Review for
"Robust and Covariance-assisted Tensor Response Regression"

Jiaxin Hu

This work proposes a tensor response regression model, CATL, with noise following tensor t distribution. The t distribution is adopted to better address the heavy tail and outlier issues in real data. The tensor coefficients are assumed to be low-ranked with decomposition factors equal to the eigenvectors of the covariance matrices.

Tensor regression with heavy tail noise is an interesting topic. However, novelty and correctness concerns arise with current manuscript. I believe that the manuscript can be improved from multiple aspects.

Before the detailed comment, it is notable that important conclusions (e.g. the MLE of coefficient) and literature comparisons rely on the reference, Wang, Zhang, and Mai (2022). However, this reference is not publicly available, which may seriously affect the manuscript evaluation. It is worthwhile to include the public access of Wang, Zhang, and Mai (2022) in the paper.

Comments:

1. (Novelty) Novelty is the biggest concern of the proposed method.

   Based on Section 3, the covariance-assisted structure is a special case of the tensor-response envelope method in Li and Zhang (2017). The remark that CATL finds a smaller subspace that envelope model is natural. Because CP decomposition is a special case of Tucker decomposition with zero-valued off-diagonal elements in the core tensor $\boldsymbol{\Theta}$. The benefits of finding a smaller subspace are not stated. The algorithm structure in Section 4 to solve the covariance-assisted optimization is also very similar with Li and Zhang (2017). In this sense, the novelty of the proposed method lie in the coefficient $\boldsymbol{B}$ estimation under the t-distributed noise.

   However, the estimation of $\boldsymbol{B}$ under t-distribution is already analyzed in the reference Wang, Zhang, and Mai (2022), based on description in Section 2.3. Unlike normal model, weighted least square estimator is used in the t model. Due to the non-accessible reference, I can not tell whether the simplified estimator $\hat{\boldsymbol{B}}$ in (4) is also used in Wang, Zhang, and Mai (2022). But it is sure that the idea of weighted estimator is not new here.

   Therefore, I feel the contribution of this work is incremental. More novelty contributions should be emphasized to impress the readers.

2. (Theory) The Theorem 1 and Corollary 1 seem incorrect. Specifically, the term $\log(n)/p$ in the error bound is counter-intuitive. With fixed dimension p, the term $\log(n)/p$ leads to a increasing error as sample size n tends to ∞. Discussions should be added to address this

abnormal term. Also, let $n > \mathcal{O}(\exp(p))$. The consistency in Corollary 1 collapses since the term $\log(n)/p \to \infty$ as $p, n \to \infty$. Hence, the $\sqrt{n}$-consistency in the explanation does not hold.

3. (Algorithm) The algorithm section should be re-organized. Several questions arise in this section

   (a) The motivation of non-iterative algorithm in unclear. Based on page 11, authors avoid the iterative algorithm by calculating the Frobenius norm rather than the Mahalanobius distance. Note that Mahalanbius distance is a weighted version of Frobenius norm with weights Σ_m. It is unclear for me that why people need iterations with Mahalanobius distance. Moreover, it is worthwhile to present the specific computation complexity of the iterative and non-iterative algorithm to support the computational efficiency of CATL.

   (b) I think it is better to show the finite sample version of the algorithm. The population version is confusing. What are the inputs and outputs of the algorithm? Why we can assume the unknown parameters Σ_m and $\boldsymbol{B}$ in the algorithm? The output in step 3 includes the rank $\tilde{R}$, and then why and how we use cross-validation to estimate the rank R in the finite sample version?

   (c) It would be better to add more implications of Lemma 1. Lemma 1 indicates that the true space is always a subspace of the algorithm outputs but does not show any algorithm guarantees.

4. (Presentation) There are many presentation issues hinder the understanding of the paper.

   (a) In page 12, what is the latent variable G_i? I do not find any explanations for G_i in previous sections. How to interpret the latent variable?

   (b) In page 7, the denominator of the weight w_i includes the square of Mahalanobius distance; in page 11, the estimated weight $\hat{w}_i$ has denominator of the Frobenius norm without square. As the Frobenius and Mahalanobius distance are of the same order, the discrepancy should be a typo.

   (c) The shorthands p, p_{-m} are defined in Definition 1 rather than the notation section.

   (d) The term "envelope rank " is not defined in the notation section.

References

Li, L. and Zhang, X. (2017). Parsimonious tensor response regression. *Journal of the American Statistical Association*, 112(519):1131–1146.