Multi-Slice Clustering (MSC) for 3-order Tensors

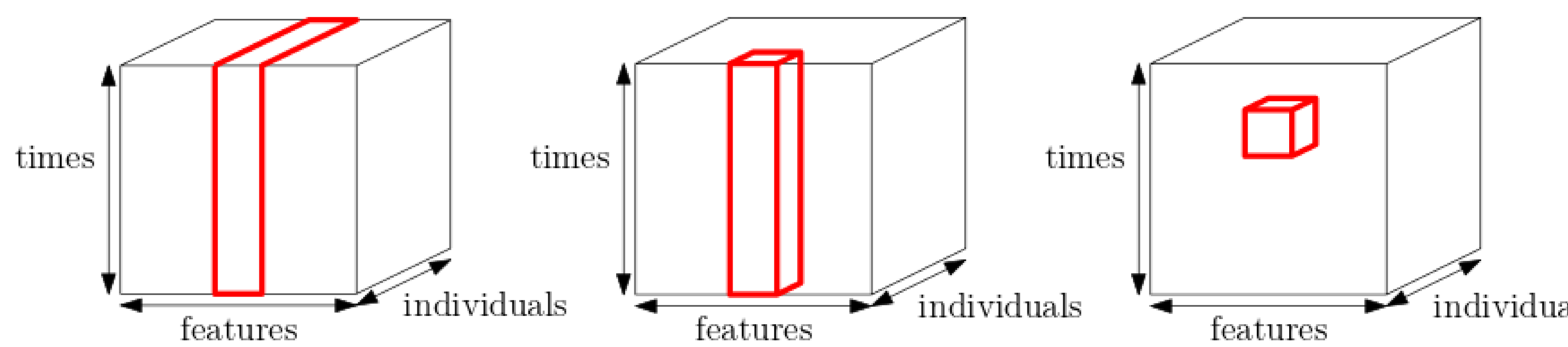
Dina Faneva Andriantsiory [1,2] Joseph Ben Geloun [1,2] Mustapha Lebbah [1,2]

[1]Université Sorbone Paris Nord [2]Laboratoire d'Informatique de Paris Nord (LIPN), UMR CNRS 7030

Motivations

The clustering algorithms

- depend on the number of clusters or the cluster size as data input. Nevertheless, for real data, these inputs might be very difficult to assess from the outset
- only guarantee a clear distinction between a given cluster and the rest of the data
- does not guarantee the strong similarity withing a cluster
- should make sure about separation and strong correlation within the cluster.



(a) Slice clustering (b) Tensor biclustering (c) Triclustering

Figure 1. 3-order tensors and clustering.

Main goal of the MSC algorithm

The MSC algorithm

- solves the parameter issue
- determines and gathers the indices of the matrix slices that are similar in each dimension of the tensor.

Theoritical guarantee

- Let $\mathcal{T} \approx \mathcal{X} + \mathcal{Z}$ where $\mathcal{X}$ is the signal tensor and $\mathcal{Z}$ is a noise tensor.
- **Assumption:** $\mathcal{X}$ is a rank one tensor $\mathcal{X} \approx \gamma \mathbf{w} \otimes \mathbf{u} \otimes \mathbf{v}$, and the entries of $\mathcal{Z}$ are i.i.d and have a standard normal distribution.

Theroem 1 :

Let $l = |J_1|$, assume that $\sqrt{\epsilon} \leq \frac{1}{m_1 - l}$. $\forall i, n \in J_1$, for $\lambda = \Omega(\mu)$, there is a constant $c_1 > 0$ such that

$$|d_i - d_n| \leq l\frac{\epsilon}{2} + \sqrt{\log(m_1 - l)} \qquad (1)$$

holds with probability at least $1 - e(m_1 - l)^{-c_1}$.

Theorem 2 :

For $i \in \bar{J}_1$, if $\lambda = \Omega(\mu m_1)$,

$$d_i \leq \frac{l}{m_1} + \sqrt{\log(m_1 - l)} \qquad (2)$$

with probability at least $1 - e(m_1 - l)^{-c_1}$ with $c_1 > 0$.

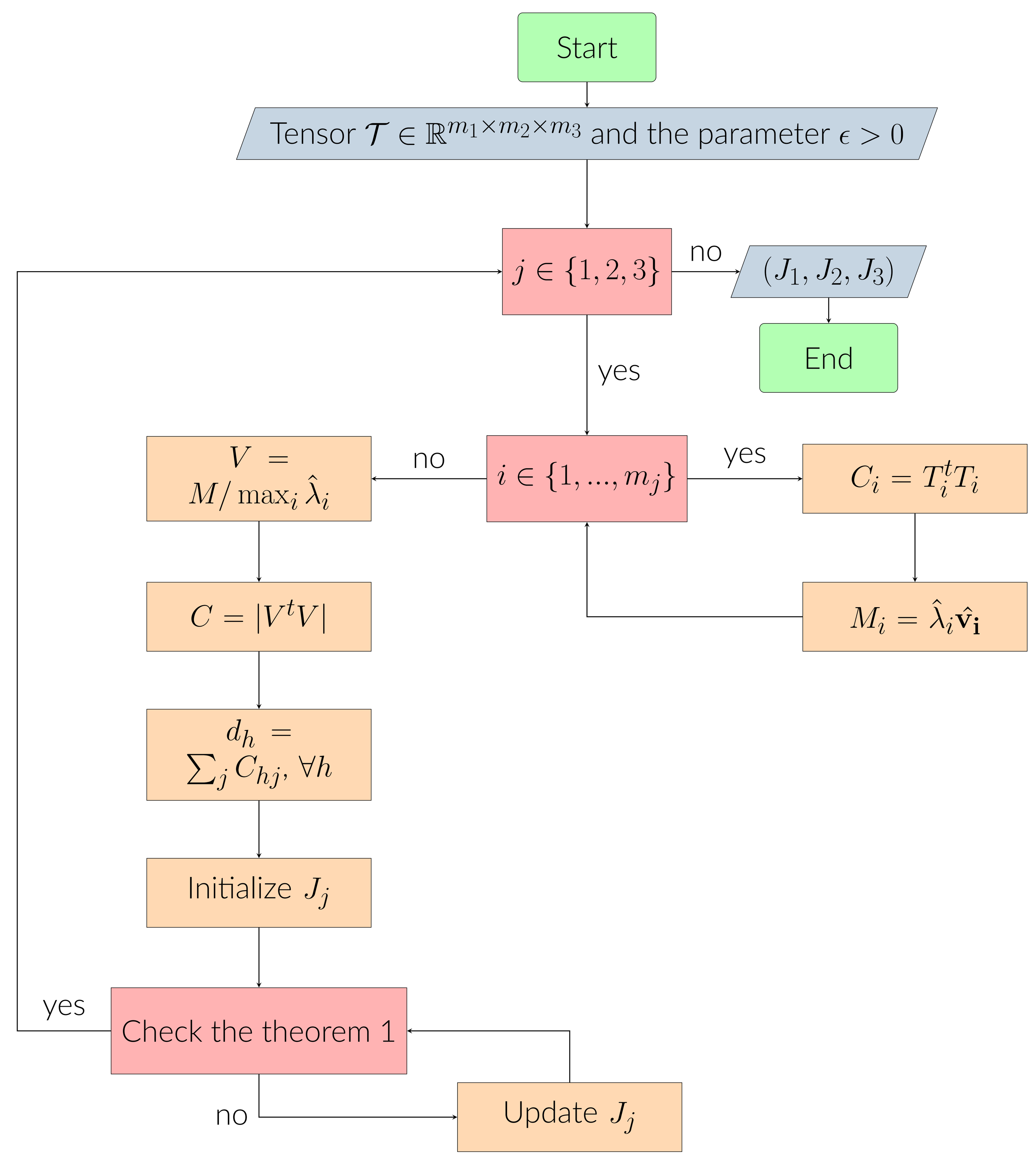
Proposition :

Let $J_1 \subset [m_1]$ the set of all indices of the cluster in the first dimension. Then

$$\text{dist}(d(J_1), d(\bar{J}_1)) \geq l(1 - \frac{\epsilon}{2} - \frac{1}{m_1}) - \sqrt{\log(m_1 - l)}, \qquad (3)$$

with probability at least $1 - e(m_1 - l)^{-c_1}$, with $c_1 > 0$

MSC clustering algorithm

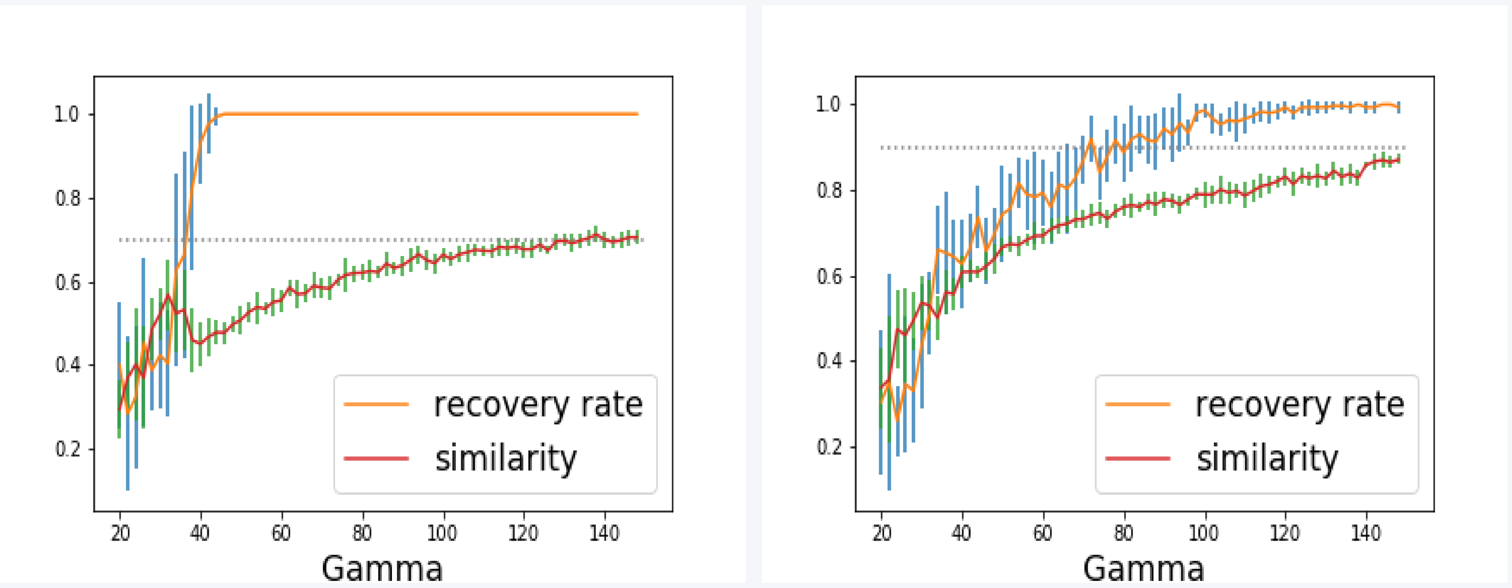


Where:

- $T_i = \mathcal{T}(i, :, :)$ or $\mathcal{T}(:, i, :)$ or $\mathcal{T}(:, i, :)$ for $j = 1, 2$ or 3, respectively.
- M_i is the i-the column of the matrix M.
- $\lambda = \max_i \hat{\lambda}_i$ where $\hat{\lambda}_i$ is the largest eigenvalue of C_i.
- J_1, J_2 and J_3 are the clusters selected from mode-1, mode-2, and mode-3 of the tensor, respectively.

Experiment results

Synthetic data set : Generated according to equation (1), with size $50 \times 50 \times 50$ and cluster size equal to 10.



(a) $\epsilon = 0.005$ (b) $\epsilon = 0.0006$

Figure 2. Recovery rate and similarity measure of tensor biclustering, for $\gamma \in [20, 150]$: (a) with $\epsilon = 0.005$ and (b) with $\epsilon = 0.0006$.

Real data set : We use the flow injection analysis (FIA) dataset, with size 12 (samples)×100 (wavelengths)×89 (times) and we choose $\epsilon = 0.00013$.

	Frobinius distance	fibers correlation
mode-1	3.08616	0.99915
mode-2	0.40307	0.97164
mode-3	1.05795	0.99059

Table 1. Similarity measure within the cluster, with MSE = 0.48360

Evaluation of the cluster quality : We add a random index to the cluster

	frobenius distance	fibers correlation	MSE of triclustering
mode-1	3.17592	0.81419	0.49257
mode-2	3.83777	0.74525	0.51133
mode-3	4.86486	0.81829	1.19725

Table 2. Similarity measure within the cluster.

Perspectives

- Apply MSC algorithm to large tensor size (using parallel and distributed system).
- Generalize MSC algorithm to multiple cluster detection.

References

[1] Dina Faneva Andriantsiory, Joseph Ben Geloun, and Mustapha Lebbah. Multi-slice clustering for 3-order tensor data - supplementary material. *arXiv preprint arXiv:2109.10803*, 2021.

[2] Dina Faneva Andriantsiory, Joseph Ben Geloun, and Mustapha Lebbah. Multi-slice clustering for 3-order tensor. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 173–178. IEEE, 2021.

[3] Soheil Feizi, Hamid Javadi, and David Tse. Tensor biclustering. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30, pages 1311–1320. Curran Associates, Inc., 2017.

[4] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

andriantsiory@lipn.univ-paris13.fr