Graphic Lasso: Accuracy with intercept

Jiaxin Hu

April 19, 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 $\tau_1 < \min\{\varphi_{\min}(\Theta_0), \varphi_{\min}(\Theta_1)\} \le \max\{\varphi_{\max}(\Theta_0), \varphi_{\max}(\Theta_1)\} < \tau_2, \tau_1, \tau_2 > 0$.

Lemma 1. *Let $Z_i \sim_{i.i.d.} \mathcal{N}(0,\Sigma)$ and $\phi_{max}(\Sigma) \le \tau < \infty$. Let $\Sigma = [\![\Sigma_{ij}]\!]$, then*

$$P\left(\left| \sum_{i=1}^{n} Z_{ij}Z_{ik} - n\Sigma_{jk} \right| \ge n\nu \right) \le c_1 e^{-c_2 n\nu^2}, \quad for \quad |\nu| \le \delta,$$

where c_1, c_2, δ depends on τ only.

1 Conjectures

Lemma 2 (Conjecture)**.** *Consider the random variables $Z_i^k \sim_{i.i.d.} \mathcal{N}(0,\Sigma_1), i = [n], k \in [K]$, where Σ_1 are positive definite with bounded singular values. Then for a non-negative sequence $c_k \ge 0, k \in [K]$, we have*

—> c_k n? sum of squares of c_k

$$P\left(\left| \sum_{k=1}^{K} \sum_{i=1}^{n} \left[c_k Z_{ij}^k Z_{il}^k - c_k \Sigma_{1,jl} \right] \right| \ge nK\nu \right) \le C_1 \exp\left(-C_2 nK v^2\right),$$

K —> sum of squares of c_k

for some v small enough.

Theorem 1.1 (Accuracy for intercept case (Conjecture))**.** *Let $\{\Theta_0, \Theta_1, U\}$ denote the true parameters. There exists a local minimizer $\{\hat{\Theta}_0, \hat{\Theta}_1, \hat{U}\}$ satisfies*

$$\max\left\{ \left\|\hat{\Theta} - \Theta\right\|_F, \left\|\hat{\Theta}_1 - \Theta_1\right\|_F, \max_{k\in[K]} |\hat{u}_k - u_k| \right\} = \mathcal{O}\left(\sqrt{\frac{p^2 \log p}{nK}} \right),$$

Marix decomposition (pertubation)

Sigma = [vec(Sigma_1),…vec(Sigma_k)]
(model) = [1, mu] *[vec(theta0), vec(theta_1)];

sample: Sigma hat

y_k sim N(0, theta0+mu_k theta1), k=1,…K

accuracy for log-likelihood estimate of (theta0,theta1,k);

(1) theta0 non-parallel to theta1
(2) sum of mu_k = 0

Proof. Define the function

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

Davis-Khanh matrix pertubation:
P = A*diag*B =LU
P+pertubation = A'diag'* B'=LU
Conclusion: llA-A'Qll_F and llB-B'Qll_F < = spectral norm of
(pertubation matrix)/min(non-zero singular value)

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

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

Consider the set $\mathcal{A} = \left\{ (\hat{U}, \hat{\Theta}_1, \hat{\Theta}_0) : \|\Delta\|_F \leq M_1 \sqrt{\frac{p^2 \log p}{nK}}, \|\Delta_1\|_F \leq \gamma_1, \max_{k \in [K]} |\hat{u}_k - u_k| \leq \gamma_2 \right\}$, where $\gamma_1, \gamma_2 = o\left(\sqrt{\frac{p^2 \log p}{nK}}\right)$. Let $\partial \mathcal{A}$ denote the boundary of $\mathcal{A}$. Therefore, we only need to prove $G(\hat{u}, \hat{\Theta}) > 0$ for the estimates $\{\hat{u}, \hat{\Theta}\} \in \partial \mathcal{A}$.

For I_1, let $\Delta = \hat{\Theta}_0 - \Theta_0$, $\Delta_1 = \hat{\Theta}_1 - \Theta_1$. Then, we have

$$\Delta_k = \Delta + u_k \Delta_1 + (\hat{u}_k - u_k) \hat{\Theta}_1.$$

Then, we have

$$\begin{aligned}
|I_1| &= |\sum_{k=1}^{K} \langle S^k - \Sigma^k, \Delta_k \rangle| \\
&\leq |\sum_{k=1}^{K} \langle S^k - \Sigma^k, \Delta \rangle| + |\sum_{k=1}^{K} \langle S^k - \Sigma^k, u_k \Delta_1 \rangle| + |\sum_{k=1}^{K} \langle S^k - \Sigma^k, (\hat{u}_k - u_k) \hat{\Theta}_1 \rangle| \\
&\leq |\langle \sum_{k=1}^{K} S^k - \Sigma^k, \Delta \rangle| + |\langle \sum_{k=1}^{K} S^k - \Sigma^k, \Delta_1 \rangle| + \max_{k \in [K]} |(\hat{u}_k - u_k)| |\langle \sum_{k=1}^{K} S^k - \Sigma^k, \hat{\Theta}_1 \rangle|.
\end{aligned}$$

Note that

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

Let $\Sigma_k = \Sigma_0 + c_k \Sigma_1$, where

$$\Sigma_0 = \Theta_0^{-1} + \min_{k \in [K]} \frac{u_k}{1 + u_k \langle \Theta_0^{-1}, \Theta_1 \rangle} \Theta_0^{-1} \Theta_1 \Theta_0^{-1}, \quad \Sigma_1 = \Theta_0^{-1} \Theta_1 \Theta_0^{-1}$$

, and

$$c_k = \frac{u_k}{1 + u_k \langle \Theta_0^{-1}, \Theta_1 \rangle} - \min_{k \in [K]} \frac{u_k}{1 + u_k \langle \Theta_0^{-1}, \Theta_1 \rangle}.$$

Now, consider random variable $Y_i^k = X_i^k + \sqrt{c_k} Z_i^k \sim_{i.i.d.} \mathcal{N}(0, \Sigma_0 + c_k \Sigma_1)$, where $X_i^k \sim_{i.i.d.} \mathcal{N}(0, \Sigma_0)$, $Z_i^k \sim_{i.i.d.} \mathcal{N}(0, \Sigma_1)$ and X_i^k is independent with Z_i^k. Then, we have

$$\frac{1}{K} \sum_{k=1}^{K} S_{ab}^k - \Sigma_{0,ab} - c_k \Sigma_{1,ab} = \frac{1}{nK} \sum_{k=1}^{K} \sum_{i=1}^{n} Y_{ia}^k Y_{ib}^k - Y_{.a}^k Y_{.b}^k - \Sigma_{0,ab} - c_k \Sigma_{1,ab}.$$

Note that $Y^k_{.a} \rightarrow_{a.s.} 0$, and

$$\begin{aligned} Y^k_{ia} Y^k_{ib} &= [X^k_{ia} + \sqrt{c_k} Z^k_{ia}][X^k_{ib} + \sqrt{c_k} Z^k_{ib}] \\ &= X^k_{ia} X^k_{ib} + c_k Z^k_{ia} Z^k_{ib} + \sqrt{c_k} X^k_{ia} Z^k_{ib} + \sqrt{c_k} Z^k_{ia} X^k_{ib}, \end{aligned}$$

where

$$\frac{1}{n}\sum_{i=1}^{n} X^k_{ia} Z^k_{ib} \rightarrow_{a.s.} 0, \quad \frac{1}{n}\sum_{i=1}^{n} Z^k_{ia} X^k_{ib} \rightarrow_{a.s.} 0.$$

Hence, with high probability, we have

$$\begin{aligned} |\frac{1}{K}\sum_{k=1}^{K} S^k_{ab} - \Sigma_{0,ab} - c_k \Sigma_{1,ab}| &= |\frac{1}{K}\sum_{k=1}^{K} X^k_{ia} X^k_{ib} + c_k Z^k_{ia} Z^k_{ib} - \Sigma_{0,ab} - c_k \Sigma_{1,ab}| \\ &\le |\frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{n} X^k_{ia} X^k_{ib} - \Sigma_{0,ab}| + |\frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{n} c_k Z^k_{ia} Z^k_{ib} - c_k \Sigma_{1,ab}|, \\ &\le C\sqrt{\frac{\log p}{nK}}, \end{aligned}$$

where C is a constant and the last inequality follows by the lemmas 1 and 2.
Therefore, we have

$$|I_1| \le C\sqrt{K}\sqrt{\frac{p^2 \log p}{n}} \left[\|\Delta\|_F + \|\Delta_1\|_F + \max_{k\in[K]} |(\hat{u}_k - u_k)| \right].$$

For I_2, consider the estimate $\{\hat{\Theta}_0, \hat{\Theta}_1, \hat{U}\} \in \partial\mathcal{A}$. By triangle inequality, we have

$$\|\Delta\|_F - \|\Delta_1\|_F - \max_{k\in[K]} |(\hat{u}_k - u_k)|\tau_2 \le \|\Delta_k\|_F \le \|\Delta\|_F + \|\Delta_1\|_F + \max_{k\in[K]} |(\hat{u}_k - u_k)|\tau_2,$$

and thus $\|\Delta_k\|_F \asymp \|\Delta\|_F$.

Therefore, for the estimate $\{\hat{\Theta}_0, \hat{\Theta}_1, \hat{U}\} \in \partial\mathcal{A}$, we have

$$\begin{aligned} G(\hat{\Theta}_0, \hat{\Theta}_1, \hat{U}) &\ge I_2 - |I_1| \\ &\ge \frac{C''}{4\tau_2^2} K \|\Delta\|_F - C'\sqrt{K}\sqrt{\frac{p^2 \log p}{n}} [\|\Delta\|_F] \\ &\ge \frac{C'' M^2}{4\tau_2^2} \frac{p^2 \log p}{n} - C' M \frac{p^2 \log p}{n} \\ &> 0, \end{aligned}$$

for M large enough.

$\square$