Estimation of intercept case

Jiaxin Hu

June 15, 2021

1 Intercept case

Suppose the K categories are clustered in R groups based on the space of precision matrix with an intercept matrix, i.e., $\Omega_k^* = \Theta_0 + \sum_{r=1}^R u_{kr}\Theta_r^*$ and $u_{kr} \neq 0$ if k-th category belongs to the r-th group and $u_{kr} = 0$ otherwise. For identifiability, we have $\|u_{\cdot r}\|_F = 1$ and $\sum_{k=1}^K u_{kr} = 0, r \in [R]$. Note that we allow $r = 0$ in this case where $I_0 = \{k \in [K] : u_{kr} = 0, \text{for all } r \in [R]\}$ and $\sum_{r=0}^R |I_r| = K$.

Assume we know the true membership matrix $\boldsymbol{U} = [\![u_{kr}]\!] \in \mathbb{R}^{K\times R}$. Consider the optimizer $\hat\Theta_{0,\lambda}, \hat\Theta_{1,\lambda}, ..., \hat\Theta_{R,\lambda})$ which satisfies

$$(\hat\Theta_{0,\lambda}, \hat\Theta_{1,\lambda}, ..., \hat\Theta_{R,\lambda}) = \underset{\Theta_r, r=0,1,...,R}{\arg\min} \quad \mathcal{L}(\Theta_r, S_k) + \lambda \mathcal{R}(\Theta_r),$$

where $\mathcal{L}$ is denoted as loss function, and $\mathcal{R}$ is the regularization term. Particularly,

$$\mathcal{L}(\Theta_r, S_k) = \sum_{r=0}^R \mathcal{L}_r(\Theta_0, \Theta_r, S_k), \quad \mathcal{R}(\Theta_r) = \sum_{r=1}^R \sum_{k\in I_r} |u_{kr}| \, \|\Theta_r\|_1 + K \|\Theta_0\|_1,$$

where

$$\mathcal{L}_0 = \sum_{k\in I_0} \langle S_k, \Theta_0\rangle - |I_0| \log\det(\Theta_0),$$

$$\mathcal{L}_r = \sum_{k\in I_r} \langle S_k, \Theta_0 + u_{kr}\Theta_r\rangle - |I_r| \log\det(\Theta_0 + u_{kr}\Theta_r), \quad r \in [R].$$

Theorem 1.1. *Suppose* $\|\Theta_r^*\|_0 \leq s$ *and* $\lambda \geq C_\lambda \max_{r\in[R]} \sqrt{\frac{\log p}{n|I_r|}}$. *Let* $\hat\Delta_r = \hat\Theta_{r,\lambda} - \Theta_r^*, r = 0, 1, ..., R$. *Assume* $||u_{kr}| - \frac{1}{\sqrt{|I_r|}}| \leq \epsilon_r$ *with* $\epsilon_r \leq \frac{1}{|I_r|}$ *and* $|I_r| = \mathcal{O}(K)$, *for* $r \in [R]$. *With high probability, we have*

$$\left\|\hat\Theta_{0,\lambda} - \Theta_0^*\right\| = \left\|\hat\Delta_0\right\|_F \leq C_0' \tau^2 \sqrt{\frac{s\log p}{Kn}}, \quad \textit{and} \quad \left\|\hat\Theta_{r,\lambda} - \Theta_r^*\right\| = \left\|\hat\Delta_r\right\|_F \leq C_r' \tau^2 \sqrt{\frac{s\log p}{n}}, r \in [R],$$

Proof. Let $\Delta_r = \hat\Theta_r - \Theta_r^*$, where $\hat\Theta_r$ are arbitrary estimates for $r = 0, 1, ..., R$. We define the function $\mathcal{F}(\Delta_r)$ as the difference between the objective functions with $\hat\Theta_r$ and the true parameter Θ_r^*. Specifically,

$$\mathcal{F}(\Delta_r) = A_1 + A_2 + \lambda A_3,$$

where

$$
\begin{aligned}
A_1 &= \sum_{r=1}^{R} A_{1r}, \quad A_{1r} = \mathcal{L}_r(\Theta_r^* + \Delta_r) - \mathcal{L}_r(\Theta_r^*), \\
A_2 &= \mathcal{L}_0(\Theta_0^* + \Delta_0) - \mathcal{L}_0(\Theta_0^*), \\
A_3 &= \mathcal{R}(\Theta_r^* + \Delta_r) - \mathcal{R}(\Theta_r^*).
\end{aligned}
$$

For $A_{1r}, r \in [R]$**:**

we have

$$
\begin{aligned}
A_{1r} &= \sum_{k \in I_r} \langle S_k, \Delta_0 + u_{kr}\Delta_r \rangle - [\log\det(\Theta_0^* + u_{kr}\Theta_r^* + \Delta_0 + u_{kr}\Delta_r) - \log\det(\Theta_0^* + u_{kr}\Theta_r^*)] \quad (1) \\
&\geq \sum_{k \in I_r} \langle S_k - \Sigma_k, \Delta_0 + u_{kr}\Delta_r \rangle + \frac{1}{4\tau^2} \sum_{k \in I_r} \|\Delta_0 + u_{kr}\Delta_r\|_F^2 .
\end{aligned}
$$

Note that

$$
\begin{aligned}
\frac{1}{4\tau^2} \sum_{k \in I_r} \|\Delta_0 + u_{kr}\Delta_r\|_F^2 &= \frac{1}{4\tau^2} \sum_{k \in I_r} \|\Delta_0\|_F^2 + u_{kr}^2 \|\Delta_r\|_F^2 + 2u_{kr}\langle \Delta_0, \Delta_r \rangle \quad (2) \\
&= \frac{1}{4\tau^2} \left(|I_r| \|\Delta_0\|_F^2 + \|\Delta_r\|_F^2 \right),
\end{aligned}
$$

by the assumption that $\|u_{.r}\|_F = 1$ and $\sum_{k \in I_r}^{K} u_{kr} = 0$ and

$$
|\sum_{k \in I_r} \langle S_k - \Sigma_k, u_{kr}\Delta_r \rangle| \leq \left\| \sum_{k \in I_r} u_{kr}(S_k - \Sigma_k) \right\|_{\max} \|\Delta_r\|_1 .
$$

Since $(S_k - \Sigma_k)$ and $(\Sigma_k - S_k)$ share the same distribution, then $\left\|\sum_{k \in I_r} u_{kr}(S_k - \Sigma_k)\right\|_{\max}$ share the same distribution with $\left\|\sum_{k \in I_r} |u_{kr}|(S_k - \Sigma_k)\right\|_{\max}$. By the assumption that $||u_{kr}| - \frac{1}{\sqrt{|I_r|}}| \leq \epsilon_r$ with $\epsilon_r \leq \frac{1}{|I_r|}$, with high probability, we have

$$
\begin{aligned}
\left\| \sum_{k \in I_r} |u_{kr}|(S_k - \Sigma_k) \right\|_{\max} &\leq \left\| \sum_{k \in I_r} \frac{1}{\sqrt{|I_r|}}(S_k - \Sigma_k) \right\|_{\max} + \sum_{k \in I_r} \epsilon_r \|S_k - \Sigma_k\|_{\max} \quad (3) \\
&\leq C_r' \sqrt{\frac{\log p}{n}} + C_r' |I_r| \epsilon_r \sqrt{\frac{\log p}{n}} \\
&\leq C_r \sqrt{\frac{\log p}{n}},
\end{aligned}
$$

where the last inequality follows by the assumption on ϵ_r. Plugging the inequalities (3) and (2) into A_{1r} (1), we have

$$
A_{1r} \geq \sum_{k \in I_r} \langle S_k - \Sigma_k, \Delta_0 \rangle + \frac{1}{4\tau^2} \left(|I_r| \|\Delta_0\|_F^2 + \|\Delta_r\|_F^2 \right) - C_r \sqrt{\frac{\log p}{n}} \|\Delta_r\|_1 .
$$

For A_2**:**

we have

$$A_2 = \sum_{k\in I_0} \langle S_k, \Delta_0\rangle - |I_0| \left[\log\det(\Theta_0^* + \Delta_0) - \log\det(\Theta_0^*)\right]$$
$$\geq \sum_{k\in I_0} \langle S_k - \Sigma_k, \Delta_0\rangle + \frac{|I_0|}{4\tau^2} \|\Delta_0\|_F^2 .$$

For A_3:

we have

$$A_3 = \sum_{r=1}^{R} \sum_{k\in I_r} |u_{kr}| \left[\|\Theta_r^* + \Delta_r\|_1 - \|\Theta_r^*\|_1\right] + K\left[\|\Theta_0^* + \Delta_0\|_1 - \|\Theta_0^*\|_1\right].$$

By the Lemma 3 in the Supplement of (Negahban et al., 2012), we have

$$\|\Theta_r^* + \Delta_r\|_1 - \|\Theta_r^*\|_1 \geq \left\|\Delta_{r,T_r^\perp}\right\|_1 - \|\Delta_{r,T_r}\|_1 \tag{4}$$

where $T_r = \{(i,j)|\Theta_{r,ij}^* \neq 0\}$, for $r = 0, 1, ..., R$. Plugging the inequality (4) into A_3, we have

$$A_3 \geq \sum_{r=1}^{R} \sum_{k\in I_r} |u_{kr}| \left[\left\|\Delta_{r,T_r^\perp}\right\|_1 - \|\Delta_{r,T_r}\|_1\right] + K\left[\left\|\Delta_{0,T_0^\perp}\right\|_1 - \|\Delta_{0,T_0}\|_1\right]$$
$$\geq C\sum_{r=1}^{R} \sqrt{|I_r|} \left[\left\|\Delta_{r,T_r^\perp}\right\|_1 - \|\Delta_{r,T_r}\|_1\right] + K\left[\left\|\Delta_{0,T_0^\perp}\right\|_1 - \|\Delta_{0,T_0}\|_1\right]$$

for some constant C, and the second inequality follows by the assumption on u_{kr}.

Plug A_1, A_2, A_3 into $\mathcal{F}$,

with high probability, we have

$$\mathcal{F}(\Delta_r) \geq \sum_{k=1}^{K} \langle S_k - \Sigma_k, \Delta_0\rangle + \frac{1}{4\tau^2} K \|\Delta_0\|_F^2 + \sum_{r=1}^{R} \|\Delta_r\|_F^2 - \sum_{r=1}^{R} C_r \sqrt{\frac{\log p}{n}} \|\Delta_r\|_1 + \lambda A_3$$
$$\geq \frac{1}{4\tau^2} K \|\Delta_0\|_F^2 + \sum_{r=1}^{R} \|\Delta_r\|_F^2 - \sum_{r=1}^{R} C_r \sqrt{\frac{\log p}{n}} \|\Delta_r\|_1 - C_0 \sqrt{\frac{K\log p}{n}} \|\Delta_0\|_1 + \lambda A_3.$$

Note that $\|\Delta_r\|_1 \leq \left\|\Delta_{r,T_r^\perp}\right\|_1 + \|\Delta_{r,T_r}\|_1$, and $\lambda \geq C_\lambda \max_{r\in[R]} \sqrt{\frac{\log p}{n|I_r|}}$. For C_λ large enough, we have

$$\mathcal{F}(\Delta_r) \geq B_1 + B_2,$$

where

$$B_1 \geq \frac{1}{4\tau^2} K \|\Delta_0\|_F^2 - C_0' \sqrt{\frac{K\log p}{n}} \|\Delta_{0,T_0}\|_1 \geq \frac{1}{4\tau^2} K \|\Delta_0\|_F^2 - C_0' \sqrt{\frac{Ks\log p}{n}} \|\Delta_{0,T_0}\|_F ,$$
$$B_2 \geq \sum_{r=1}^{R} \left\{ \|\Delta_r\|_F^2 - C_r' \sqrt{\frac{\log p}{n}} \|\Delta_{r,T_r}\|_1 \right\} \geq \sum_{r=1}^{R} \left\{ \|\Delta_r\|_F^2 - C_r' \sqrt{\frac{s\log p}{n}} \|\Delta_{r,T_r}\|_F \right\},$$

where C_0', C_r' are constant independent with K by the assumption that $|I_r| = \mathcal{O}(K)$. To let $\mathcal{F}(\Delta_r) > 0$, it is enough to let $B_1 > 0$ and $B_2 > 0$. Therefore, we have

$$(\Delta_0, \Delta_1, ..., \Delta_R) \in \left\{ \|\Delta_0\|_F \geq C_0' \tau^2 \sqrt{\frac{s \log p}{Kn}} \right\} \times \left\{ \|\Delta_r\|_F \geq C_r' \tau^2 \sqrt{\frac{s \log p}{n}}, r \in [R] \right\},$$

which implies that

$$\left\|\hat{\Delta}_0\right\|_F \leq C_0' \tau^2 \sqrt{\frac{s \log p}{Kn}}, \quad \text{and} \quad \left\|\hat{\Delta}_r\right\|_F \leq C_r' \tau^2 \sqrt{\frac{s \log p}{n}}, r \in [R],$$

with high probability.

□

References

Negahban, S. N., Ravikumar, P., Wainwright, M. J., Yu, B., et al. (2012). A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. *Statistical science*, 27(4):538–557.