Review for

“Sharp rates of convergence for the tensor graphical Lasso estimator”

Jiaxin Hu

This work focuses on the theoretical results for the global minimizer in tensor graphical Lasso (TeraLasso) model. Specifically, the improved error bound drops the undesirable factor $\log p$ by a new concentration inequality for diagonal elements, where $p = \prod_k d_k$ is the product of the order-K tensor dimensions. This improvement thus provides the guarantee for the empirical performance of two-way ($K = 2$) TeraLasso, which is unresolved in previous work Greenewald et al. (2019).

The new results are clearly stated and the theoretical analyses seem solid. However, the biggest concern lies in the theoretical and practical impact of this work. Hence, I am on the fence of this paper.

Major concerns:

1. (Impact) The new theoretical technique is only applied to sharpen the upper error bounds for the global minimizer of TeraLasso model with convex (ℓ_1) penalty. More influenced aspects can be discussed.
    - (Algorithm guarantee) The impact of the new results to the practical algorithm, TG-ISTA in Greenewald et al. (2019), is not introduced in present manuscript. Based on the analysis in Section 8, supplementary material of Greenewald et al. (2019), the number of iterations for TG-ISTA to achieve a desirable optimization error and the choice of optimal contraction factor seem relate to the statistical error of global minimizer. Adding more discussion on the influence to algorithm guarantees may strengthen the practical impact of the new theoretical techniques.
    - (Nonconvex optimization) Previous work Greenewald et al. (2019) also provided the error bound for TeraLasso with nonconvex penalty and indicated that both convex and nonconvex regularizers remain effectively the same bound. It would be interesting to discuss whether the improved convex rate may imply the improvement for the nonconvex optimization.
    - ($K = 2$) When $K = 2$, the TeraLasso model seeks the same estimator as two-ways models (Kalaitzis et al., 2013; Li et al., 2022; Yoon and Kim, 2022). As the new error rate provides a deeper understanding of two-way model convergence, it is curious to see the comparison of present result with the theoretical results in related works. Also, it would be interesting to see the empirical performance of all two-way algorithms compared with the present sharp error rate.

2. (Aspect ratio) The statements about aspect ratio and optimality under single-sample ($n = 1$) and balanced ($d_1 \asymp \cdots \asymp d_K$) case in Section 3 are a little vague for me. In TeraLasso model, the ratio between the number of parameters and sample size would be $\sum_k d_k^2/p \asymp d_{\max}/m_{\min}$, which is equal to the aspect ratio defined below Theorem 5. In this sense, it seems that the aspect ratio refers to the optimal rate in general case (even without sparsity) while the statement below Corollary 6 says the convergence achieves optimal with a large sparsity. Similar optimality statements is provided for the discussion below Theorem 8. In addition, optimality is usually supported by upper and lower error bounds of the estimator, whereas the present work only discusses the upper bound. More discussions about the aspect ratio and optimality would be helpful.

3. (Simulation) In previous work Greenewald et al. (2019) Section 6.1, the empirical optimization errors are shown to be much lower than the statistical error with more iterations. It would be interesting to see whether this phenomenon remains with a sharper bound, which may also directly illustrate the improvement of the error bound.

References

Greenewald, K., Zhou, S., and Hero III, A. (2019). Tensor graphical lasso (teralasso). *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 81(5):901–931.

Kalaitzis, A., Lafferty, J., Lawrence, N. D., and Zhou, S. (2013). The bigraphical lasso. In *International Conference on Machine Learning*, pages 1229–1237. PMLR.

Li, S., López-Garcıa, M., Lawrence, N. D., and Cutillo, L. (2022). Two-way sparse network inference for count data. In *International Conference on Artificial Intelligence and Statistics*, pages 10924–10938. PMLR.

Yoon, J. H. and Kim, S. (2022). Eiglasso for scalable sparse kronecker-sum inverse covariance estimation. *The Journal of Machine Learning Research*, 23(1):4733–4771.