

Graphic Lasso: two precision matrices

Jiaxin Hu

January 16, 2021

1 Consistency

Suppose K categories are clustered by two groups with precision matrices Θ_1, Θ_2 . The model becomes

$$\Omega^k = \mathbf{I}_k \Theta_1 + (1 - \mathbf{I}_k) \Theta_2, \quad k = 1, \dots, K,$$

where $\mathbf{I}_k = \mathbf{I}$ (k -th category belongs to group 1) are indicator functions. The model is identifiable since the indicator functions can be replaced by a membership matrix. Consider the optimization problem

$$\begin{aligned} \min_{\Theta_1, \Theta_2, \mathbf{I}_k} \quad & \sum_{k=1}^K \text{tr}(S^k \Omega^k) - \log |\Omega^k| \\ \text{s.t.} \quad & \Omega^k = \mathbf{I}_k \Theta_1 + (1 - \mathbf{I}_k) \Theta_2, \quad k = 1, \dots, K, \\ & \|\Theta_i\|_0 \leq b, \quad i = 1, 2. \end{aligned}$$

Theorem 1.1. *Let $(\Theta_1, \Theta_2, \mathbf{I}_k)$ be the true precision matrices and the membership. Suppose $0 < \tau_1 < \phi_{\min}(\Theta_i) \leq \phi_{\max}(\Theta_0) < \tau_2 < \infty$, where $i = 1, 2$ and τ_1, τ_2 are positive constants. For the estimation $(\hat{\Theta}_1, \hat{\Theta}_2, \hat{\mathbf{I}}_k)$ such that $\sum_{k=1}^K \text{tr}(S^k \hat{\Omega}^k) - \log |\hat{\Omega}^k| \leq \sum_{k=1}^K \text{tr}(S^k \Omega^k) - \log |\Omega^k|$, we have the following accuracy with probability tending to 1*

$$\sum_{k=1}^K \left\| \hat{\Omega}^k - \Omega^k \right\| \leq \sqrt{K} C'' \left[C \sqrt{\frac{b \log p}{n}} + C' \sqrt{\frac{p \log p}{n}} \right]. \quad (1)$$

Proof. Let Σ^1, Σ^2 denote the true covariance matrices. Define the sets $A_{11} = \{k : \hat{I}_k = I_k = 1\}$, $A_{12} = \{k : \hat{I}_k = 1, I_k = 0\}$, $A_{21} = \{k : \hat{I}_k = 0, I_k = 1\}$ and $A_{22} = \{k : \hat{I}_k = I_k = 0\}$. Correspondingly, we define $\Delta_{11} = \hat{\Theta}_1 - \Theta_1$, $\Delta_{12} = \hat{\Theta}_1 - \Theta_2$, $\Delta_{21} = \hat{\Theta}_2 - \Theta_1$, and $\Delta_{22} = \hat{\Theta}_2 - \Theta_2$. Let $\Delta^k = \hat{\Omega}^k - \Omega \in \{\Delta_{11}, \Delta_{12}, \Delta_{21}, \Delta_{22}\}$. The value of Δ^k depends on the true and estimated membership of k . Consider the function

$$G(\Delta_{11}, \Delta_{12}, \Delta_{21}, \Delta_{22}) = I_1 + I_2,$$

where

$$\begin{aligned} I_1 &= \sum_{k \in A_{11}} \text{tr}((S^k - \Sigma^1) \Delta_{11}) + \sum_{k \in A_{12}} \text{tr}((S^k - \Sigma^2) \Delta_{12}) + \sum_{k \in A_{21}} \text{tr}((S^k - \Sigma^1) \Delta_{21}) + \sum_{k \in A_{22}} \text{tr}((S^k - \Sigma^2) \Delta_{22}) \\ &= I_{11} + I_{12} + I_{21} + I_{22}, \end{aligned}$$

and

$$I_2 = |A_{11}|f(\Delta_{11}, \Theta_1) + |A_{12}|f(\Delta_{12}, \Theta_2) + |A_{21}|f(\Delta_{21}, \Theta_1) + |A_{22}|f(\Delta_{22}, \Theta_2),$$

with $f(\Delta, \Theta) = (\tilde{\Delta})^T \int_0^1 (1-v)(\Theta + v\Delta)^{-1} \otimes (\Theta + v\Delta)^{-1} dv \tilde{\Delta}$.

Recall the result in the common precision matrix case. For each $I_{ij}, i, j = 1, 2$, we have

$$\frac{1}{|A_{ij}|} |I_{ij}| = \text{tr} \left(\left(\frac{1}{|A_{ij}|} \sum_{k \in A_{ij}} S^k - \Sigma^j \right) \Delta_{ij} \right) \leq C_{ij} \sqrt{\frac{\log p}{n|A_{ij}|}} |\Delta_{ij}^-|_1 + C'_{ij} \sqrt{\frac{p \log p}{n|A_{ij}|}} \|\Delta_{ij}\|_F.$$

Let $T_j = \{(k, l) : \Theta_{j,kl} \neq 0\}, j = 1, 2$. We have $|\Delta_{ij}^-|_1 = |\Delta_{T_j, ij}^-|_1 + |\Delta_{T_j^c, ij}^-|_1$. Note that $|\Delta_{T_j, ij}^-|_0, |\Delta_{T_j^c, ij}^-|_0 \leq b$ and $|\Delta_{T_j, ij}^-|_1, |\Delta_{T_j^c, ij}^-|_1 \leq \sqrt{b} \|\Delta_{ij}\|_F$. Then, we have

$$|I_{ij}| \leq \sqrt{|A_{ij}|} \left[C_{ij} \sqrt{\frac{b \log p}{n}} + C'_{ij} \sqrt{\frac{p \log p}{n}} \right] \|\Delta_{ij}\|_F.$$

On the other hand, the lower bound for I_2 is

$$I_2 \leq \frac{1}{4\tau_2^2} \sum_{ij} |A_{ij}| \|\Delta_{ij}\|_F^2.$$

To let $G \leq 0$, we have $I_2 \leq |I_1| \leq \sum_{ij} |I_{ij}|$. Plug the upper bound for $|I_{ij}|$ and the lower bound for I_2 , we have

$$\frac{1}{4\tau_2^2} \sum_{ij} |A_{ij}| \|\Delta_{ij}\|_F^2 \leq \left[C \sqrt{\frac{b \log p}{n}} + C' \sqrt{\frac{p \log p}{n}} \right] \sum_{ij} \sqrt{|A_{ij}|} \|\Delta_{ij}\|_F.$$

By Cauchy Schwartz inequality, we have

$$\sum_{ij} |A_{ij}| \|\Delta_{ij}\|_F^2 \geq \frac{1}{4} \left(\sum_{ij} \sqrt{|A_{ij}|} \|\Delta_{ij}\|_F \right)^2.$$

Thus, we have

$$\sum_{ij} \sqrt{|A_{ij}|} \|\Delta_{ij}\|_F \leq 4C'' \left[C \sqrt{\frac{b \log p}{n}} + C' \sqrt{\frac{p \log p}{n}} \right].$$

Multiply $\max \sqrt{|A_{ij}|}$ on both sides. We obtain the accuracy

$$\begin{aligned} \sum_{k=1}^K \left\| \hat{\Omega}^k - \Omega^k \right\|_F &= \sum_{ij} |A_{ij}| \|\Delta_{ij}\|_F \\ &\leq \max \sqrt{|A_{ij}|} \sum_{ij} \sqrt{|A_{ij}|} \|\Delta_{ij}\|_F \\ &\leq 4 \max \sqrt{|A_{ij}|} C'' \left[C \sqrt{\frac{b \log p}{n}} + C' \sqrt{\frac{p \log p}{n}} \right]. \end{aligned}$$

□

Remark 1. In two group case with equal group size, we have $\max \sqrt{|A_{ij}|} \leq \sqrt{\frac{K}{2}}$. Then, we obtain the accuracy (1) in Theorem 1.1. If we have r groups and each group has equal number of categories, the number 4 should be replaced by $r(r-1)$ and $\max \sqrt{|A_{ij}|} \leq \sqrt{\frac{K}{r}}$. Thus the accuracy is of order $\mathcal{O}(\sqrt{K}r^{3/2})$.