Review for "Generalized Low-rank plus Sparse Tensor Estimation by Fast Riemannian Optimization"

Chanwoo Lee,

This paper provides a generalized low-rank plus sparse tensor model to account for heterogeneous signals that low-rank tensor model fails to explain. The authors propose a fast algorithm combining the Riemannian gradient descent and gradient pruning procedure. Under some conditions on ground truth tensors and considered loss function, the final estimates enjoy the statistical guarantees. Furthermore, stricter theoretical guarantees are established applying the general theoretical results to specific cases including the robust sub-Gaussian tensor PCA, tensor PCA under heavy-tailed noise, and binary tensor models. The real data analysis on international commodity trade flows is provided based on the suggested model.

Major points:

1. What is the benefits of the suggested method compared to alternating optimization approach? In updating low-rank tensor $\widehat{\mathcal{T}}_\ell$, the authors perform Riemannian gradient descent. Since this gradient procedure does not guarantee low-rankness of the output tensor, they project the tensor to the smooth manifold $\mathbb{M}_{\boldsymbol{r}}$. Instead of the projection, the alternating optimization updates each factor $\mathcal{C}, \boldsymbol{U}_1, \ldots, \boldsymbol{U}_m$ of low-rank tensor $\mathcal{T}$ with other factors being fixed. This approach automatically guarantees the low-rankness of the output tensor. Since both methods require the initial points to be close to ground truth points, I wonder the advantages of the projection approach compared to the alternating optimization.

2. How to set the unknown hyperparameters: rank $\boldsymbol{r}$, the sparsity parameter α, and μ_1 in Assumption 1? Except Section 5.2. where α is set according to Theorem 5.3, setting the hyperparameters is not fully discussed. Unlike other tuning parameters such as γ and k_{pr}, the hyperparamters are from the ground truth model and need to be estimated from the observed tensor $\mathcal{A}$ when the hyperparameters are unknown. Because of the increased model complexity and relationship among hyperparameters, I guess estimating the unknown hyperparameters in their setting is not as straightforward as in low-rank tensor model (for example, one can simply use spectral method, cross-validation, or Bayesian Information Criterion to set the rank). It would have been great to introduce general rules for setting those hyperparameters when they are unknown.

   In addition, the parameter μ_1 is introduced to handle identifiability issue between low-rank tensor $\mathcal{T}^*$ and sparse tensor $\mathcal{S}^*$. In this sense, I guess that μ_1 is related to the rank $\boldsymbol{r}$ and the sparsity parameter α, which becomes a constraint when searching the unknown hyperparameters. Therefore, I wonder how those hyperparameters are related.

3. The suggested algorithm uses gradient pruning with $\gamma\alpha$ not α where γ is a tuning parameter such that $\gamma > 1$. I am curious why the gradient pruning with exact α leads to divergence of error bound in Theorem 4.1, which is somewhat counter intuitive to me. Similarly, how do I intuitively understand the condition $\gamma > 1 + c_{4,m}^{-1} b_u^4 b_\ell^4$?

4. The sup-norm error of $\hat{\mathcal{S}}_{l_{\max}}$ in Theorem 4.3 is great but the argument for recovering the support of $\mathcal{S}^*$ based on Theorem 4.3. is not convincing. The condition that all magnitudes of non-zero entries in $\mathcal{S}^*$ should be two times greater than the sup-norm error in Theorem 4.3., seems too strict. This is because the definition of sparse tensor has nothing to do with the magnitude of its entries. The authors should provide some cases that this condition is

reasonable showing that the sup-norm error is negligibly small under some cases.

Overall, I think this is good paper. The suggested model incorporates low-rank model and constructed algorithm has theoretical guarantees. But it is also true that there is some room for improvement in this paper. I cannot quite figure out whether it rises to the JASA level.

Minor points:

1. At line 43 ,page 15, there seems to be a typo: $d^* := d_j d_j^{-1} = d_1 \cdots d_m$.