# Review for

## "Sparse logistic tensor decomposition for binary data"

This work proposes a sparse logistic tensor decomposition method and applies the majorization-minimization (MM) algorithm to solve the regularized likelihood optimization problem.

This paper generalizes previous work (Allen, 2012; Sun et al., 2017) to the binary case. The extension via MM approach is standard. Theoretically, no algorithm accuracy is provided. Numerically, there is no comparison with the competitive method (Kolda and Hong, 2020) that tackles the same problem. Hence, I am on the fence with respect to this paper.

1. (Novelty) Novelty is the biggest concern of this work. The extension from Allen (2012); Sun et al. (2017) to the binary case via MM approach is natural. Theoretically, the estimation accuracy for the algorithm output is not provided, and the main theorems are straightforward. Numerically, there is no comparison with Kolda and Hong (2020) that tackles the same sparse low-rank tensor decomposition problem. It is worthwhile to show the unique contribution in this binary extension and thus emphasize the novelty.

2. (Theorem) The theoretical results do not indicate the accuracy of the algorithm outputs. Specifically, Theorems 1, 2, and 4 show the algorithm convergence to the local minimum. However, we do not know the performance of local minimum compared with true parameters and the number of iterations or the speed for the algorithm to achieve the local minimum. Therefore, these theorems do not guarantee the performance of the algorithm.

   In addition, Theorem 2 and 3 focus only on the rank-1 optimization. The algorithm performance with $R > 1$ and $L$ initialization is not guaranteed.

   Also, some theorem statements are vague. Theorem 2 and 3 show the analytical solution to the sub-optimization problem. However, the true parameters $u, v, w$ are involved in the estimations, and it is doubtful whether the presented $\hat{u}, \hat{v}, \hat{w}$ are the solution of the joint optimization over $u, v, w$. Theorem 1 and 4 show the quadratic error "majorizes" the negative likelihood. Is there any rigorous definition of "majorize"?

3. (Model) More discussion on the offset $\mu$ would be helpful. The extra offset $\mu$ leads to the worry for the model identifiability and extra computational steps. It is beneficial to explain the necessity of including the offset in the model, and the validation of identifiability and computational cost with offset.

4. (Experiment) Several problems arise in the numerical experiments.

   - (Comparison) The competitive method (Kolda and Hong, 2020) tackles the same problem. It is worthwhile to involve the competitive method in all numerical and theoretical comparisons. It is also helpful if the comparison is non-applicable.

Also, in the comparison of Table 1, TSP and TTP are given true sparsity parameters, which is unfair for other methods. Parameter tuning procedures should be included for fair comparison.

- (Computational time) In computation time comparison, it would be helpful to report the time for initialization and tuning sparsity parameter, since these step are unavoidable in practice. Further, the time to determine $R$, determine $L$, and find $L$ initializations are also critical evaluate the computational efficiency in the realistic applications.

In addition, it is helpful to report the theoretical computational complexity of the proposed MM-based method and compare with others.

It is also curious why the iteration number and iteration time for BR are missing in Table 2.

- (Model misspecification) In practice, with perturbed observation, we do not know the sparsity information. It would be helpful to test the robustness of the proposed method to the model misspecification. For example, what are the performances of TSP and TTP with dense low-rank tensors. and how about other methods?

- (Tensor completion) It is also helpful to include the experiments to evaluate the algorithm performance on tensor completion problem, since authors provide theoretical analysis on tensor completion in Section 5.

- (Tuning parameter) Authors provide three methods to select rank and tune parameters. It would be helpful to provide some guidance on which criteria should be used in practice.

## References


Allen, G. (2012). Sparse higher-order principal components analysis. In *Artificial Intelligence and Statistics*, pages 27–36. PMLR.

Kolda, T. G. and Hong, D. (2020). Stochastic gradients for large-scale tensor decomposition. *SIAM Journal on Mathematics of Data Science*, 2(4):1066–1095.

Sun, W. W., Lu, J., Liu, H., and Cheng, G. (2017). Provable sparse tensor decomposition. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):899–916.