Graphic Lasso: Possible Accuracy for Multi-Layer Model

Jiaxin Hu

January 5, 2021

1 A simple extension

Let $Q(\Omega) = \text{tr}(S\Omega) - \log|\Omega|$. Assume the rank of decomposition r is known. Consider the constrained optimization problem

$$\begin{aligned} \min_{\mathcal{C}} \quad & \sum_{k=1}^{K} \left[Q(\Omega^k) \right] \\ s.t. \quad & \Omega^k = \Theta_0 + \sum_{l=1}^{r} u_{lk}\Theta_l, \quad \text{for} \quad k = 1, ..., K, \\ & \|\Theta_l\|_0 \leq b, \quad \text{for} \quad l = 1, ..., r, \\ & \|\Theta_0\|_0 \leq b_0, \end{aligned}$$

where a, b, b_0 are fixed positive constants, $|\cdot|_0$ refers to the vector L_0 norm, and $\|\cdot\|_0$ refers to the matrix L_0 norm. For simplicity, let $\hat{\mathcal{C}} = \{\hat{\Theta}_0, \hat{\Theta}_1, ..., \hat{\Theta}_r, \hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_r\}$ denote the estimation, and $\hat{\Omega}^k = \hat{\Theta}_0 + \sum_{l=1}^{r} \hat{u}_{lk}\hat{\Theta}_l$ for $k = 1, ..., K$.

For true precision matrices Ω^k, let $T^k = \{(j, j')|\omega^k_{j,j'} \neq 0\}$ and $q^k = |T^k|$. Let $T = T^1 \cup \cdots \cup T^k$ and $q = |T|$.

Theorem 1.1. *Suppose two assumptions hold. Let $\{\Omega^k\}$ denote the true precision matrices. For the estimation $\hat{\mathcal{C}}$ such that $\sum_{k=1}^{K} \left[Q(\hat{\Omega}^k) \right] \leq \sum_{k=1}^{K} \left[Q(\Omega^k) \right]$ and satisfies the constrains, the following accuracy bound holds with probability tending to 1.*

$$\sum_{k=1}^{K} \left\| \hat{\Omega}^k - \Omega^k \right\|_F = \mathcal{O}_p \left[\left\{ \frac{(p+q)\log p}{n} \right\}^{1/2} \right].$$

Proof. Let Ω^k denote the true precision matrices for $k = 1, ..., K$. Consider the estimation $\hat{\mathcal{C}}$ such that $\sum_{k=1}^{K} \left[Q(\hat{\Omega}^k) \right] \leq \sum_{k=1}^{K} \left[Q(\Omega^k) \right]$. Let $\Delta^k = \hat{\Omega}^k - \Omega^k$. Define the function

$$G(\left\{\Delta^k\right\}) = \sum_{k=1}^{K} \text{tr}(S(\Omega^k + \Delta^k)) - \text{tr}(\Omega^k) - \log|\Omega^k + \Delta^k| + \log|\Omega^k| = I_1 + I_2,$$

where

$$I_1 = \sum_{k=1}^{K} \text{tr}((S^k - \Sigma^k)\Delta^k), \quad I_2 = \sum_{k=1}^{K} (\tilde{\Delta}^k)^T \int_0^1 (1-v)(\Omega^k + v\Delta^k)^{-1} \otimes (\Omega^k + v\Delta^k)^{-1} dv \tilde{\Delta}^k.$$

With probability tending to 1, we have

$$I_1 \le C_1 \left(\frac{\log p}{n}\right)^{1/2} \sum_{k=1}^{K} \left(|\Delta^k_{T^k}|_1 + |\Delta^k_{T^{k,c}}|_1\right) + C_2 \left(\frac{p \log p}{n}\right)^{1/2} \sum_{k=1}^{K} \left\|\Delta^k\right\|_F, \quad I_2 \ge \frac{1}{4\tau_2^2} \sum_{k=1}^{K} \left\|\Delta^k\right\|_F^2.$$

Note that $|\Delta^k_{T^k}|_1 \le q^{1/2} \left\|\Delta^k\right\|_F$. Then, we only need to deal with $|\Delta^k_{T^{k,c}}|_1$. Rewrite the term, we have

$$|\Delta^k_{T^{k,c}}|_1 = |\hat{\Theta}_{0,T^{k,c}} + \hat{u}_{1k}\hat{\Theta}_{1,T^{k,c}} + \cdots + \hat{u}_{rk}\hat{\Theta}_{r,T^{k,c}}|_1 \le (b_0 + rb) \left\|\Delta^k\right\|_{\max} \le (b_0 + rb) \left\|\Delta^k\right\|_F.$$

Then, by Guo et al, we have

$$\sum_{k=1}^{K} \left\|\Delta^k\right\|_F = \sum_{k=1}^{K} \left\|\hat{\Omega}^k - \Omega^k\right\|_F = \mathcal{O}_p \left[\left\{\frac{(p+q)\log p}{n}\right\}^{1/2}\right]. \tag{1}$$

□

Remark 1. Note that q can be replaced by $\max_k q^k$, where $q^k \le (b_0 + rb)$ for all $k = 1, ..., K$. Also, the accuracy (1) holds when q^k are fixed. Otherwise, the accuracy is of order $\mathcal{O}_p \left[q \left\{\frac{\log p}{n}\right\}^{1/2}\right]$.