

Comparison of estimation error versus dimension

- We simulate signal tensors under a wide range of complexity.

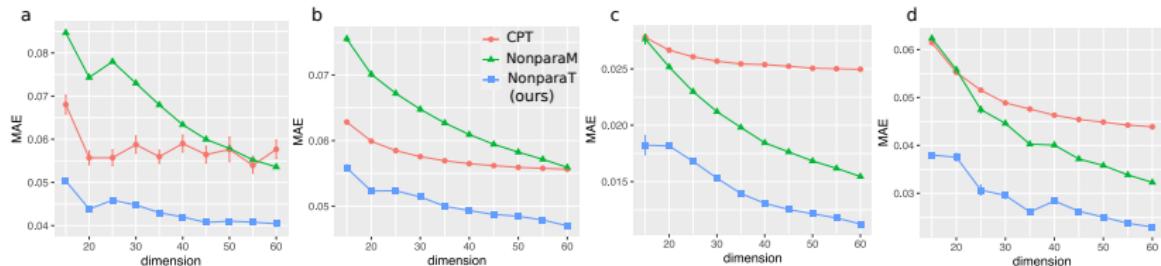
Simulation	Signal Tensor Θ	Rank	Sign Rank	Global α	CDF	Noise
1	$\mathcal{C} \times M_1 \times M_2 \times M_3$	3^3	$\leq 3^3$	∞		Uniform $[-0.3, 0.3]$
2	$ \mathbf{a} \otimes \mathbf{1} \otimes \mathbf{1} - \mathbf{1} \otimes \mathbf{a} \otimes \mathbf{1} $	d	≤ 3	1		Normal $\mathcal{N}(0, 0.15)$
3	$\log(0.5 + Z_{\max})$	$\geq d$	2	1		Uniform $[-0.1, 0.1]$
4	$2.5 - \exp(\mathcal{Z}_{\min}^{1/3})$	$\geq d$	2	1		Normal $\mathcal{N}(0, 0.15)$

Comparison of estimation error versus dimension

- We simulate signal tensors under a wide range of complexity.

Simulation	Signal Tensor Θ	Rank	Sign Rank	Global α	CDF	Noise
1	$\mathcal{C} \times M_1 \times M_2 \times M_3$	3^3	$\leq 3^3$	∞		Uniform $[-0.3, 0.3]$
2	$ \mathbf{a} \otimes \mathbf{1} \otimes \mathbf{1} - \mathbf{1} \otimes \mathbf{a} \otimes \mathbf{1} $	d	≤ 3	1		Normal $\mathcal{N}(0, 0.15)$
3	$\log(0.5 + Z_{\max})$	$\geq d$	2	1		Uniform $[-0.1, 0.1]$
4	$2.5 - \exp(Z_{\min}^{1/3})$	$\geq d$	2	1		Normal $\mathcal{N}(0, 0.15)$

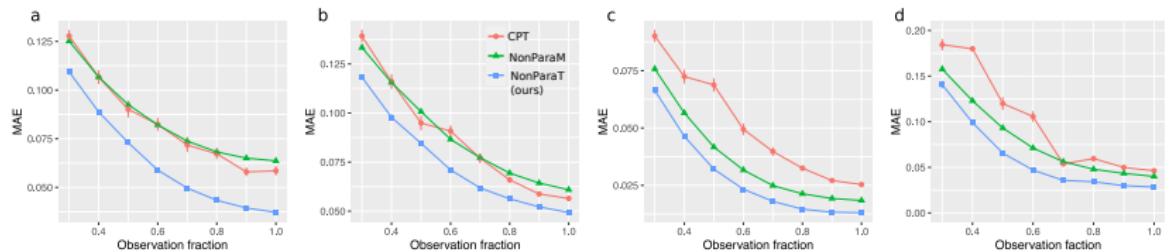
- Our method (**NonparaT**) achieves the best performance, whereas the second best method is low-rank CP tensor (**CPT**) for models 1-2, and matrix nonparametric method (**NonParaM**) for models 3-4.



Estimation error versus observation fraction

Simulation	Signal Tensor Θ	Rank	Sign Rank	Global α	CDF	Noise
1	$\mathcal{C} \times M_1 \times M_2 \times M_3$	3^3	$\leq 3^3$	∞		Uniform $[-0.3, 0.3]$
2	$ \mathbf{a} \otimes \mathbf{1} \otimes \mathbf{1} - \mathbf{1} \otimes \mathbf{a} \otimes \mathbf{1} $	d	≤ 3	1		Normal $\mathcal{N}(0, 0.15)$
3	$\log(0.5 + Z_{\max})$	$\geq d$	2	1		Uniform $[-0.1, 0.1]$
4	$2.5 - \exp(Z_{\min}^{1/3})$	$\geq d$	2	1		Normal $\mathcal{N}(0, 0.15)$

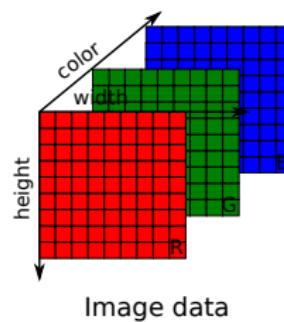
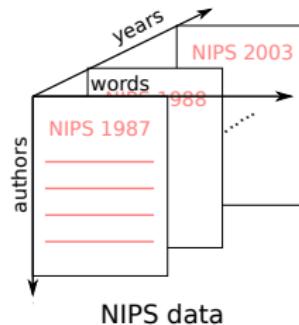
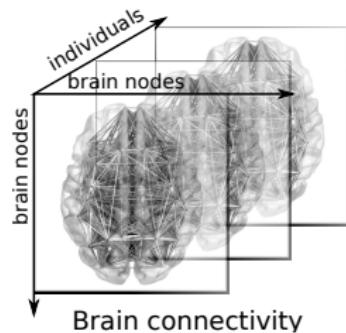
- Our method (NonparaT) achieves the best performance in completion.



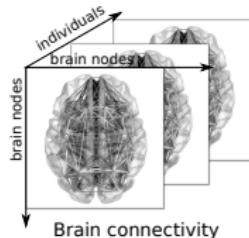
Data application

We apply our method to three datasets:

- ▶ The human brain connectivity data (Want et al '17) consists of 68 brain regions for 114 individuals along with their IQ scores.
- ▶ The NIPS dataset (Globerson et al '07) consistsis of word occurrence counts in papers published from 1987 to 2003.
- ▶ The 3-channel image data is from licensed google image file.

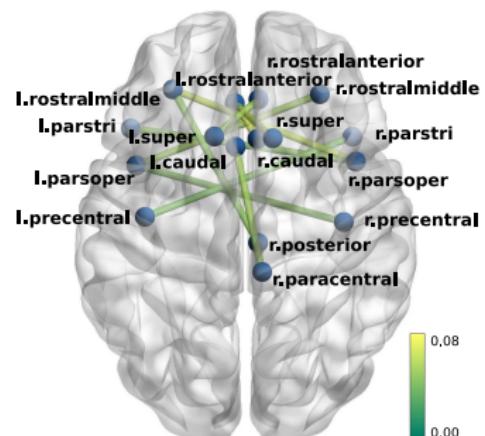


Data application: Brain connectivity

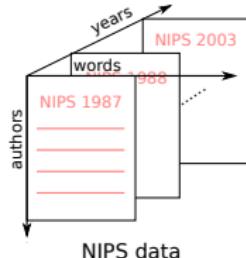


- ▶ The human brain connectivity data consists of 68 brain regions for 114 individuals along with their IQ scores.
- ▶ Data tensor $\mathcal{Y} \in \{0, 1\}^{68 \times 68 \times 114}$.

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

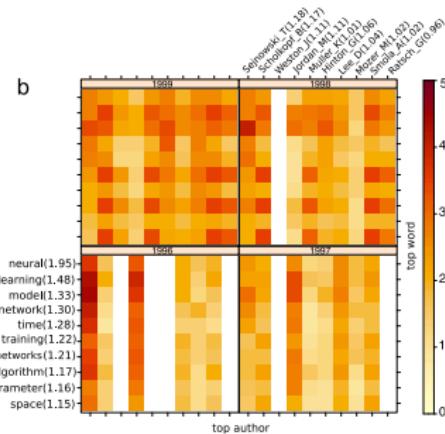


Data application: NIPS



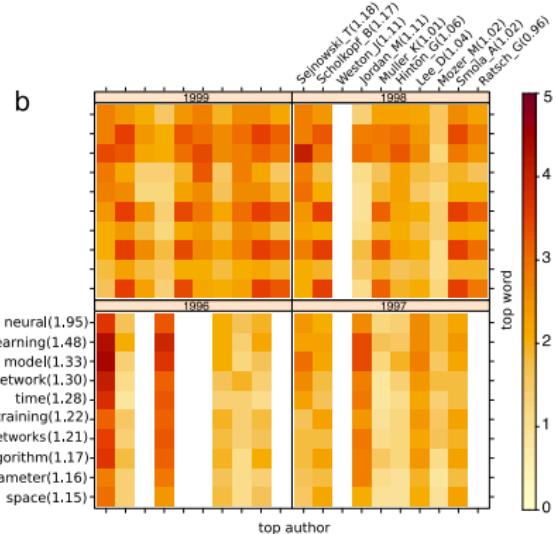
- ▶ The NIPS dataset consists of word occurrence counts in papers published from 1987 to 2003.
- ▶ Data tensor $\mathcal{Y} \in \mathbb{R}^{100 \times 200 \times 17}$.

- ▶ We examine the estimated signal tensor $\hat{\Theta}$.
- ▶ Most frequent words are consistent with active topics in NIPS conference.
- ▶ Strong heterogeneity among word occurrences across authors and years.



Data application: NIPS

- ▶ Top words: *neural* (1.95), *learning* (1.48), *network* (1.21), *training* (1.22), *parameter* (1.16).
- ▶ Top authors: *T. Sejnowski* (1.18), *B. Schölkopf* (1.17), *M. Jordan* (1.11), and *G. Hinton* (1.06),
- ▶ Top combinations:
 $(\text{training}, \text{algorithm}) \times (\text{B. Schölkopf}, \text{A. Smola})$
 $\times (1998, 1999)$,
 $(\text{model}) \times (\text{M. Jordan}) \times (1996)$.
- ▶ Similar word patterns in 1998-1999:
 $(\text{B. Schölkopf}, \text{Möller K}, \text{A. Smola}, \text{G. Rätsch}, \text{J. Weston}) \Rightarrow \text{Co-authors}$.



Numbers in parentheses denote marginal averages based on $\hat{\Theta}$

details

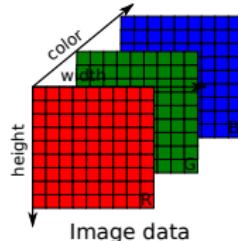
Data application: Brain connectivity + NIPS

- ▶ Our method achieves better test performance than low-rank methods.

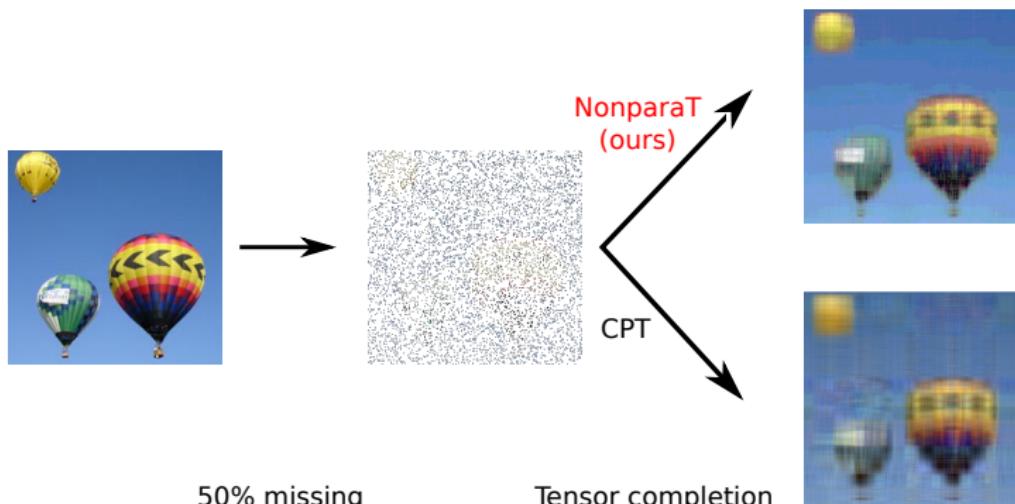
MRN-114 brain connectivity dataset					
Method	$r = 3$	$r = 6$	$r = 9$	$r = 12$	$r = 15$
NonparaT (Ours)	0.18(0.001)	0.14(0.001)	0.12(0.001)	0.12(0.001)	0.11(0.001)
Low-rank CPT	0.26(0.006)	0.23(0.006)	0.22(0.004)	0.21(0.006)	0.20(0.008)
NIPS word occurrence dataset					
Method	$r = 3$	$r = 6$	$r = 9$	$r = 12$	$r = 15$
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 based on 5-folded cross-validations.
Standard errors are reported in parenthesis.

Data application: Image



- ▶ The original data is from licensed google image file.
- ▶ Data tensor $\mathcal{Y} \in [0, 1]^{217 \times 217 \times 3}$.
- ▶ We assess completion performance by sampling 50% entries in the original image tensor.



Summary

Tensor analysis provides a rich source of

- ▶ fundamental problems in data science.
- ▶ new tools for long-standing questions.
- ▶ potentials for new applications.

Our general strategy is to carve out a broad range of **specially-structured tensors** that are useful in practice, and to develop efficient statistical methods for analyzing these high-dimensional tensor data.

References:

- ▶ Beyond the sign: nonparametric tensor completion from sign series.

Acknowledgment: NSF DMS 1915978 and Grant from Wisconsin Alumni Research Foundation.

Appendix

Algorithm 1 Nonparametric tensor completion

Input: Noisy and incomplete data tensor \mathcal{Y}_Ω , rank r , resolution parameter H .

```
1: for  $\pi \in \mathcal{H} = \{-1, \dots, -\frac{1}{H}, 0, \frac{1}{H}, \dots, 1\}$  do
2:   Random initialization of tensor factors  $\mathbf{A}_k = [\mathbf{a}_1^{(k)}, \dots, \mathbf{a}_r^{(k)}] \in \mathbb{R}^{d_k \times r}$  for all  $k \in [K]$ .
3:   while not convergence do
4:     for  $k = 1, \dots, K$  do
5:       Update  $\mathbf{A}_k$  while holding others fixed:
6:        $\mathbf{A}_k \leftarrow \arg \min_{\mathbf{A}_k \in \mathbb{R}^{d_k \times r}} \sum_{\omega \in \Omega} |\mathcal{Y}(\omega) - \pi| F(\mathcal{Z}(\omega) \text{sgn}(\mathcal{Y}(\omega) - \pi)),$ 
7:       where  $F(\cdot)$  is the large-margin loss, and  $\mathcal{Z} = \sum_{s \in [r]} \mathbf{a}_s^{(1)} \otimes \dots \otimes \mathbf{a}_s^{(K)}$  is a rank- $r$  tensor.
8:     end for
9:   end while
10:  Return  $\mathcal{Z}_\pi \leftarrow \sum_{s \in [r]} \mathbf{a}_s^{(1)} \otimes \dots \otimes \mathbf{a}_s^{(K)}$ .
11: end for
```

Output: Estimated signal tensor $\hat{\Theta} = \frac{1}{2H+1} \sum_{\pi \in \mathcal{H}} \text{sgn}(\mathcal{Z}_\pi)$.

NIPS dataset

1998:

- ▶ Kernel PCA and De-Noising in Feature Spaces. Sebastian Mika, **B. Schölkopf**, **A. Smola**, **K. Müller**, M. Scholz, **G. Rätsch**
- ▶ Shrinking the Tube: A New **Support Vector Regression Algorithm**. **B. Schölkopf**, P. Bartlett, **A. Smola**, R. C. Williamson
- ▶ Semiparametric **Support Vector** and Linear Programming Machines. **A. Smola**, T-T. Frieb, **B. Schölkopf**
- ▶ Regularizing AdaBoost. **G. Rätsch**, T. Onoda, **K. Müller**

1999:

- ▶ v-Arc: Ensemble Learning in the Presence of Outliers. **G. Rätsch**, **B. Schölkopf**, **A. Smola**, **K. Müller**, T. Onoda, S. Mika
- ▶ Invariant Feature Extraction and Classification in Kernel Spaces. S. Mika, **G. Rätsch**, J. Weston, **B. Schölkopf**, **A. Smola**, K Müller
- ▶ **Support Vector Method** for Novelty Detection. **B. Schölkopf**, R. C. Williamson, **A. Smola**, J. S. Taylor, J. Platt.

1996:

- ▶ A Variational Principle for **Model**-based Morphing. L. Saul, **M. Jordan**
- ▶ Hidden Markov Decision Trees. **M. Jordan**, Z. Ghahramani, L. Saul
- ▶ Triangulation by Continuous Embedding. M. Meila, **M. Jordan**
- ▶ Recursive Algorithms for Approximating Probabilities in Graphical **Models**. T. Jaakkola, **M. Jordan**

[Back to NIPS analysis](#)

References

- ▶ Oseledets, Ivan V. Tensor-train decomposition. SIAM Journal on Scientific Computing 33.5 (2011): 2295-2317.
- ▶ M. Wang and Y. Zeng. Multiway clustering via tensor block models. Advances in Neural Information Processing Systems 32 (NeurIPS), 715-725, 2019
- ▶ R. Han, Y. Luo, M. Wang, and A. R. Zhang. Exact clustering in tensor block model: Statistical optimality and computational limit. Under review, 2020.