# Tensor Clustering

Jiaxin Hu

July 15, 2021

## 1 Model

Consider an order-d tensor $\mathcal{Y} \in \mathbb{R}^{p_1 \times \cdots \times p_d}$. Suppose there are $r_k$ clustering in the $k$-th mode for $k \in [d]$. Let $z_k \in [r_k]^{p_k}$ denote the clustering assignment and $\theta_k = (\theta_{k,1}, ..., \theta_{k,p_k}) \in \mathbb{R}^{p_k}$ denote the degree-corrected parameters on the $k$-th mode. Consider the degree-corrected tensor block model

$$\mathcal{Y}_{i_1,...,i_d} = \theta_{1,i_1} \cdots \theta_{d,i_d} \mathcal{S}_{(z_1)_{i_1},...,(z_d)_{i_d}} + \mathcal{E}_{i_1,...,i_d}, \tag{1}$$

where $\mathcal{S} \in \mathbb{R}^{r_1 \times \cdots \times r_d}$ is the core tensor collected the block means, and $\mathcal{E}_{i_1,...,i_d}$ are independent mean zero Gaussian random noise with variance $\sigma^2$. Rewrite the model (1) in the tensor form

$$\mathcal{Y} = \mathcal{S} \times_1 \Theta_1 \boldsymbol{M}_1 \times_2 \cdots \times_d \Theta_d \boldsymbol{M}_d + \mathcal{E},$$

where $\Theta_k = \text{diag}(\theta_k) \in \mathbb{R}^{p_k \times p_k}$ and $\boldsymbol{M}_k \in \mathbb{R}^{p_k \times r_k}$ is the membership matrix generated by $z_k$.

## 2 Parameter Space

Suppose $(z_1, ..., z_d)$ belongs to the parameter space $\mathcal{P}_z(r_1, ..., r_d, \beta)$ for $\beta > 1$ where

$$\mathcal{P}_z(r_1, ..., r_d, \beta) = \left\{ (z_1, ..., z_d) : \frac{p_k}{\beta r_k} \leq \sum_{j=1}^{p_k} \mathbf{1}\left\{ (z_k)_j = a \right\} \leq \frac{\beta p_k}{r_k}, \quad \text{for all} \quad a \in [r_k], k \in [d] \right\}.$$

With given assignment $z$, define the minimal gap

$$\Delta_k^2 = \min_{j,j' \in [p_k], a \neq b \in [r_k]} \left\| (\theta_{k,j}^*)^2 \mathcal{M}_k(\mathcal{S}^*)_{a,:} - (\theta_{k,j'}^*)^2 \mathcal{M}_k(\mathcal{S}^*)_{b,:} \right\|_F^2, \quad \Delta_{\min}^2 = \min_{k \in [d]} \Delta_k^2.$$

Then core tensor and degree-corrected parameter $\mathcal{S}, \theta$ belongs to the space $\mathcal{P}(z, m, M)$ where

$$\mathcal{P}(z, m, M) = \left\{ (\mathcal{S}, \theta) : 0 < m \leq \theta_{k,j} \leq M, \ \frac{\sum_{j=1}^{p_k} \theta_{k,j} \mathbf{1}\left\{ (z_k)_j = a \right\}}{\sum_{j=1}^{p_k} \mathbf{1}\left\{ (z_k)_j = a \right\}} = 1, \Delta_{\min}^2 > 0, \text{ for all } a \in [r_k], k \in [d] \right\}$$

# 3 Main results (conjecture)

Similarly with (Gao and Zhang, 2019), the misclassification loss of MLE would be bounded by the oracle misclassification loss. So, the goal is to find 1) the relationship between loss and oracle loss; 2) the upper bound for oracle loss.

First, we define the loss

$$\ell_k(z) = \|\mathcal{S}^* \times_k (\Theta_k^* \boldsymbol{M}_k - \Theta_k^* \boldsymbol{M}_k^*)\|_F^2 = \sum_{j=1}^{p_k} (\theta_{k,j}^*)^2 \left\| \mathcal{M}_k(\mathcal{S}^*)_{(z_k)_j,:} - \mathcal{M}_k(\mathcal{S}^*)_{(z_k^*)_j,:} \right\|_F^2,$$

and the hamming distance

$$h_k(z) = \sum_{j=1}^{p_k} \mathbf{1}\left\{(z_k)_j \neq (z_k^*)_j\right\}.$$

Therefore, we have

$$h_k(z)\Delta_k^2 \leq \ell_k(z).$$

Next, we define the oracle estimator $\tilde{\theta}, \tilde{\mathcal{S}}$, such that

$$(\tilde{\theta}, \tilde{\mathcal{S}}) = \underset{(\theta,\mathcal{S}) \in \mathcal{P}(z^*, m, M)}{\arg\min} \|\mathcal{Y} - \mathcal{S} \times_1 \Theta_1 \boldsymbol{M}_1^* \times_2 \cdots \times_d \Theta_d \boldsymbol{M}_d^*\|_F^2.$$

Though the true assignment is known, it is still possible to misclassfy the clusters due to the random samples. Without the loss of generality, consider the mode 1 clustering and node $j$ is misclassified to group $b$ while $(z_1^*)_j = a$. Then, we have

$$\begin{aligned} &\left\| \left\{ \mathcal{Y} \times_1 \tilde{\Theta}_1^{-1} \times_2 \cdots \times_d \Theta_d^{-1} \right\}_{j:} \times_1 \boldsymbol{W}_1^T \times_2 \cdots \times_d \boldsymbol{W}_d^T - \tilde{\mathcal{S}}_{a:} \right\|_F^2 \\ &\quad \geq \left\| \left\{ \mathcal{Y} \times_1 \tilde{\Theta}_1^{-1} \times_2 \cdots \times_d \Theta_d^{-1} \right\}_{j:} \times_1 \boldsymbol{W}_1^T \times_2 \cdots \times_d \boldsymbol{W}_d^T - \tilde{\mathcal{S}}_{b:} \right\|_F^2 \end{aligned} \tag{2}$$

where $\boldsymbol{W}_d = \boldsymbol{M}_k^* (\mathrm{diag}(\mathbf{1}_{p_k}^T \boldsymbol{M}_k^*))^{-1}$. Note that the event (2) is equivalent to

$$\begin{aligned} &\langle \mathcal{E} \times_1 \boldsymbol{W}_1^T \tilde{\Theta}_1^{-1} \times_2 \cdots \times \boldsymbol{W}_d^T \tilde{\Theta}_d^{-1}, \tilde{\mathcal{S}}_{b:} - \tilde{\mathcal{S}}_{a:} \rangle \\ &\quad \leq \frac{1}{2} \left\{ - \left\| \left[ \mathcal{S}^* \times_1 \tilde{\Theta}_1^{-1} \Theta_1^* \times_2 \cdots \times_d \tilde{\Theta}_d^{-1} \Theta_d^* \right]_{a:} - \tilde{\mathcal{S}}_{a:} \right\|_F^2 + \left\| \left[ \mathcal{S}^* \times_1 \tilde{\Theta}_1^{-1} \Theta_1^* \times_2 \cdots \times_d \tilde{\Theta}_d^{-1} \Theta_d^* \right]_{a:} - \tilde{\mathcal{S}}_{b:} \right\|_F^2 \right\} \\ &\quad \approx -\frac{1}{2} \|S_{a:}^* - S_{b:}^*\|_F^2 \end{aligned}$$

# References

Gao, C. and Zhang, A. Y. (2019). Iterative algorithm for discrete structure recovery. arXiv preprint arXiv:1911.01018.