Graphic Lasso: Accuracy with intercept

Jiaxin Hu

April 11, 2021

Consider the model

$$\Omega_k = \Theta_0 + \sum_{l=1}^{r} u_{kl}\Theta_l, \quad k \in [K].$$

Let $U = [\![u_{kl}]\!] \in \mathbb{R}^{K\times r}$ be the membership matrix and u_l denote the l-th column of U. The optimization problem is stated as

$$\begin{aligned}
\min_{\{U,\Theta\}} \quad & \mathcal{L}(U,\Theta) = \sum_{k=1}^{K} \langle S^k, \Omega^k\rangle - \log\det(\Omega^k),\\
s.t. \quad & \Omega^k = \Theta_0 + u_k\Theta_1, \quad k = 1, ..., K,\\
& \|U\|_F = 1, \sum_{k=1}^{K} u_k = 0,
\end{aligned}$$

where Θ_0, Θ_1 are positive definite and $\tau_1 < \min\{\varphi_{\min}(\Theta_0), \varphi_{\min}(\Theta_1)\} \leq \max\{\varphi_{\max}(\Theta_0), \varphi_{\max}(\Theta_1)\} < \tau_2, \tau_1, \tau_2 > 0$.

Lemma 1 (Accuracy with intercept)**.** *Let $\{U, \Theta_l\}$ denote the true parameter. Suppose the estimate $\{\hat{U}, \hat{\Theta}_l\}$ satisfies $\mathcal{L}(\hat{U}, \hat{\Theta}_l) \leq \mathcal{L}(U, \Theta_l)$. Then, with probability tends to 1, we have the accuracy rate*

$$\sum_{k=1}^{K} \|\Delta_k\|_F \leq KC\sqrt{\frac{\log p}{n}}.$$

Remark 1. Note that by matrix inverse lemma, we have

$$\Sigma^k = (\Theta_0 + u_k\Theta_1)^{-1} = \Theta_0^{-1} + \frac{u_k}{1 + u_k\langle\Theta_0^{-1}, \Theta_1\rangle}\Theta_0^{-1}\Theta_1\Theta_0^{-1}.$$

Since $\sum_{k=1}^{K} \frac{u_k}{1+u_k\langle\Theta_0^{-1},\Theta_1\rangle}$ may not equal to 0, the true covariance matrices do not share the same pattern and we can not use all sample covariance matrices S^k to improve the estimation error from K to $\sqrt{K}$.

Proof. Define the function

$$G(\hat{U}, \hat{\Theta}_l) = \sum_{k=1}^{K} \langle S^k, \hat{\Theta}_0 + \hat{u}_k\hat{\Theta}_1 - \Theta_0 - u_k\Theta_1\rangle - \log\det(\hat{\Theta}_0 + \hat{u}_k\hat{\Theta}_1) + \log\det(\Theta_0 + u_k\Theta_1).$$

Let $\Delta_k = \hat{\Theta}_0 + \hat{u}_k\hat{\Theta}_1 - \Theta_0 - u_k\Theta_1$. By Taylor Expansion, we have

$$-\log\det(\hat{\Theta}_0 + \hat{u}_k\hat{\Theta}_1) + \log\det(\Theta_0 + u_k\Theta_1) \geq -\langle(\Theta_0 + u_k\Theta_1)^{-1}, \Delta_k\rangle + \frac{1}{4\tau_2^2}\|\Delta_k\|_F^2.$$

Let $\Sigma^k = (\Theta_0 + u_k\Theta_1)^{-1}$ denote the true precision matrix. Then, we have

$$G(\hat{U}, \hat{\Theta}_l) \geq \sum_{k=1}^{K} \langle S^k - \Sigma^k, \Delta_k \rangle + \frac{1}{4\tau_2^2} \left\| \Delta_k \right\|_F^2 = I_1 + I_2.$$

By Lemma 1 in A.J. Rothman et al, we have

$$\max |S_{jk}^k - \Sigma^{jk}| \leq C\sqrt{\frac{\log p}{n}},$$

with high probability. Therefore, we have

$$0 \geq \frac{1}{4\tau_2^2} \frac{1}{K} (\sum_{k=1}^{K} \left\| \Delta_k \right\|_F)^2 - C\sqrt{\frac{\log p}{n}} \sum_{k=1}^{K} \left\| \Delta_k \right\|_F ,$$

which implies that

$$\sum_{k=1}^{K} \left\| \Delta_k \right\|_F \leq KC\sqrt{\frac{\log p}{n}}.$$

$\square$