Beyond matrices: Statistical foundations and algorithms for tensor learning

Miaoyan Wang

Department of Statistics
University of Wisconsin-Madison

APS March Meeting
March 2022

Session aim

Exploratory Data Analysis (EDA):

- dimension reduction, clustering, visualization.

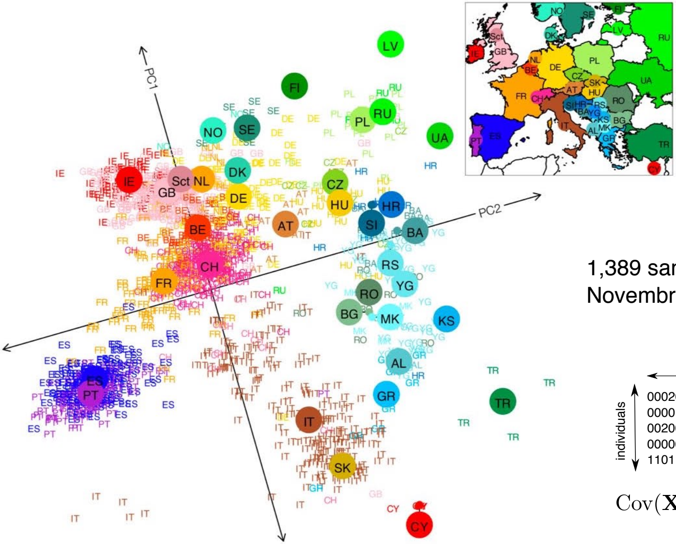
Statistical Machine Learning:

- tensor decomposition, high-dimensional statistics.

Application to Domain Sciences:

- gene expression analysis, population structure, neuroimaging.

A successful story: PCA of Europeans



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

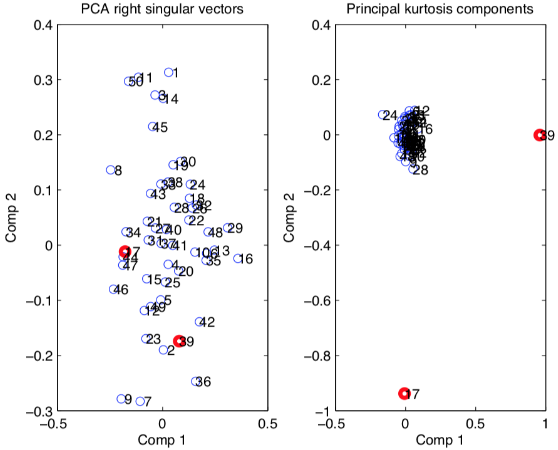
SNPs

individuals

00020110000011111000000...
00001100000012011000000...
00200111012001010011011 1...
00000000011121010010111 0...
11011011101111012000100 1...

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

Matrix methods are powerful, however...



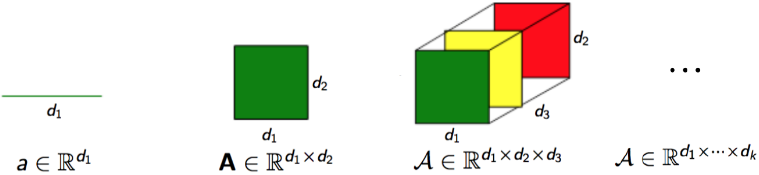
All Gaussian except points 17 and 39.

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

Figure credit: Jason Morton and Lek-Heng Lim (2009, 2015 Rasmus Bro).

What is a tensor?

- Tensors are generalizations of vectors and matrices:



- An order-k tensor $\mathcal{A} = [\![a_{i_1 \dots i_k}]\!] \in \mathbb{F}^{d_1 \times \cdots \times d_k}$ is a hypermatrix with dimensions $(d_1, \dots, d_k)$ and entries $a_{i_1 \dots i_k} \in \mathbb{F}$.
- This talk will focus on $\mathbb{F} = \mathbb{R}$ or $\{0, 1\}$.
- We focus on tensor of order 3 or greater, also known as higher-order tensors.

Tensors in statistical modeling

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

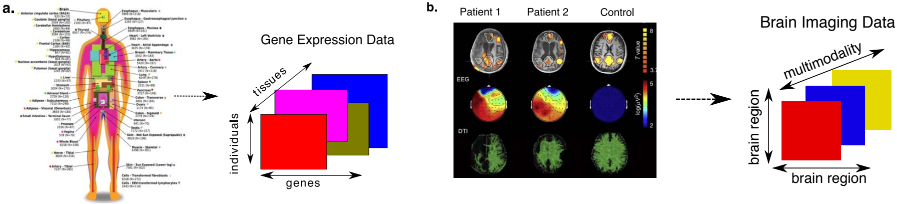
- Longitudinal social network data $\{\boldsymbol{Y}_t : t = 1, ..., n\}$
- Spatio-temporal transcriptome data
- Joint probability table of a set of variables $\mathbb{P}(X_1, X_2, X_3)$
- Higher-order moments in topic models
- Markov models for the phylogenetic tree $K_{1,3}$

Anandkumar 2014, Hoff 2015, Montanari-Richard 2014

M. Yuan et al 2017, Mossel et al 2004, McCullagh 1987

Tensors in science

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



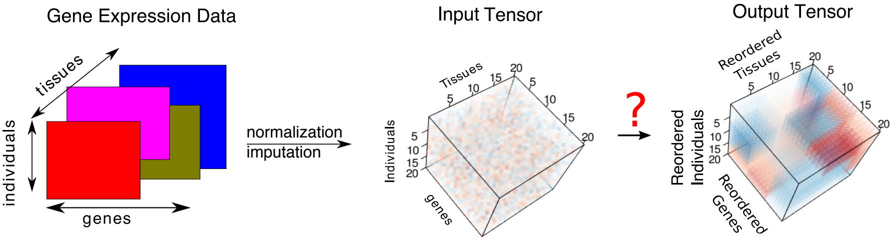
Dimension reduction of tensor data

We seek to identify interpretable low-dimensional representations from noisy, high-dimensional, higher-order tensor data.

Why study tensors?

Tensors provide a rich source of

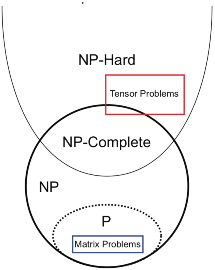
- fundamental problems in data science.
- new tools for long-standing questions.
- huge potentials for new applications.



My research

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
- Block-structure
- smooth, ...

Breaking previous limits

My group is developing a framework of statistical models, efficient algorithms, and fundamental theory to analyze large-dimension tensor data.

Talk outline

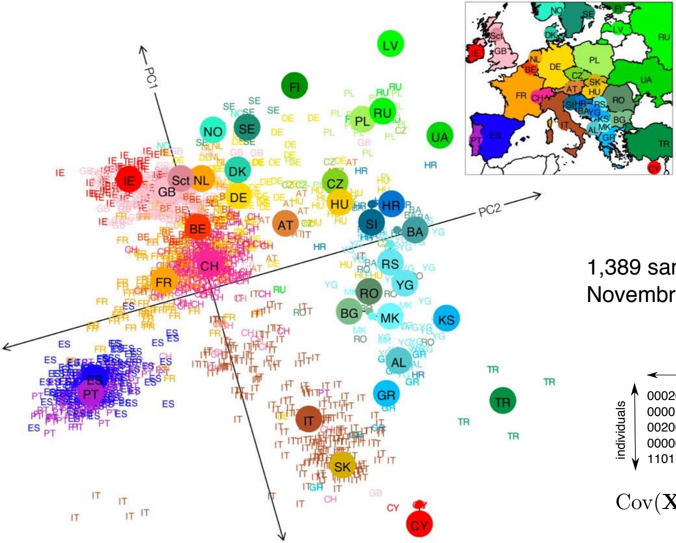
Prohibitive Computational Complexity

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

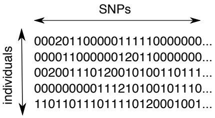
Topics I will address:

- Warm-up: PCA and matrix spectral methods
- Beyond matrices: tensor spectral methods
  - Orthogonally decomposable tensors (continuous, parametric structure)
  - Stochastic block tensors (discrete, parametric structure)

A successful story: PCA of Europeans



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



00020110000011111000000...
00001100000012011000000...
00200111012001010011011...
00000000011121010010111 0...
11011011101111012000100 1...

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

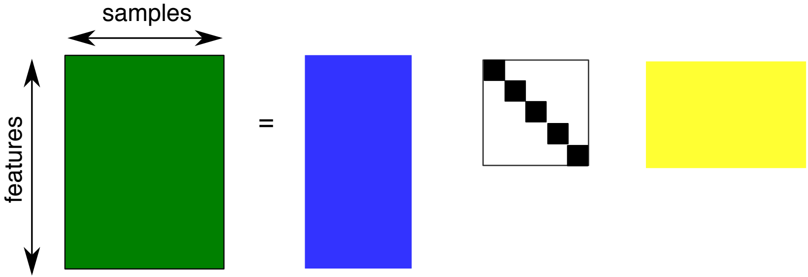
Standard Principal Components Analysis (PCA)

- Create an $n \times m$ matrix, X, of centered and standardized genotype values, and from this, a individual-by-individual correlation matrix:

$$\hat{\Phi} = \frac{1}{m} X X^T$$

- $\hat{\Phi}_{ij}$ is an estimate of the genome-wide correlation between individuals i and j.
- PCA is performed by obtaining the eigen-decomposition $\hat{\Phi}$.
- Top eigenvectors (PCs) are used as surrogates for population structure.

Review: Matrix SVD for biclustering



$$\mathbf{X} = \mathbf{U} \quad \Lambda \quad \mathbf{V}^T$$

$$= \sum_{i=1}^{r} \lambda_i \mathbf{u}_i \mathbf{v}_i^T$$

- Columns of $\mathbf{U}$ describe patterns across samples
- Columns of $\mathbf{V}^T$ describe patterns across features

Outline

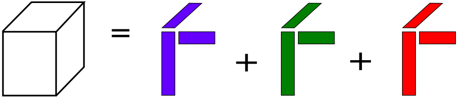
- Warm-up: PCA and matrix spectral methods
- Beyond matrices: tensor spectral methods
    - Orthogonal decomposable tensors (continuous, parametric structure)
    - Stochastic block tensors (discrete, parametric structure)

Tensor rank decomposition

- Tensor analogue of matrix SVD:

$$\mathcal{A} = \sum_{r=1}^{R} \lambda_r \boldsymbol{a}_r \otimes \boldsymbol{b}_r \otimes \boldsymbol{c}_r,$$

  where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_R > 0$, $\|\boldsymbol{a}_r\|_2 = \|\boldsymbol{b}_r\|_2 = \|\boldsymbol{c}_r\|_2 = 1$.
- Minimal R is called the rank of tensor.
- Example: an order-3 rank-3 symmetric tensor.

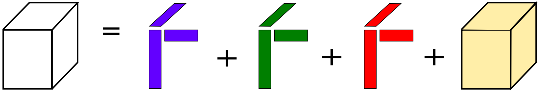


$$\mathcal{A} = \lambda_1 \mathbf{u}_1^{\otimes 3} + \lambda_2 \mathbf{u}_2^{\otimes 3} + \lambda_3 \mathbf{u}_3^{\otimes 3}$$

Illness of low-rank approximation

- A noisy rank-R tensor:

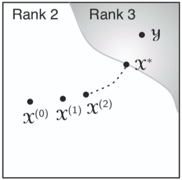
$$\underbrace{\mathcal{Y}}_{\text{observation}} = \underbrace{\sum_{r=1}^{R} \lambda_r \boldsymbol{a}_r \otimes \boldsymbol{b}_r \otimes \boldsymbol{c}_r}_{\text{unknown signal } \mathcal{A}} + \underbrace{\mathcal{E}}_{\text{noise}},$$



- Maximizing the likelihood under the i.i.d. Gaussian error model is equivalent to minimizing the squared loss:

$$\hat{\mathcal{A}} = \underset{\text{rank}(\mathcal{A}) \leq R}{\operatorname{argmin}} \|\mathcal{Y} - \mathcal{A}\|_F^2$$

- ill-posed problem. Optimizer may not exist!



[Silva and Lim 2008; Kolda and Bader 2009] example

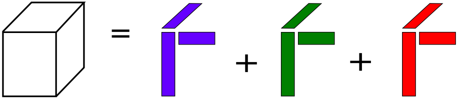
SOD tensors

- Solution: impose manageable structure to the signal tensor.
- A tensor $\mathcal{A}$ is called symmetric and orthogonally decomposable (SOD) if

$$\mathcal{A} = \sum_{i=1}^{r} \lambda_i \mathbf{u}_i^{\otimes k},$$

where $\{\mathbf{u}_i\}$ are orthonormal vectors in $\mathbb{R}^d$ and $\{\lambda_i\}$ are non-zero scalars.

- For example, $k = 3$ and $r = 3$:



$$\mathcal{A} = \lambda_1 \mathbf{u}_1^{\otimes 3} + \lambda_2 \mathbf{u}_2^{\otimes 3} + \lambda_3 \mathbf{u}_3^{\otimes 3}$$

- Is every symmetric tensor orthogonally decomposable?

example

- Fortunately, uniqueness comes for free, under the existence assumption.

Example

Orthogonal decomposition does not exist for the following 2-by-2-by-2 symmetric tensor:

$$\mathcal{A}(:,:,1) = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix},$$

$$\mathcal{A}(:,:,2) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

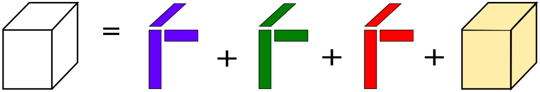
- $\mathcal{A} \in \mathbb{R}^{2\times 2\times 2}$.
- Rank$(\mathcal{A}) = 3$.
- Rank $>$ dimension. Fundamentally different from matrix spectral method.

Specially-structured tensors

Orthogonally decomposable tensors $\subsetneq$ General tensors

Tensor decomposition

- Nearly SOD tensors:
$$\tilde{A} = \sum_{i=1}^{r} \lambda_i \mathbf{u}_i^{\otimes k} + \mathcal{E},$$
where $\mathcal{E} \in \mathbb{R}^{d\times\cdots\times d}$ is a symmetric but otherwise arbitrary tensor with $\|\mathcal{E}\|_2 \leq \varepsilon$.
- For example, $k = 3$ and $r = 3$:



$$\tilde{A} = \lambda_1 \mathbf{u}_1^{\otimes 3} + \lambda_2 \mathbf{u}_2^{\otimes 3} + \lambda_3 \mathbf{u}_3^{\otimes 3} + \mathcal{E}$$

Key question

Can we recover the vectors $\{\mathbf{u}_i\}$ from the noisy observation $\tilde{A}$?

Our results

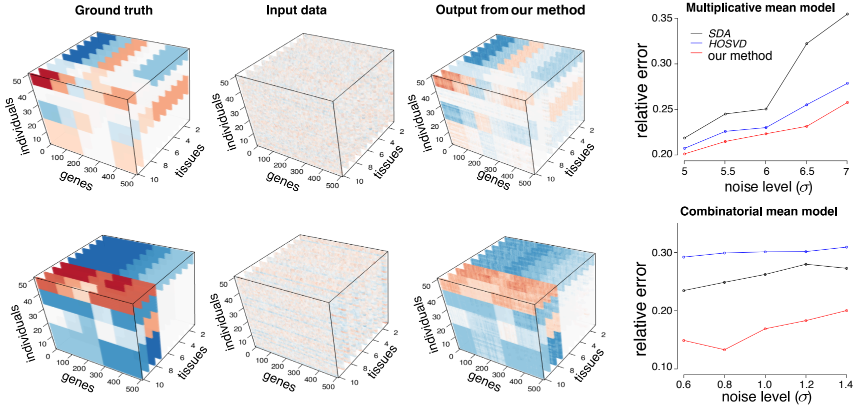
- The error bound in tensor decomposition does not depend on the eigenvalue gap ⇒ more stable than matrix decomposition.

Method	Noise threshold $(\varepsilon/\lvert\lambda\rvert_{\min} \leq)$	Recovery accuracy $(\lVert\hat{\mathbf{u}}_i - \mathbf{u}_i\rVert_2 \leq)$
Power iteration (Anandkumar et al, 2014)	$O(d^{-1})$ for order 3	$\frac{8\varepsilon}{\lambda_i}$
Joint diagonalization (Kuleshov et al, 2015)	–	$\frac{2\varepsilon\sqrt{\lVert\boldsymbol{\lambda}\rVert_1\lambda_{\max}}}{\lambda_i^2} + o(\varepsilon)$
Two-mode HOSVD W. and Song, 2018	$\boldsymbol{\tilde{O}(d^{-1/2})}$ **for order 3** $\tilde{O}(d^{-(k-2)/2})$ for order k	$\boldsymbol{\frac{2\varepsilon}{\lambda_i}} + o(\boldsymbol{\varepsilon})$

- Our perturbation bounds can be viewed as a higher-order analogue of Davis-Kahan theorem.

Performance comparison

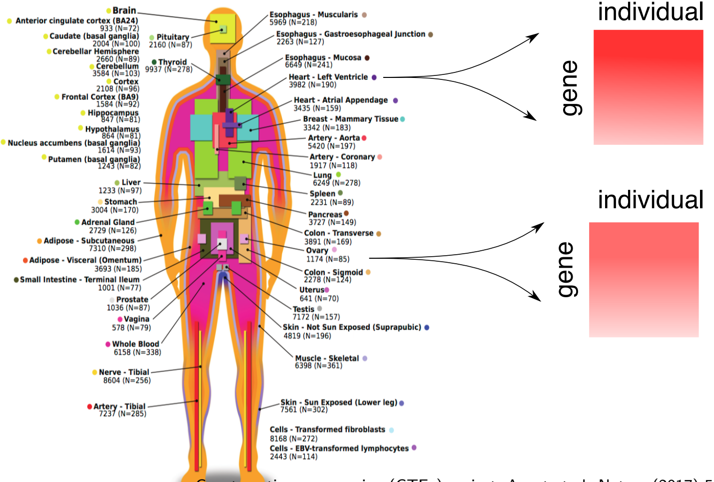
Our algorithm recovers low-rank tensor signals with higher accuracy compared to existing methods.



SDA: sparse decomposition of array (Hore et al. Nat. Gen. 2016);
HOSVD: higher-order singular value decomposition (Omberg et al. PNAS. 2007).

Multi-tissue gene expression analysis

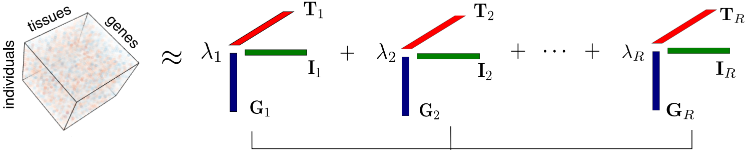
- Central dogma of genetics: DNA $\xrightarrow{\text{transcription}}$ RNA $\xrightarrow{\text{translation}}$ protein.
- GTEx RNA-seq data: expression profiles ($\sim 25,000$ genes) measured from 544 individuals across 53 tissues.

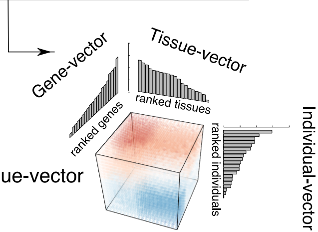


Genotype-tissue expression (GTEx) project, Aguet et al. Nature (2017) 550, 204-213.

Multi-tissue gene expression analysis

We modified the earlier tensor decomposition for three-way clustering of multi-tissue multi-individual gene expression data.





$$\mathcal{Y} \approx \sum_{r=1}^{R} \lambda_r \mathbf{T}_r \otimes \mathbf{G}_r \otimes \mathbf{I}_r$$

Each gene expression moduel consists of

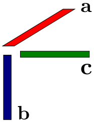
λ_r : Singular-value

$\mathbf{T}_r \geq 0$ entriwise : Tissue-vector

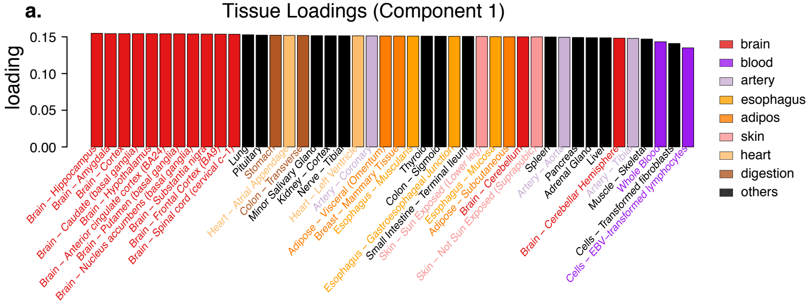
$\mathbf{G}_r$: Gene-vector

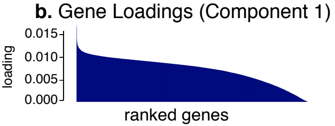
$\mathbf{I}_r$: Individual-vector

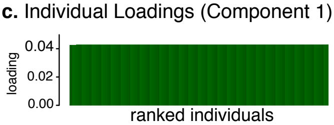
Component I: shared, global expression



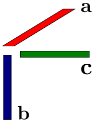
- Both tissue- and individual-loadings are essentially flat ⇒ this component captures global expression common to all samples.


a. Tissue Loadings (Component 1)

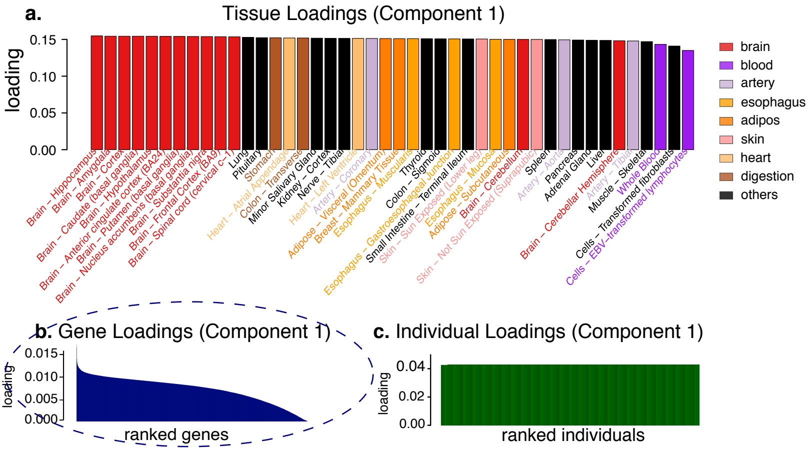

b. Gene Loadings (Component 1)


c. Individual Loadings (Component 1)

Component I: shared, global expression

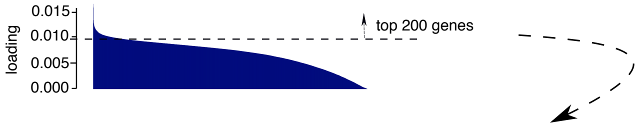


- Both tissue- and individual-loadings are essentially flat ⇒ this component captures global expression common to all samples.

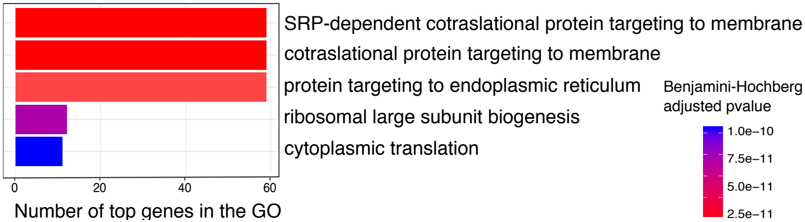


Component I: shared, global expression

- Top genes are mainly mitochondrial genes (15/20 top genes); other non-mitochondrial genes include *ACTB*, *EEF1A1*, and *EEF2*.

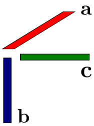


b. Gene Loadings (Component 1)

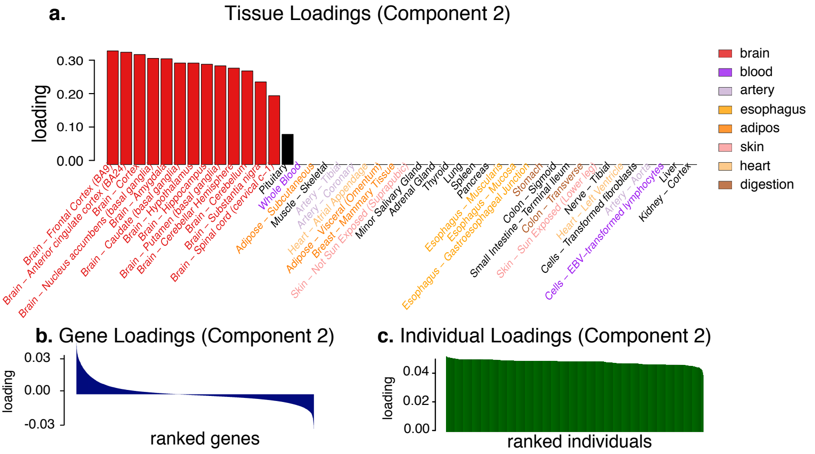


d. Enriched gene ontologies (GOs) among top genes (Component 1)

Component II: brain tissues



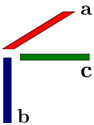
- Tissue vector clearly separates brain tissues from non-brain tissues.


a. Tissue Loadings (Component 2)

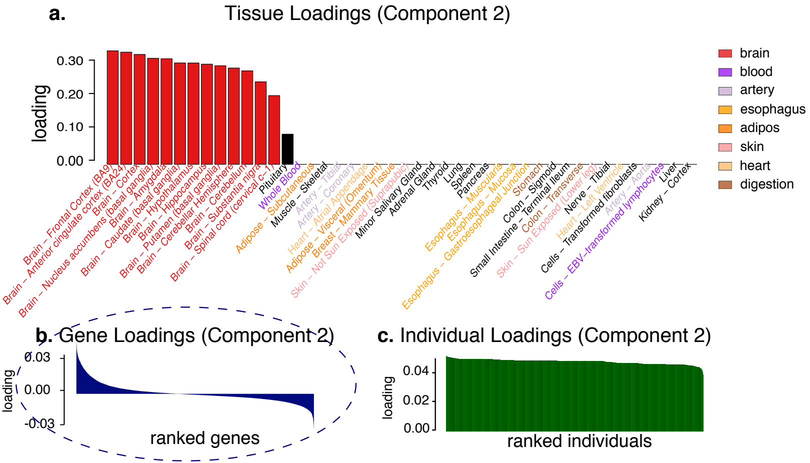
b. Gene Loadings (Component 2)

c. Individual Loadings (Component 2)

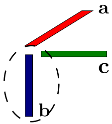
Component II: brain tissues



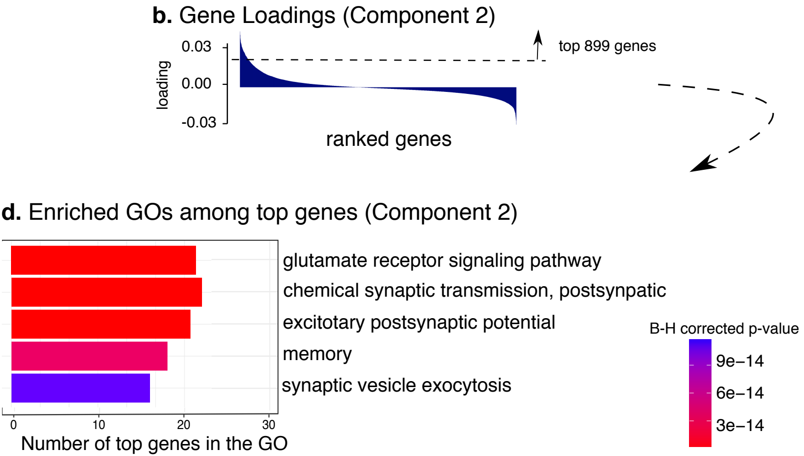
- Tissue vector clearly separates brain tissues from non-brain tissues.



Component II: brain tissues



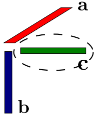
- Enriched GOs are mostly related to glutamate receptor signaling pathway, chemical synaptic transmission, and memory.



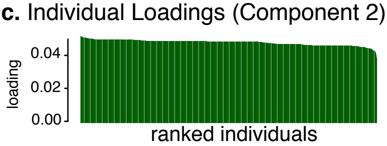
b. Gene Loadings (Component 2)

d. Enriched GOs among top genes (Component 2)

Component II: brain tissues



- Age explains more variation (24.4%) than gender (0.3%) or ethnicity (4.3%).


c. Individual Loadings (Component 2)

d.

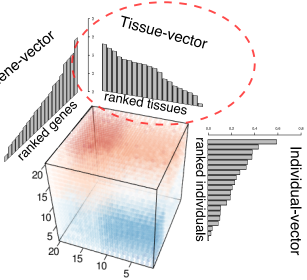
Linear model: individual-vector 2 $\sim$ age+gender+ethnicity

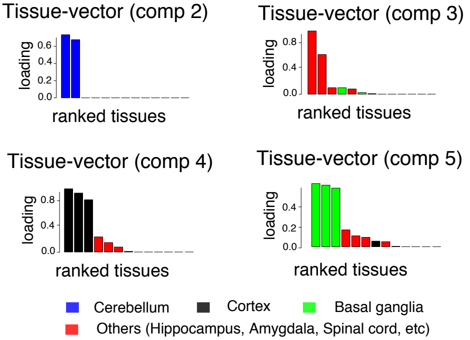
Attribute	age	gender	ethnicity
Coefficient estimate	-7.3e-04	-8.6e-05	3.1e-04
P-value	**<2e-16**	0.12	2.3e-08
Proportion of variance explained	**24.4%**	0.3%	4.3%

Gene expression in the brain

- We analyze 13 brain tissues (“subtensor”) and apply our tensor method to this subtensor.
- Expression modules are spatially restricted to specific brain regions, such as the two cerebellum tissues, three cortex tissues, and three basal ganglia tissues.

Recall: i^{th} tensor component, i = 1, 2, ...





Gene expression in the brain

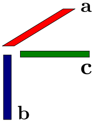
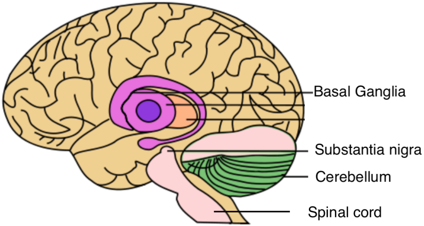


Table 1: Three-way clustering analysis of gene expression in the brain

Tissue	Gene	Individual		
enriched region	enriched ontology	variance explained by		
		age	gender	ethnicity
cerebellum	dorsal spinal cord development	0.0%	8.0%	0.2%
cortex	behavior defense response	16.7%	0.6%	1.4%
basal ganglia	forebrain generation of neurons	1.3%	0.8%	1.7%
others	embryonic skeletal system morphogenesis	10.5%	0.7%	5.2%

Red number indicates p-value < 0.001.

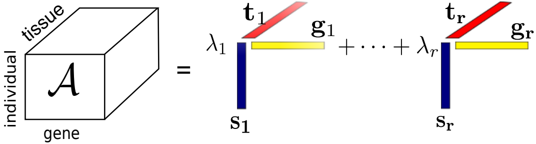


Outline

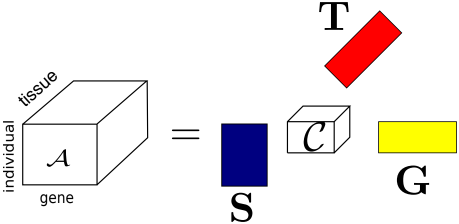
- Warm-up: PCA and matrix spectral methods
- Beyond matrices: tensor spectral methods
  - Orthogonal decomposable tensors (continuous, parametric structure)
  - Stochastic block tensors (discrete, parametric structure)

Various notions of low-rankness

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



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



- Others: tensor train [Oseledet '11], shape-constrained decomposition [Lusch, Chi et al '19], etc.

We will focus on block tensors in this part

- Block structure is an art:

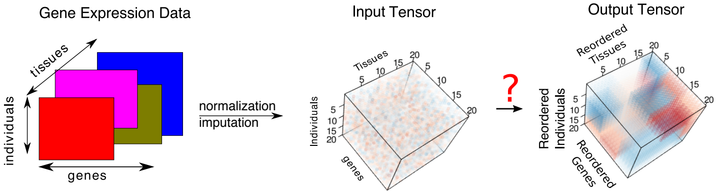
Piet Mondrian (Dutch, 1871-1944)

Composition No. 1, The Art Institute of Chicago

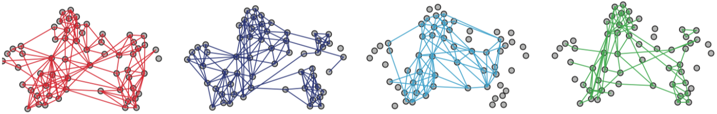
Block structure implies low-rankness, but directly applying low-rank estimation to a block tensor yields an inferior estimator.

Why block tensors?

- Many data tensors are expected to exhibit underlying block structure.
- Examples of block tensors:
  - multiway clustering



  - multi-layer community detection (Figure source: Li, Xu, et al, 2019)

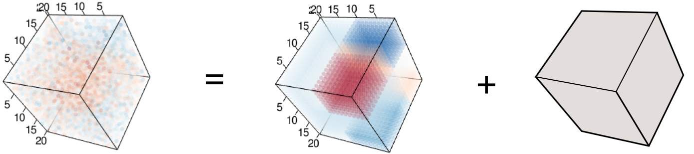


(a) Work. (b) Lunch. (c) Facebook. (d) Friend.

Tensor block model

- Consider an order-K data tensor $\mathcal{Y} \in \mathbb{R}^{d\times\cdots\times d}$. Assume symmetry for simplicity.
- Unknown block assignment function $z\colon [d] \to [r]$.
- We model the data tensor by an unknown mean tensor $\mathcal{S} \in \mathbb{R}^{r\times\cdots\times r}$ and unknown block assignments z; that is, for all $(i_1,\ldots,i_K) \in [d]^K$,

$$\mathcal{Y}(i_1,\ldots,i_K) = \mathcal{S}(z(i_1),\ldots,z(i_K)) + \mathcal{E}(i_1,\ldots,i_K),$$



Observation

signal

sub-Guassian noise

- (Clustering) How to recover the block assignment function z?
- (Denoising) How to estimate the underlying mean tensor $\mathcal{S}$?

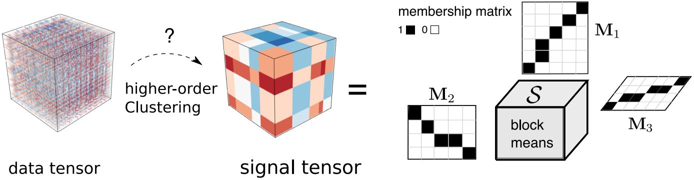
Equivalent representation

- Tensor block model can be viewed as a super sparse, discrete Tucker model:

$$\mathcal{Y} = \underbrace{\mathcal{S} \times_1 \boldsymbol{M}_1 \times_2 \cdots \times_K \boldsymbol{M}_K}_{=:\ \text{unknown signal tensor}} + \mathcal{E},$$

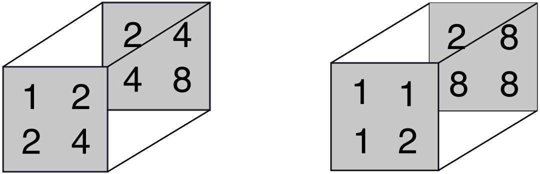
  where

  - $\mathcal{Y}$: the data tensor.
  - $\mathcal{S}$: a core tensor consisting of block means.
  - $\boldsymbol{M}_k \in \{0,1\}^{d\times r}$: membership matrix indicating the block allocation along mode k.
  - $\mathcal{E}$: a noise tensor with i.i.d. $N(0,\sigma^2)$ entries



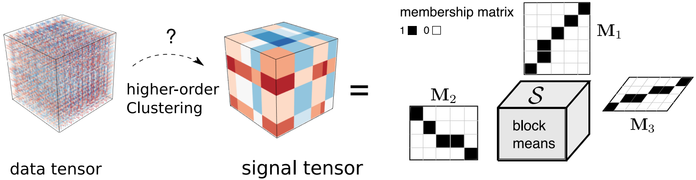
Identifiability in noiseless case

- Intuition: $({\color{red}r}, r, \ldots, r)$ blocks should not be merged to $({\color{red}r-1}, r, \ldots, r)$ blocks.
- Matrices $(K = 2)$: core matrix $\mathcal{S}$ has no two identical rows/columns.
- General K: unfolded core tensor $\text{Unfold}_k(\mathcal{S})$ has no two identical rows.
- Block gap condition is a weaker assumption than eigen gap condition.



Stability in the noisy case

- Given a noisy tensor data $\mathcal{Y}$ from the model
$$\mathcal{Y}(i_1,\ldots,i_K)=\mathcal{S}(z(i_1),\ldots,z(i_K))+\mathcal{E}(i_1,\ldots,i_K).$$
  Goal: recover block assignments $z\colon [d]\to[r]$ and block means $\mathcal{S}$.



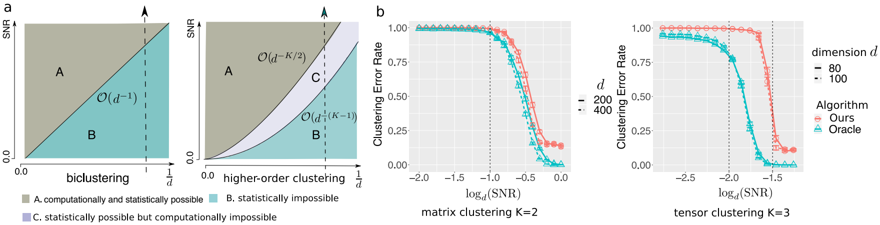
A key quantity for the hardness (WZ'19; HLWZ'21)

The statistical and computational hardness of inference is completely determined by the following signal-to-noise ratio (SNR)
$$\text{SNR}=\min_{k\in[K]}\Delta_k^2/\sigma^2.$$

Statistical-computational gaps for higher-order tensors

Tensor vs. matrices (Han-Luo-W.-Zhang'21)

A statistical-computational gap emerges for tensors of order 3 or greater.



- SNR $\gg d^{-K/2}$: our polynomial algorithm achieves exact clustering;
- SNR $\ll d^{-(K-1)}$: no algorithm can achieve exact clustering;
- When $d^{-(K-1)} \ll$ SNR $\ll d^{-K/2}$, no polynomial-time algorithm exists for exact clustering under planted clique detection hardness conjecture.

Polynomial algorithms for higher-order clustering

- In moderate-to-high SNR regime, a well-designed local solution achieves similar statistical accuracy as the global MLE.

Statistical and computational accuracy (Han-Luo-W-Zhang'21)

The output from Algorithm 1 satisfies

$$\text{Error at step } (t+1) \lesssim \underbrace{\sigma^2 \exp\left(-\frac{d^{K-1}}{r^{K-1}}\mathsf{SNR}\right)}_{\text{statistical error}} + \underbrace{0.5^t \text{ Error at step } 0}_{\text{computational error}}.$$

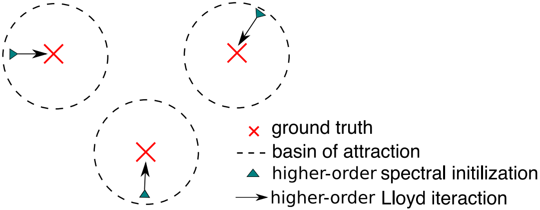
Algorithm 1 Multiway clustering based on tensor block model

```
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 $\hat{\boldsymbol{M}}_k \in \{0,1\}^{r\times d}$.
Initialize the marginal clustering by higher-order spectral clustering.
for $t = 1,2,\ldots$ do
    Update the core tensor $\hat{\mathcal{S}}$ using the sample averages within each multi-way block.
    Update each membership matrix $\hat{\boldsymbol{M}}_k$ by a small nearest neighbor search over $r$ discrete points.
end for
```

Take-away

A large variety of structured tensor problems exhibit the two-component error and certain benign non-convex optimization landscape, e.g.,

- tensor SVD
- tensor completion
- sparse tensor models
- tensor regression



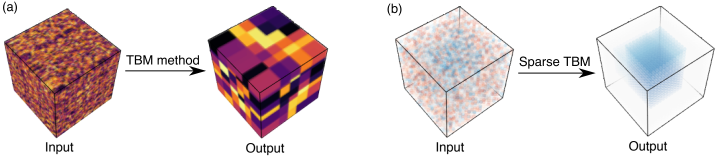
Open problems

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

Data analysis

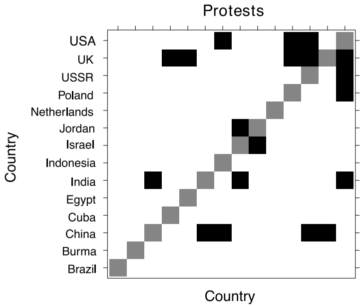
- **Nations** (Nickel et al., 2011). Binary tensor consisting of 56 political relations among multiple countries.

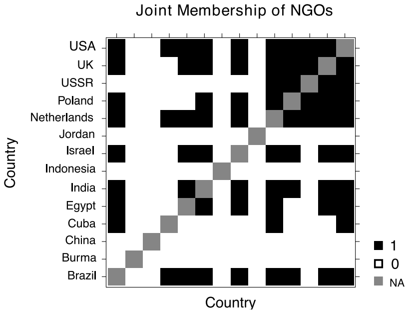
- **Multi-tissue gene expression** (W., Fischer, and Song, 2019). Continuous tensor consisting of expression values from 53 tissues, $\sim$ 500 individuals, and 10K genes.



Political relationships in 1950-1965

- 56 types of relationships among 14 countries between 1950-1965.
- Nations data can be organized as a tensor $\mathcal{Y} \in \{0, 1\}^{14 \times 14 \times 56}$. A tensor entry is 1 if the relation holds between two countries.
- Relations: “sends tourists to”, “exports books to”, “joint membership of NGOs”, “conferences”, etc.
- Countries: USSR, Poland, China, UK, Brazil, etc.





Real data analysis

Table: Comparison of goodness-of-fit in the real data analysis

Dataset	TBM	TBM-sparse	CP+k-means	Tucker+k-means	CoTeC [Jegelka et al]
Nations	0.439	0.433	0.324	0.253	0.419
Brain expression	0.856	0.855	0.576	0.434	0.849

Nations data:

- The 14 countries are naturally partitioned into 5 clusters:
  - Cluster 1 (neutral): Brazil, Egypt, India, Israel, Netherlands.
  - Cluster 2 (neutral): Burma, Indonesia, Jordan.
  - Cluster 3 (eastern): China, Cuba, Poland, USSR.
  - Two western blocs: {USA} and {UK}.
- Two major relationship clusters involving neutral and western blocs:
  - Exports-related activities {reltreaties, book translations, relexports...}
  - NGO-related activities {relintergovorgs, relngo, ngoorgs3,...}.

Summary

References:

- J. Hu and **M. Wang**. Multiway spherical clustering via degree-corrected tensor block models. AISTATS 2021. arxiv. 2201.07401
- R. Han, Y. Luo, **M. Wang**, and A. R. Zhang. Exact clustering in tensor block model: Statistical optimality and computational limit. arxiv. 2012.09996.
- **M. Wang**, J. Fischer, and Y. S. Song. Three-way Clustering of Multi-tissue Multi-individual Gene Expression Data Using Semi-nonnegative Tensor Decomposition. Annals of Applied Statistics. Vol. 13, No. 2, 1103-1127, 2019.
- **M. Wang** and Y. S. Song. Tensor Decomposition via Two-Mode Higher-Order SVD (HOSVD). AISTATS 2017.
- **M. Wang**, K. Dao Duc, J. Fischer, and Y. S. Song. Operator Norm Inequalities Between Tensor Unfoldings on the Partition Lattice. Linear Algebra and its Applications, Vol. 520, 44-66, 2017.

Our general strategy is to develop efficient statistical methods for analyzing a broad range of specially-structured tensors that are useful in practice.

Acknowledgment: NSF CAREER DMS-2141865, NSF DMS-1915978, DMS-202323, and NSF EF-2133740.

Non-Existence of MLE

- Rank($\mathcal{A}_n$)$= 2$ for all $n = 1, 2, 3, \ldots$.

$$\mathcal{A}_n = n\left(\mathbf{u} + \frac{1}{n}\mathbf{v}\right) \otimes \left(\mathbf{u} + \frac{1}{n}\mathbf{v}\right) \otimes \left(\mathbf{u} + \frac{1}{n}\mathbf{v}\right) - n\mathbf{u} \otimes \mathbf{u} \otimes \mathbf{u}.$$

- However, Rank($\lim_{n\to\infty} \mathcal{A}_n$)$= 3$.

$$\mathcal{A}_n \to \mathcal{A}^* = \mathbf{u} \otimes \mathbf{u} \otimes \mathbf{v} + \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{u} + \mathbf{v} \otimes \mathbf{u} \otimes \mathbf{u}.$$

back

Non-Existence of MLE

- Rank($\mathcal{A}_n$)$= 2$ for all $n = 1, 2, 3, \ldots$.

$$\mathcal{A}_n = n\left(\mathbf{u} + \frac{1}{n}\mathbf{v}\right) \otimes \left(\mathbf{u} + \frac{1}{n}\mathbf{v}\right) \otimes \left(\mathbf{u} + \frac{1}{n}\mathbf{v}\right) - n\mathbf{u} \otimes \mathbf{u} \otimes \mathbf{u}.$$

- However, Rank($\lim_{n\to\infty} \mathcal{A}_n$)$= 3$.

$$\mathcal{A}_n \to \mathcal{A}^* = \mathbf{u} \otimes \mathbf{u} \otimes \mathbf{v} + \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{u} + \mathbf{v} \otimes \mathbf{u} \otimes \mathbf{u}.$$

- For every signal tensor $\mathcal{X} \in \mathbb{R}^{2\times 2\times 2}$ with Rank($\mathcal{X}$) $= 2$, under Gaussian observation model,

$$\mathbb{P}(\hat{\boldsymbol{X}}^{\text{MLE}} \text{ is not well defined}) > 0.$$

back

Unfolding of an order-k tensor

- **General Unfolding**. The set of all possible unfoldings of an order-k tensor is in one-to-one correspondence with the set $\mathcal{P}_{[k]}$ of all partitions of $[k] = \{1, \ldots, k\}$.
- For $\pi = \{B_1, \ldots, B_\ell\} \in \mathcal{P}_{[k]}$, $\mathrm{Unfold}_\pi(\mathcal{A})$ is obtained by combining the modes in each block B_n into a single mode.

Example. An order-4 tensor $\mathcal{A} = [\![a_{i_1 i_2 i_3 i_4}]\!] \in \mathbb{R}^{2\times2\times2\times2}$ with $a_{i_1 i_2 i_3 i_4} = \begin{cases} 1 & \text{if } i_1 = i_2 = i_3 = i_4 \\ 0 & \text{otherwise} \end{cases}$ can be matricized into

- 2×2^3 matrix: $\text{Unfold}_{[1|234]}(\mathcal{A}) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$.
- $2^2 \times 2^2$ matrix: $\text{Unfold}_{[12|34]}(\mathcal{A}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

back