# Multiway clustering of 3-order tensor via affinity matrix

1. **Summary and Contributions: Briefly summarize the paper and its contributions**

   This work propose a new multiway clustering method for order-3 tensors via affinity matrix clustering on every mode. Numerical experiments are implemented to compare the proposed algorithm with competitive methods.

2. **Strengths: Please describe the strengths of the work according (but not limited) to the following criteria: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the AISTATS community.**

   The cross similarity issue among multiple eigenvectors is the key difficulty to extend the rank-1 affinity matrix clustering to the rank-d case. This is an interesting phenomenon that may also occur in other spectral methods.

3. **Weaknesses: Please describe the limitations of this work according (but not limited) to the following criteria: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the AISTATS community.**

   Based on my understanding, the model, method, and motivation are not clear enough to address the multiway clustering problem, which is the main weakness of this work. The numerical results are not convincing for me. Comparisons with several competitive prior works are missing. Moreover, the proposed method is restricted to order-3 tensors, rather than general tensors with order $\geq 3$, which limits the practical usage of the method in many real-life applications with higher-order tensors.

   See additional comments and relation to prior work for details.

4. **Correctness: Are the method and claims correct? Is the empirical methodology correct?**

   Part of method description and numerical results are not sound for me; see additional comments for details.

5. **Clarity: Is the paper well written? Does it clearly state its contributions, notation and results?**

   Most parts of the manuscript are well-written. Parts of procedure description and numerical experiments are unclear. See additional comments for details.

6. **Relation to prior work: Is it clearly discussed how this work differs from or relates to prior work in the literature?**

   More multiway/tensor clustering methods should be discussed and compared:

[1] Han, Rungang, Yuetian Luo, Miaoyan Wang, and Anru R. Zhang. "Exact clustering in tensor block model: Statistical optimality and computational limit. Journal of the Royal Statistical Society: Series B. 2022.

[2] Jing, Bing-Yi, Ting Li, Zhongyuan Lyu, and Dong Xia. "Community detection on mixture multilayer networks via regularized tensor decomposition." The Annals of Statistics 49, no. 6 (2021): 3181-3205.

[3] Lyu, Zhongyuan, and Dong Xia. "Optimal Clustering by Lloyd Algorithm for Low-Rank Mixture Model." arXiv preprint arXiv:2207.04600 (2022).

[4] Ke, Zheng Tracy, Feng Shi, and Dong Xia. "Community detection for hypergraph networks via regularized tensor power iteration." arXiv preprint arXiv:1909.06503 (2019).

[5] Hu, Jiaxin, and Miaoyan Wang. "Multiway Spherical Clustering via Degree-Corrected Tensor Block Models." In International Conference on Artificial Intelligence and Statistics, pp. 1078-1119. PMLR, 2022.

Particularly, the method in [1] is proven to achieve the optimal error rate under the TBM structure, which should be the most competitive method for the proposed method. Both [2] and [3] consider the order-3 multiway clustering under the mixture model; both [4] and [5] tackle the same multiway clustering problem with individual heterogeneity.

7. **Additional Comments: Add your additional comments, feedback and suggestions for improvement, as well as any further questions for the authors.**

Methodology:

- It is confusing that the cluster assignments are not involved in the main models (2) and (3). Then, how should we estimate the clustering assignment and evaluate the clustering results if there is no clustering structure underlying the data tensors?

- The affinity matrix clustering seems not well-motivated in the multiway clustering context. Consider the rank-1 and noiseless ($\mathcal{Z} = 0$) case. All slices have the same top eigenvector $\mathbf{w}_1$ with distinct eigenvalues $\hat{\gamma}_i = \gamma_1 \mathbf{u}_1(i)$. The information affinity matrix $C'$ as well as the clustering result depend only on the first mode factor $\mathbf{u}_1$ . Hence, why we generate the affinity matrix with term $|\langle \hat{\mathbf{w}}_i, \hat{\mathbf{w}}_j \rangle|$, which introduces extra noise in noisy clustering? Moreover, it is doubtful whether one can obtain satisfactory clustering performance of $m_1$ nodes only based on the information in an $m_1$-dimensional vector, even under the noiseless case. Similar issue applies for the rank-d cases. More intuition and explanations for the affinity matrix clustering should be added.

- The cross similarity issue could be considered more carefully. My understanding for the cross similarity is following:

Suppose the tensor slices have all distinct non-zero eigenvalues. Then, by standard eigen-decomposition, the eigenvectors $\mathbf{x}_1$ is orthogonal to $\mathbf{x}_2$. There should have no cross similarity issue under the noiseless case, neither under the noisy case with controlled level of noise.

Suppose the tensor slices have degenerate non-zero eigenvalues; i.e., one eigenvalue corresponds multiple eigenvectors. Under the noise case, cross similarity occurs due to the non-unique eigen-decomposition.

Current arguments in Section 2.3 do not essentially explain the occurrence of cross similarity. If my understanding is true, authors may adopt more advanced techniques to address the eigenvalue multiplicity issue, rather than propose two algorithms. Again, the affinity matrix construction problem in the last point should be fixed firstly.

Numerical experiment:

- It is unclear how to generate the clustering assignments based on description in Section 4.1.

- The result in Figure 3 seems suspicious. Under the setting $d = c = 1$, all methods should achieve accuracy 1 since there is only 1 cluster in the model and all nodes belong to the cluster. However, the results with $c = 1$ are close to 0, which contradicts to the intuition and lacks proper interpretation. In addition, MCAM results should also shown in Figure 3 to obtain the conclusion that MCAM is robust under low-rankness.

- The comparison with TBM in Figure 2 is unfair. The signal in TBM is characterized by the Euclidean gap between the tensor slices, rather than the magnitude of the largest eigenvalue. TBM also suffers model misspecification under the CP model (11) because TBM assumes a special Tucker structure with binary membership factor matrices.

- The algorithm output $\mathcal{C}$ is not defined; adding the specific formula of ARI is helpful.

## References