Multiway clustering via tensor block models

Miaoyan Wang Yuchen Zeng
miaoyan.wang@wisc.edu yzeng58@wisc.edu

Department of Statistics
University of Wisconsin-Madison

Neural information Processing Systems, 2019

October 26, 2019

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

In many applications, the data tensors are often expected to have underlying block structure modulo some unknown reordering along each of its modes.

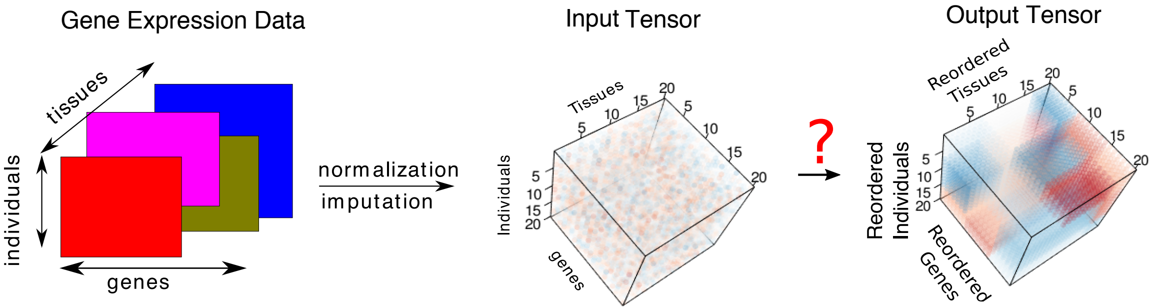

Figure: One application of high-order tensors: *Gene expression* data. [Wang et al, 2017]

State-of-art

Note: a block structure automatically implies low-rankness.

- Existing multiway clustering methods:
  - typically take a two-step procedure; [Kolda et al, 2008][Wang et al, 2015][Hore et al, 2016][Wang et al, 2018]
- Our approach:
  - takes a single shot to perform estimation and clustering simultaneously;

Suppose that the k-th mode of the tensor consists of R_k clusters, where $k \in [K]$.

Notations:

- $\mathcal{Y} = [\![y_{i_1,\ldots,i_K}]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_K}$: an order-$K$, $(d_1, \ldots, d_K)$-dimensional data tensor.
- $\mathcal{C} = [\![c_{r_1,\ldots,r_K}]\!] \in \mathbb{R}^{R_1 \times \cdots \times R_K}$: a core tensor consisting of block means.
- $\boldsymbol{M}_k \in \{0,1\}^{d_k \times R_k}$: a membership matrix indicating the block allocations along mode k for $k \in [K]$.
- $\mathcal{E} = [\![\varepsilon_{i_1,\ldots,i_K}]\!]$: the noise tensor consisting of i.i.d mean-zero sub-Gaussian entries.

Tensor form:

$$\mathcal{Y} = \underbrace{\mathcal{C} \times_1 \boldsymbol{M}_1 \times_2 \cdots \times_K \boldsymbol{M}_K}_{\text{defined as } \Theta} + \mathcal{E}$$

Entry-wise form:

$$y_{i_1,\ldots,i_K} = c_{r_1,\ldots,r_K} + \varepsilon_{i_1,\ldots,i_K}, \quad \text{for } (i_1, \ldots, i_K) \in [d_1] \times \cdots \times [d_K],$$

Assume the clustering size is known.

- $\mathcal{P}$: the parameter space consists of all tensors of block structure.

$$\hat{\Theta} = \underset{\Theta \in \mathcal{P}}{\arg\min} \left\{ \|\mathcal{Y} - \Theta\|_F^2 \right\}$$

Covergence rate:

Method	Tucker [Zhang et al, 2018]	CoCo [Chi et al, 2018]	Tensor block model (this paper)
Recovery error (MSE)	dR	d^{K-1}	$d \log R$

Table: Comparison of various tensor decomposition methods when $d_1 = \cdots = d_K = d,\ R_1 = \cdots = R_K = R$.

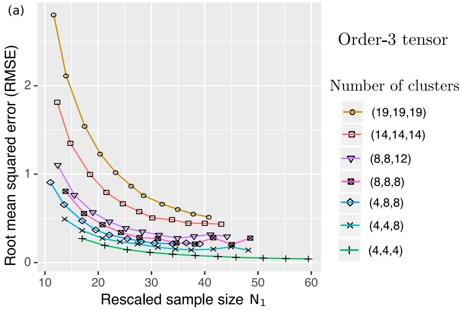

Figure: Estimation error for order-3 block tensors with Gaussian noise.

Conclusion: the empirical mean squared error decreases at a rate of reciprocal of rescaled sample size. This is consistent with our theoretical result.

More information

- Read our paper: `https://arxiv.org/abs/1906.03807`;
- Install our tensor clustering package (tensorsparse): `https://cran.r-project.org/web/packages/tensorsparse/index.html`.