Comments about *Paper Sketch for AISTATS*

Zhuoyan Xu

This note follows the the note [*Paper Sketch for AISTATS Tentative title: "Binary tensor regression with multi-mode features*].

1 Sqrate root of matrix

According to your property 1, we have

$$(\boldsymbol{X}^T\boldsymbol{X})^{\frac{1}{2}} = \Delta Q^T$$

According to the definition of square root of matrix we have:

$$B = A^{\frac{1}{2}} \Longleftrightarrow A = BB$$

Since $\boldsymbol{X}^T\boldsymbol{X} = Q\Delta^2 Q^T$, I suppose $(\boldsymbol{X}^T\boldsymbol{X})^{\frac{1}{2}} = Q\Delta Q^T$. Are these two forms equivalent?

2 RIP Assumption

Could you explain why there is a term **d** in assumption? This term didn't show up in following proof.

3 Sharper Bound

In your main result of theorem 1, the RIP parameter in numerator might be sharpened to $1+\delta_{r,\alpha}$ from $1+\delta_{2r,2\alpha}$.

The main result in page 3 can be sharpened to:

$$\left\|\hat{\mathcal{B}}_{MLE} - \mathcal{B}_{true}\right\|_F \leq \frac{C_\alpha}{\prod_k d_k}\sqrt{\frac{(1+\delta_{r,\alpha})}{(1-\delta_{2r,2\alpha})^2}\frac{\prod_{k=1}^K r_k}{r_{\max}}\sum_{k=1}^K p_k}$$

In proof of lemma 2 on page 5, consider last inequality of first formula:

$$\|\mathcal{E}\|_\sigma \times \sqrt{\frac{\prod_k r_k}{r_{\max}}} \times \left\|\mathcal{B} \times_1 \left(\tilde{\boldsymbol{X}}_1^T \boldsymbol{X}_1\right) \times_2 \cdots \times_K \left(\tilde{\boldsymbol{X}}_K^T \boldsymbol{X}_K\right)\right\|_F$$

We have:

$$\left\|\mathcal{B} \times_1 \left(\tilde{\boldsymbol{X}}_1^T \boldsymbol{X}_1\right) \times_2 \cdots \times_K \left(\tilde{\boldsymbol{X}}_K^T \boldsymbol{X}_K\right)\right\|_F \leq \|\mathcal{B}\|_F \|\tilde{\boldsymbol{X}}_1^T \boldsymbol{X}_1\|_2 \cdots \|\tilde{\boldsymbol{X}}_K^T \boldsymbol{X}_K\|_2$$

where $\|A\|_2$ denote operator norm of matrix A.

Since:

$$\tilde{\boldsymbol{X}}_1^T \boldsymbol{X}_1 = (\boldsymbol{X}_1^T \boldsymbol{X}_1)^{\frac{1}{2}}$$
$$\|\tilde{\boldsymbol{X}}_1^T \boldsymbol{X}_1\|_2 = \|(\boldsymbol{X}_1^T \boldsymbol{X}_1)^{\frac{1}{2}}\|_2 = \|\boldsymbol{X}\|_2$$

Then:

$$\left\|\mathcal{B} \times_1 \left(\tilde{\boldsymbol{X}}_1^T \boldsymbol{X}_1\right) \times_2 \cdots \times_K \left(\tilde{\boldsymbol{X}}_K^T \boldsymbol{X}_K\right)\right\|_F \leq \|\mathcal{B}\|_F \|\boldsymbol{X}_1\|_2 \cdots \|\boldsymbol{X}_K\|_2$$

Since for every $\mathcal{B}$, we have:

$$|\mathcal{B} \times_1 \boldsymbol{X}_1 \times_2 \cdots \times_K \boldsymbol{X}_K\|_F \leq \|\mathcal{B}\|_F \|\boldsymbol{X}_1\|_2 \cdots \|\boldsymbol{X}_K\|_2$$

According to RIP assumption, X must satisfy:

$$|\mathcal{B} \times_1 \boldsymbol{X}_1 \times_2 \cdots \times_K \boldsymbol{X}_K\|_F \leq \sqrt{1+\delta_{\boldsymbol{r},\alpha}} \|\mathcal{B}\|_F$$

Then for the X that satisfies the RIP assumption, we have:

$$\|\mathcal{B}\|_F \|\boldsymbol{X}_1\|_2 \cdots \|\boldsymbol{X}_K\|_2 \leq \sqrt{1+\delta_{\boldsymbol{r},\alpha}} \|\mathcal{B}\|_F$$
$$\|\boldsymbol{X}_1\|_2 \cdots \|\boldsymbol{X}_K\|_2 \leq \sqrt{1+\delta_{\boldsymbol{r},\alpha}}$$

Thus we have inequality:

$$\|\mathcal{E}\|_\sigma \times \sqrt{\frac{\prod_k r_k}{r_{\max}}} \times \left\|\mathcal{B} \times_1 \left(\tilde{\boldsymbol{X}}_1^T \boldsymbol{X}_1\right) \times_2 \cdots \times_K \left(\tilde{\boldsymbol{X}}_K^T \boldsymbol{X}_K\right)\right\|_F \leq \sqrt{\frac{\prod_k r_k}{r_{\max}}} \times \|\mathcal{E}\|_\sigma \times \sqrt{1+\delta_{r,\alpha}} \|\mathcal{B}\|_F$$

When we replace $\mathcal{B}$ with $\hat{\mathcal{B}}_{\text{MLE}} - \mathcal{B}_{\text{true}}$, we can keep the same RIP parameter.

$$\begin{aligned}\left\|(\hat{\mathcal{B}}_{\text{MLE}} - \mathcal{B}_{\text{true}}) \times_1 \left(\tilde{\boldsymbol{X}}_1^T \boldsymbol{X}_1\right) \times_2 \cdots \times_K \left(\tilde{\boldsymbol{X}}_K^T \boldsymbol{X}_K\right)\right\|_F &\leq \|\hat{\mathcal{B}}_{\text{MLE}} - \mathcal{B}_{\text{true}}\|_F \|\boldsymbol{X}_1\|_2 \cdots \|\boldsymbol{X}_K\|_2 \\ &\leq \sqrt{1+\delta_{r,\alpha}} \|\hat{\mathcal{B}}_{\text{MLE}} - \mathcal{B}_{\text{true}}\|_F\end{aligned}$$

Thus on proof of theorem 1 on page 2, we have:

$$\begin{aligned}&(1-\delta_{2r,2\alpha}) \left\|\left(\hat{\mathcal{B}}_{\text{MLE}} - \mathcal{B}_{\text{true}}\right)\right\|_F^2 \\ &\leq \left\|\left(\hat{\mathcal{B}}_{\text{MLE}} - \mathcal{B}_{\text{true}}\right) \times_1 \boldsymbol{X}_1 \times_2 \cdots \times_K \boldsymbol{X}_K\right\|_F^2 \\ &\leq C_\alpha \times \left\|\hat{\mathcal{B}}_{\text{MLE}} - \mathcal{B}_{\text{true}}\right\|_F \times \sqrt{(1+\delta_{r,\alpha}) \frac{\prod_k r_k}{r_{\max}} \sum_k p_k}\end{aligned}$$

4 $sG(\sigma)$ in Lemma 3

In lemma 3 you said $\mathcal{E} = \mathcal{S} \times_1 \tilde{\boldsymbol{X}}_1^T \times_2 \cdots \times_K \tilde{\boldsymbol{X}}_K^T$ is an $sG(\sigma)$ tensor. Does this notation mean the entries in $\mathcal{E}$ are mutually independent? There might still exist dependence among entries. We can just invoke spectral norm theorem.