# Prove Idea of statistical upper bound for tucker decomposition and some simulation result

Jiaxin Hu    Zhuoyan Xu

Date: 2019.8.1

# 1 Prove Idea of upper bound

## 1.1 Matrix form of linear term

Consider the prove idea of Theorem 1, we want to extent it to a *tucker decomposition* version. Due to $\mathcal{L}_Y(\hat{\Theta}) - \mathcal{L}_Y(\Theta_{true}) \geq 0$ and:

$$\mathcal{L}_Y(\hat{\Theta}) - \mathcal{L}_Y(\Theta_{true}) = \left|\langle \mathcal{S}_Y(\Theta_{true}), (\hat{\Theta} - \Theta_{true})\rangle\right| + \frac{1}{2}\mathbf{vec}(\hat{\Theta} - \Theta_{true})^T \mathcal{H}_Y(\tilde{\Theta})\mathbf{vec}(\hat{\Theta} - \Theta_{true})$$

Since

$$\mathbf{vec}(\hat{\Theta} - \Theta_{true})^T \mathcal{H}_Y(\tilde{\Theta})\mathbf{vec}(\hat{\Theta} - \Theta_{true}) \leq \gamma_\alpha \|\hat{\Theta} - \Theta_{true}\|_F^2$$

Therefore, the key point is to find $\left|\langle \mathcal{S}_Y(\Theta_{true}), (\hat{\Theta} - \Theta_{true})\rangle\right| \leq f(\|\hat{\Theta} - \Theta_{true}\|_F)$.
For $\langle \cdot, \cdot \rangle$ is entry-wise inner product, so we can easily transfer the tensor inner product to matrix inner product:

$$\langle \mathcal{S}_Y(\Theta_{true}), (\hat{\Theta} - \Theta_{true})\rangle = \langle \mathcal{S}_Y(\Theta_{true})_{(1)}, (\hat{\Theta} - \Theta_{true})_{(1)}\rangle$$

where $A_{(n)}$ refers to the *mode-n matricization* of an order $K$ tensor. Then according to the fact that:

$$\langle X, Y \rangle \leq \|X\|_\sigma \|Y\|_*$$

where $\|\cdot\|_\sigma$ is spectral norm and $\|\cdot\|_*$ is nuclear norm for matrix. So,

$$\langle \mathcal{S}_Y(\Theta_{true})_{(1)}, (\hat{\Theta} - \Theta_{true})_{(1)}\rangle \leq \|\mathcal{S}_Y(\Theta_{true})_{(1)}\|_\sigma \|(\hat{\Theta} - \Theta_{true})_{(1)}\|_*$$

Due to the fact that :

$$rank_T(A) = (rank(A_{(1)}), \ldots, rank(A_{(K)})); \quad \|X\|_* \leq \sqrt{rank(X)}\|X\|_F$$

Then we get:

$$\|(\hat{\Theta} - \Theta_{true})_{(1)}\|_* \leq \sqrt{2r_1}\|(\hat{\Theta} - \Theta_{true})_{(1)}\|_F = \sqrt{2r_1}\|(\hat{\Theta} - \Theta_{true})\|_F$$

As to the spectral norm $\|\mathcal{S}_Y(\Theta_{true})_{(1)}\|_\sigma$, $\mathcal{S}_Y(\Theta_{true})_{(1)} \in \mathbb{R}^{d_1 \times \prod_{k=2}^K d_k}$. Because matrix is the special case of tensor, use Lemma 1 by Tomioka and Suzuki and Lemma 2, we can get:

$$\mathbb{P}(\|\mathcal{S}_Y(\Theta_{true})_{(1)}\|_\sigma \geq C' L_\alpha \sqrt{d_1 + \prod_{k=2}^K d_k}) \leq exp(-Clog2(d_1 + \prod_{k=2}^K d_k))$$

Then we can consider that:

$$\|\mathcal{S}_Y(\Theta_{true})_{(1)}\|_\sigma \leq C' L_\alpha \sqrt{d_1 + \prod_{k=2}^K d_k}$$

with probability $1 - exp(-Clog2(d_1 + \prod_{k=2}^K d_k))$. Plus $K$ inequalities,

$$\frac{\langle \mathcal{S}_Y(\Theta_{true}), (\hat{\Theta} - \Theta_{true})\rangle}{\sqrt{\prod_{k=1}^K d_k}} \leq \sqrt{2} C' L_\alpha \frac{1}{K} \sum_{k=1}^K \sqrt{\frac{d_k r_k + \prod_{k' \neq k}^K d_{k'} r_k}{\prod_{k=1}^K d_k}} \|(\hat{\Theta} - \Theta_{true})\|_F$$

That would lead to :

$$\begin{aligned} \frac{1}{\sqrt{\prod_{k=1}^K d_k}} \|(\hat{\Theta} - \Theta_{true})\|_F &\leq 2\sqrt{2} \frac{C' L_\alpha}{K \gamma_\alpha} \sum_{k=1}^K \sqrt{\frac{d_k r_k + \prod_{k' \neq k}^K d_{k'} r_k}{\prod_{k=1}^K d_k}} \\ &= 2\sqrt{2} \frac{C' L_\alpha}{K \gamma_\alpha} \sum_{k=1}^K \sqrt{\frac{d_k r_k}{\prod_{k=1}^K d_k} + \frac{r_k}{d_k}} \end{aligned}$$

with probability $min_k\{1 - exp(-Clog2(d_k + \prod_{k' \neq k}^K d_{k'}))\}$.

## 1.2 All spectral norm of linear term

Consider:

$$\begin{aligned} \langle \mathcal{S}_Y(\Theta_{true}), (\hat{\Theta} - \Theta_{true})\rangle \leq& \|\mathcal{S}_Y(\Theta_{true})\|_\sigma \|(\hat{\Theta} - \Theta_{true})\|_* \\ \|(\hat{\Theta} - \Theta_{true})\|_* \leq& 2R \|(\hat{\Theta} - \Theta_{true})\|_\sigma \end{aligned}$$

Consider $\hat{\Theta} - \Theta_{true}$ as random variables(need to be verified), where

$$\mathbb{E}[\hat{\Theta} - \Theta_{true}] = 0 \ , \quad |\hat{\Theta} - \Theta_{true}| \leq 2\alpha$$

with probability at least $1 - \exp(-C_2 \log K \sum_k d_k)$:

$$\left\|(\hat{\Theta} - \Theta_{true})\right\|_\sigma \leq 2C_3 \alpha \log K \sqrt{\sum_k d_k} \tag{1}$$

Then, with probability at least $\max\{1-\exp\left(-C_1 \log K \sum_k d_k\right), 1-\exp\left(-C_2 \log K \sum_k d_k\right)\}$:

$$|\langle S_{\mathcal{Y}}\left(\Theta_{\text{ true }}\right), \Theta - \Theta_{\text{ true }}\rangle| \leq C_2 C_3 \alpha L_\alpha R \sum_{k=1}^{K} d_k$$

Henceforth, with probability at least $\max\{1-\exp\left(-C_1 \log K \sum_k d_k\right), 1-\exp\left(-C_2 \log K \sum_k d_k\right)\}$:

$$\frac{1}{\sqrt{\prod_k d_k}} \left\| \hat{\Theta} - \Theta_{\text{ true }} \right\|_F \leq= \sqrt{2 C_2 C_3 \alpha \frac{L_\alpha R}{\gamma_\alpha} \frac{\sum_{k=1}^{K} d_k}{\Pi_k d_k}}$$

# 2 Simulation

We've go through the simulation. The algorithm we use is matrix form GLM, we consider each iteration consists of one update for each of factor matrices and core tensor. Since most of our result are based on 10 or 20 iterations, which is not enough to get MLE, neither the results are confidential. When we set iteration time to 100, the logLik of model still increase. As shown in figure 2, which is not a rare case.

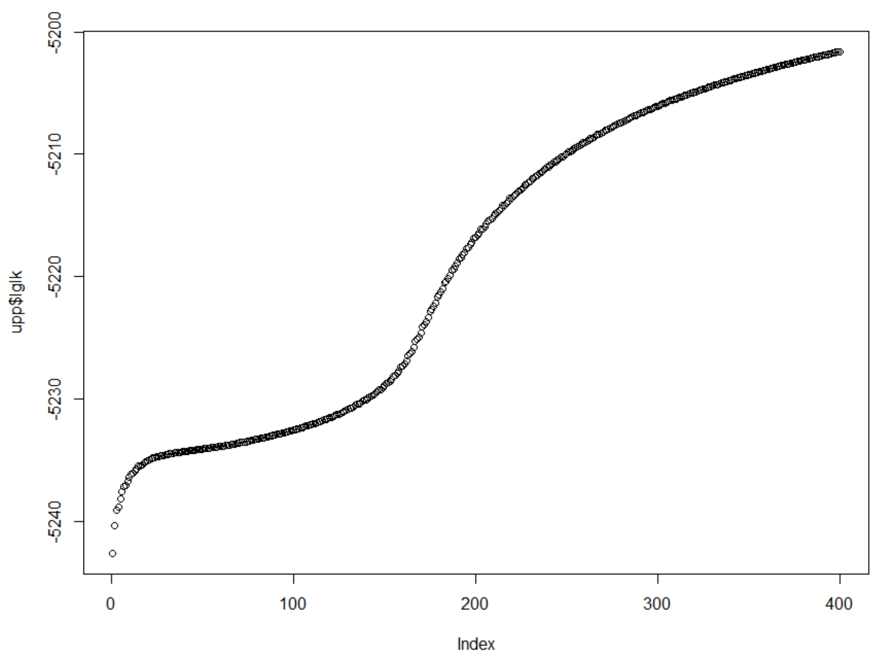


Figure 1: logLik

If we want to compute the MLE, we need to set iteration times as 100 or 1000 until it get maximum logLik(or until it converge with a selected tolerance), in which case the extreme large entry case occurs.

The ultimate goal is to find some bound of MSE betweem MLE and true parameter, where MLE and true parameter all comes from a parameter space with restriction on rank or max norm. We may need to modify the algorithm to search MLE in a restriction area(either max norm or rank or other restrictions).