

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 Sqrare root of matrix

According to your property 1, we have

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

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

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

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

2 RIP Assumption

Could you explain why there is a term \mathbf{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:

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

We have:

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

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

Since:

$$\begin{aligned} \tilde{\mathbf{X}}_1^T \mathbf{X}_1 &= (\mathbf{X}_1^T \mathbf{X}_1)^{\frac{1}{2}} \\ \|\tilde{\mathbf{X}}_1^T \mathbf{X}_1\|_2 &= \|(\mathbf{X}_1^T \mathbf{X}_1)^{\frac{1}{2}}\|_2 = \|\mathbf{X}_1\|_2 \end{aligned}$$

Then:

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

According to RIP assumption:

$$\begin{aligned} \|\mathcal{B} \times_1 \mathbf{X}_1 \times_2 \cdots \times_K \mathbf{X}_K\|_F^2 &\leq \|\mathcal{B}\|_F \|\mathbf{X}_1\|_2 \cdots \|\mathbf{X}_K\|_2 \leq (1 + \delta_{r,\alpha}) \|\mathcal{B}\|_F^2 \\ \|\mathbf{X}_1\|_2 \cdots \|\mathbf{X}_K\|_2 &\leq 1 + \delta_{r,\alpha} \end{aligned}$$

Thus we have inequality:

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

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

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

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

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

4 $sG(\sigma)$ in Lemma 3

In lemma 3 you said $\mathcal{E} = \mathcal{S} \times_1 \tilde{\mathbf{X}}_1^T \times_2 \cdots \times_K \tilde{\mathbf{X}}_K^T$ 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.