Initialization convergence

Jiaxin Hu

July 27, 2021

1 Model&Algorithm

Suppose we have p nodes from r communities and observe the adjacent tensor $\mathcal{Y} \in \{0,1\}^{p\times p\times p}$ whose entry $\mathcal{Y}_{ijk}$ refers to the connection of the triplet (i,j,k). Let $\theta = (\theta_1,...,\theta_p) \in \mathbb{R}^p$ denote the degree-corrected parameters and $z = (z_1,...,z_p) \in [r]^p$ denote the clustering assignment. Consider the hDCBM model

$$\mathbb{E}[\mathcal{Y}] = \mathcal{X} = \mathcal{S} \times_1 \Theta \boldsymbol{M} \times_2 \Theta \boldsymbol{M} \times_3 \Theta \boldsymbol{M},$$

where $\mathcal{S} \in \mathbb{R}^{r\times r\times r}$ is the symmetric core tensor, $\Theta = \text{diag}(\theta) \in \mathbb{R}^{p\times p}$ and $\boldsymbol{M} \in \mathbb{R}^{p\times r}$ is the hard membership matrix based on z. Consider the parameter space $\mathcal{P}_z(r,\beta)$ for $\beta > 1$ where

$$\mathcal{P}_z(r,\beta) = \left\{ z : \frac{p}{\beta r} \leq \sum_{j=1}^{p} \mathbf{1}\{z_j = a\} \leq \frac{\beta p}{r}, \quad \text{for all} \quad a \in [r] \right\},$$

and the parameter space $\mathcal{P}(\delta, \Delta_{\min}, \alpha)$ for $(\Theta, \mathcal{S})$, where

$$\begin{aligned}\mathcal{P}(\delta, \Delta_{\min}, \alpha) = \Bigg\{ (\mathcal{S},\Theta) : &\min_{u\neq u' \in [r]} \min_{v,w\in[p]} (\mathcal{S}_{uvw} - \mathcal{S}_{u'vw})^2 = \Delta_{\min}^2, \\ &\sum_{i=1}^{p} \theta_i^2 \in [1-\delta, 1+\delta], \\ &\max_{u,v,w\in[r]} \mathcal{S}_{uvw} \leq \alpha \Bigg\}.\end{aligned}$$

See Algorithm 1 for detailed algorithm. The convergence of the initialization is stated below:

Theorem 1.1 (Initialization convergence(conjecture))**.** *Suppose $\delta = o(1), \Delta_{\min} > 0, \|\theta\|_{\max} = o(p/r)$. Let $\tau = C_1\sqrt{rp^2\alpha \|\theta\|_{\max}^3}$ for sufficiently large $C_1 > 0$ in Algorithm 1. There exists constants C, C' such that with probability at least $1 - p^{-1+C'}$,*

$$\sum_{j:\hat{z}_j \neq z_j^*} \theta_j^2 \leq Cf(p,r,\delta,M,\Delta_{\min},\beta,\alpha) rp^2\alpha \|\theta\|_{\max}^3 .$$

Few comments:

1. The Algorithm 1 may not be optimal since it implements marginal clustering and does not take the advantages in estimation error. Moreover, the proposition 1 in Han et al. (2020) has an exponential rate which implies an exponential convergence in initialization. Is that because their initialization also uses the hard membership structure implicitly?

2. Varimax maybe another choice to cover the membership from singular space instead of k-means.

2 Key components to obtain the initialization convergence

This section shows the key components to prove Theorem 1.1.

2.1 Gap-free estimation error bound(Conjecture)

Lemma 1. *Assume $\|\theta\|_{\max} = o(p/r)$. Let $\tau = C_1\sqrt{rp^2\alpha \|\theta\|_{\max}^3}$ for sufficiently large $C_1 > 0$ in Algorithm 1. Then for any constant $C' > 0$, there exists some $C > 0$ such that*

$$\left\|\hat{\mathbf{Y}}_k - \mathcal{M}_k(\mathcal{X})\right\|_F \leq C\sqrt{rp^2\alpha \|\theta\|_{\max}^3},$$

with probability $1 - p^{-1+C'}$.

Proof Sketch of Lemma 1. Note that

$$\left\|\hat{\mathbf{Y}}_k - T_\tau(\mathcal{Y})\right\|_F^2 \leq \|\mathcal{M}_k(\mathcal{X}) - T_\tau(\mathcal{Y})\|_F^2 .$$

After arrangement, we have

$$\begin{aligned}\left\|\hat{\mathbf{Y}}_k - \mathcal{M}_k(\mathcal{X})\right\|_F^2 &\leq 2\langle \hat{\mathbf{Y}}_k - \mathcal{M}_k(\mathcal{X}), T_\tau(\mathcal{Y}) - \mathcal{M}_k(\mathcal{X})\rangle \\ &\leq 2\left\|\hat{\mathbf{Y}}_k - \mathcal{M}_k(\mathcal{X})\right\|_* \|T_\tau(\mathcal{Y}) - \mathcal{M}_k(\mathcal{X})\|_2 ,\end{aligned}$$

which implies

$$\left\|\hat{\mathbf{Y}}_k - \mathcal{M}_k(\mathcal{X})\right\|_F \leq 2\sqrt{r}\,\|T_\tau(\mathcal{Y}) - \mathcal{M}_k(\mathcal{X})\|_2 .$$

If we have

$$\|T_\tau(\mathcal{Y}) - \mathcal{M}_k(\mathcal{X})\|_2 \leq C\sqrt{p^2\alpha \|\theta\|_{\max}^3},$$

with probability $1 - n^{-1+C'}$, then we are done. □

2.2 Measurement of misclassification

For any set $S \subset [p]$, note that

$$\sum_{j\in S} \|(\mathcal{M}_k(\mathcal{X}))_j\|_F^2 = \sum_{j\in S} \theta_j^2 \left(\sum_{k,l\in[p]} \left[\theta_k \theta_l \mathcal{S}_{z_j^* z_k^* z_l^*}\right]^2 \right)$$
$$\geq \sum_{j\in S} \theta_j^2 \frac{(1-\delta)^4 r}{\beta p}.$$

Hence, to bound $\sum_{j\in S} \theta_j^2$, it is sufficient to bound $\sum_{j\in S} \|(\mathcal{M}_k(\mathcal{X}))_j\|_F^2$.

2.3 Weighted k-means

Let $\boldsymbol{X}_{kj}^s = \frac{(\mathcal{M}_k(\mathcal{X}))_j}{\|(\mathcal{M}_k(\mathcal{X}))_j\|_F} = V_{z_j^*}$, for $k \in [3]$. Note that the weighted k-means implies that

$$\sum_{j=1}^p \left\|\hat{\mathbf{Y}}_{kj}\right\|_F^2 \left\|(\hat{\mathbf{Y}}_{kj}^s) - \hat{x}_{(\hat{z})_j}\right\|_F^2 \leq M \sum_{j=1}^p \left\|\hat{\mathbf{Y}}_{kj}\right\|_F^2 \left\|(\hat{\mathbf{Y}}_{kj}^s) - V_{z_j^*}\right\|_F^2,$$

where $\hat{z}, \hat{x}_{(\hat{z})_j}$ are the estimated assignment and centroids. Hence, we may bound the term $\left\|(\hat{\mathbf{Y}}_{kj}^s)^T - \hat{x}_{(\hat{z})_j}\right\|_F^2$ by the easier term $\left\|(\hat{\mathbf{Y}}_{kj}^s)^T - V_{z_j^*}\right\|_F^2$. Particularly, we have

$$\sum_{j=1}^p \left\|\hat{\mathbf{Y}}_{kj}\right\|_F^2 \left\|(\hat{\mathbf{Y}}_{kj}^s) - V_{z_j^*}\right\|_F^2 \leq 2\sum_{j=1}^p \|\mathbf{Y}_{kj} - \mathcal{M}_k(\mathcal{X})_j\|_F^2 = 2\left\|\hat{\mathbf{Y}}_k - \mathcal{M}_k(\mathcal{X})\right\|_F,$$

which is bounded in Lemma 1.

2.4 Quantify the number of misclassification (Conjecture)

For simplicity, we ignore the permutation of assignment here. Similarly with Lemma 6 in Gao et al. (2018), let $S = \{j \in [p] : \left\|\hat{x}_{\hat{z}_j} - V_{z_j^*}\right\|_F \geq c\Delta_{\min}\}$. Then, we may have

$$\sum_{j:\ \hat{z}_j \neq z_j^*} \theta_j^2 \leq C \sum_{j\in S} \theta_j^2,$$

for some constant C.

2.5 Assemble

Now, to bound the desire misclassification rate $\sum_{j:\ \hat{z}_j\neq z_j^*} \theta_j^2$, we only need to bound the $\sum_{j\in S} \|(\mathcal{M}_k(\mathcal{X}))_j\|_F^2$.

Note that

$$\begin{aligned}
\sum_{j \in S} \|(\mathcal{M}_k(\mathcal{X}))_j\|_F^2 &\leq C \sum_{j \in S} \left\| \hat{\boldsymbol{Y}}_{kj} \right\|_F^2 + C \sum_{j \in S} \left\| \hat{\boldsymbol{Y}}_{kj} - (\mathcal{M}_k(\mathcal{X}))_j \right\|_F^2 \\
&\leq C \sum_{j \in S} \left\| \hat{\boldsymbol{Y}}_{kj} \right\|_F^2 + C \left\| \hat{\boldsymbol{Y}}_k - (\mathcal{M}_k(\mathcal{X})) \right\|_F^2,
\end{aligned}$$

where the second term is bounded by Lemma 1. For the first term, note that

$$\begin{aligned}
\sum_{j \in S} \left\| \hat{\boldsymbol{Y}}_{kj} \right\|_F^2 &\leq \frac{1}{c^2 \Delta_{\min}^2} \sum_{j \in S} \left\| \hat{\boldsymbol{Y}}_{kj} \right\|_F^2 \left\| \hat{x}_{\hat{z}_j} - V_{z_j^*} \right\|_F^2 \\
&\leq \frac{1}{c^2 \Delta_{\min}^2} \sum_{j \in S} \left\| \hat{\boldsymbol{Y}}_{kj} \right\|_F^2 \left[\left\| (\hat{\boldsymbol{Y}}_{kj}^s) - \hat{x}_{(\hat{z})_j} \right\|_F^2 + \left\| (\hat{\boldsymbol{Y}}_{kj}^s) - V_{z_j^*} \right\|_F^2 \right] \\
&\leq \frac{c'}{\Delta_{\min}^2} \sum_{i=1}^{n} \left\| \hat{\boldsymbol{Y}}_{kj} \right\|_F^2 \left\| (\hat{\boldsymbol{Y}}_{kj}^s) - V_{z_j^*} \right\|_F^2 \\
&\leq C' \left\| \hat{\boldsymbol{Y}}_k - (\mathcal{M}_k(\mathcal{X})) \right\|_F^2,
\end{aligned}$$

following by the facts in section 2.3. Then, we know that the misclassification is bounded by the estimation error with polynomial rate,

$$\sum_{j : \hat{z}_j \neq z_j^*} \theta_j^2 \leq C f(p, r, \delta, M, \Delta_{\min}, \beta, \alpha) r p^2 \alpha \|\theta\|_{\max}^3,$$

with probability at least $1 - p^{-1+C'}$.

Algorithm 1 High-order weighted k-means clustering

Input: Observation $\mathcal{Y} \in \{0,1\}^{p\times\cdots\times p}$, r, relaxation factor in k-means $M > 1$, SCORE normalization function h, tuning parameter τ.

for $k \in [3]$ **do**

  Define $T_\tau(\mathcal{Y}) \in \{0,1\}^{p\times p\times p}$ by replacing the i-th slices on k-th mode of $\mathcal{Y}$ whose ℓ_1 norm is larger than τ by zeros for each $i \in [p]$.

  Solve

$$\hat{\mathbf{Y}}_k = \underset{\operatorname{rank}(\boldsymbol{Y})\le r}{\arg\min} \|\mathcal{M}_k(T_\tau(\mathcal{Y})) - \boldsymbol{Y}\|_F^2$$

  Let $\hat{\mathbf{Y}}_{kj}$ denote the rows of $\hat{\mathbf{Y}}_k$ for $j \in [p]$. Define $S_0 = \{j \in [p] : \left\|\hat{\mathbf{Y}}_{kj}\right\|_F = 0\}$. Set $(z_k^{(0)})_j = 0$ for $j \in S_0$ and obtain the SCORE normalized $\hat{\mathbf{Y}}_k^s$ via $\hat{\mathbf{Y}}_{kj}^s = \frac{\hat{\mathbf{Y}}_{kj}}{h(\hat{\mathbf{Y}}_{kj})}$ for $j \in S_0^c$.

  Find the initial assignment $z_k^{(0)} \in [r]^p$ and centroids $\hat{x}_1, ..., \hat{x}_r \in \mathbb{R}^{p^2}$ such that

$$\sum_{j=1}^{p} h(\hat{\mathbf{Y}}_{kj})^2 \left\|(\hat{\mathbf{Y}}_{kj}^s)^T - \hat{x}_{(z_k^{(0)})_j}\right\|_F^2 \le M \min_{x_1,\dots,x_{r_k}, z_k} \sum_{j=1}^{p} h(\hat{\mathbf{Y}}_{kj})^2 \left\|(\hat{\mathbf{Y}}_{kj}^s)^T - \hat{x}_{(z_k^{(0)})_j}\right\|_F^2$$

end for

Find the average of $z_k^{(0)}, k \in [3]$, $z^{(0)}$.

Output: $\left\{z^{(0)} \in [r]^p\right\}$

References

Gao, C., Ma, Z., Zhang, A. Y., and Zhou, H. H. (2018). Community detection in degree-corrected block models. The Annals of Statistics, 46(5):2153–2185.

Han, R., Luo, Y., Wang, M., and Zhang, A. R. (2020). Exact clustering in tensor block model: Statistical optimality and computational limit. arXiv preprint arXiv:2012.09996.