

Graphic Lasso: Accuracy for Precision matrix estimation

Jiaxin Hu

February 27, 2021

1 Penalized Optimization

Consider the following penalized optimization problem.

$$\begin{aligned} \min_{\mathbf{U}, \Theta^l} \quad & \sum_{k=1}^K \langle S^k, \Omega^k \rangle - \log \det(\Omega^k) + \lambda \|\Omega^k\|_1, \\ \text{s.t.} \quad & \Omega^k = \sum_{l=1}^r u_{kl} \Theta^l, \quad k = 1, \dots, K \\ & \lambda > 0, \quad \mathbf{U} \text{ is a membership matrix.} \end{aligned}$$

Theorem 1.1 (Accuracy for precision matrix estimation with unknown hard membership). *Let $\{\mathbf{U}, \Theta^l\}$ denote the true membership matrix and the true precision matrices. Suppose $0 < \tau_1 < \min_{l \in [r]} \phi_{\min}(\Theta^l) \leq \max_{l \in [r]} \phi_{\max}(\Theta^l) < \tau_2 < \infty$, where $\phi(\cdot)$ is the singular-value of a matrix and τ_1, τ_2 are positive constants. Assume the sample size for k -th category is $n_k = n, k \in [K]$. Consider the estimation $\{\hat{\mathbf{U}}, \hat{\Theta}^k\}$ with smaller objective value than true parameters. Suppose $\lambda \leq C \sqrt{\frac{\log p}{nK}}$ for some constant C , we have the following accuracy with probability tending to 1*

Special case $K=1 \rightarrow$ should reduce to classical graphical lasso $(p+q)\log p$

$$\sum_{k=1}^K \|\hat{\Omega}^k - \Omega^k\|_F \leq 4\tau_2^2 r^2 \sqrt{K} C \sqrt{\frac{p \log p}{n}}. \quad (1)$$

Proof. First, we introduce some notations.

1. Let $\mathcal{L}(\mathbf{U}, \Theta^l) = \sum_{k=1}^K \langle S^k, \Omega^k \rangle - \log \det(\Omega^k) + \lambda \|\Omega^k\|_1$, where $\Omega^k = \sum_{l=1}^r u_{kl} \Theta^l, k = 1, \dots, K$.
2. Let D denote the confusion matrix between the estimation $\{\hat{\mathbf{U}}, \hat{\Theta}^k\}$ and the true parameters, in which $D_{al} = \sum_{k=1}^K \mathbf{I}\{u_{ka} = \hat{u}_{kl} = 1\}$.
3. Let I_l denote the index set of the categories in the k -th group based on the parameter \mathbf{U} , i.e., $I_l = \{k : u_{kl} = 1\}$. Let \hat{I}_l denote the estimated set by $\hat{\mathbf{U}}$. Note that $|\hat{I}_l \cap I_a| = D_{al}$.

Then, we define the function $G(\hat{\mathbf{U}}, \hat{\Theta}^k)$ (not the G in the clustering accuracy proof! This is the G in Guo's paper), where

$$\begin{aligned} G(\hat{\mathbf{U}}, \hat{\Theta}^k) &= \mathcal{L}(\hat{\mathbf{U}}, \hat{\Theta}^l) - \mathcal{L}(\mathbf{U}, \Theta^l) \\ &= \sum_{l=1}^r \sum_{a=1}^r \sum_{k \in \hat{I}_l \cap I_a} \langle S^k, \hat{\Theta}^l \rangle - \langle S^k, \Theta^a \rangle - \log \det(\hat{\Theta}^l) + \log \det(\Theta^a) + \lambda \|\hat{\Theta}^l\|_1 - \lambda \|\Theta^a\|_1 \\ &= \sum_{l=1}^r \sum_{a=1}^r G_{al}(\hat{\mathbf{U}}, \hat{\Theta}^l). \end{aligned}$$

Define $\Delta_{al} = \hat{\Theta}^l - \Theta^a, l \in [r], a \in [r]$. Consider an arbitrary pair of a, l . By Taylor Expansion, we have

$$G_{al}(\hat{\mathbf{U}}, \hat{\Theta}^l) = A_{al,1} + A_{al,2} + A_{al,3},$$

where

$$\begin{aligned} A_{al,1} &= \sum_{k \in \hat{I}_l \cap I_a} \langle S^k - \Sigma^a, \Delta_{al} \rangle, \\ A_{al,2} &= D_{al}(\text{vec}(\Delta_{al}))^T \int_0^1 (1-v)(\Theta^a + v\Delta_{al})^{-1} \otimes (\Theta^a + v\Delta_{al})^{-1} dv \text{vec}(\Delta_{al}), \\ A_{al,3} &= \lambda D_{al} \left(\left\| \hat{\Theta}^l \right\|_1 - \left\| \Theta^a \right\|_1 \right). \end{aligned}$$

According to the proofs in Note 0115 and Note 0113, we have the upper bound

$$|A_{al,1}| \leq \sqrt{D_{al}} C \sqrt{\frac{p \log p}{n}} \|\Delta_{al}\|_F, \quad (2)$$

where the constant C is related to the τ_2 by Lemma 1 in A.J. Rothman et al, and the lower bound

Is C related to the ground truth sparsity level?

$$A_{al,2} \geq \frac{1}{4\tau_2^2} D_{al} \|\Delta_{al}\|_F^2. \quad (3)$$

Also, note that

$$A_{al,3} \leq \lambda D_{al} \left\| \hat{\Theta}^l - \Theta^a \right\|_1 \leq \lambda D_{al} \sqrt{p} \|\Delta_{al}\|_F. \quad (4)$$

Since the estimation $\{\hat{\mathbf{U}}, \hat{\Theta}^k\}$ has smaller objective value than true parameters, we have

$$G(\hat{\mathbf{U}}, \hat{\Theta}^k) \leq 0,$$

which implies that

$$\sum_{l=1}^r \sum_{a=1}^r A_{al,2} \leq \sum_{l=1}^r \sum_{a=1}^r |A_{al,1}| - A_{al,3}.$$

Plugging the bounds (2), (3), and (4), we have

$$\frac{1}{4\tau^2} \sum_{l=1}^r \sum_{a=1}^r D_{al} \|\Delta_{al}\|_F^2 \leq C \sum_{l=1}^r \sum_{a=1}^r \sqrt{D_{al}} \sqrt{\frac{p \log p}{n}} \|\Delta_{al}\|_F - \lambda \sqrt{p} \sum_{l=1}^r \sum_{a=1}^r D_{al} \|\Delta_{al}\|_F. \quad (5)$$

Note that the inequality (5) makes sense only when the right hand side is larger than 0. Hence, we consider the constrain

$$\lambda \leq C \sqrt{\frac{\log p}{\max_{a,l \in [r]} D_{al} n}}. \quad (6)$$

By Cauchy Schwartz inequality, we have

overall approach is reasonable.

$$\sum_{l=1}^r \sum_{a=1}^r D_{al} \|\Delta_{al}\|_F^2 \geq \frac{1}{r^2} \left(\sum_{l=1}^r \sum_{a=1}^r \sqrt{D_{al}} \|\Delta_{al}\|_F \right)^2. \quad (7)$$

Then, plugging the inequality (7) and the constrain (6) into the inequality (5), we obtain that

$$\sum_{l=1}^r \sum_{a=1}^r \sqrt{D_{al}} \|\Delta_{al}\|_F \leq 4r^2 \tau_2^2 C' \sqrt{\frac{p \log p}{n}}.$$

Then, we have the accuracy for precision matrices

$$\sum_{k=1}^K \left\| \hat{\Omega}^k - \Omega^k \right\|_F = \sum_{l=1}^r \sum_{a=1}^r D_{al} \|\Delta_{al}\|_F \leq 4r^2 \tau_2^2 C' \max_{a,l \in [r]} \sqrt{D_{al}} \sqrt{\frac{p \log p}{n}}.$$

Note that $\max_{a,l \in [r]} \sqrt{D_{al}} \leq \sqrt{K}$. Replacing $\max_{a,l \in [r]} \sqrt{D_{al}}$ by \sqrt{K} , we obtain the constrain for λ and the accuracy rate (1) in the Theorem. \square