Graphic Lasso: Accuracy with intercept

Jiaxin Hu

April 20, 2021

1 Problem statement

1.1 Original problem

Consider the model

$$\Omega_k = \Theta_0 + u_k\Theta_1, \quad k \in [K]. \tag{1}$$

Let $u = [\![u_k]\!]$. The parameters $\{\Theta_0, \Theta_1, u\}$ should satisfy the following conditions for identifiability.

1. Θ_0, Θ_1 are positive definite and $\tau_1 < \min\{\varphi_{\min}(\Theta_0), \varphi_{\min}(\Theta_1)\} \leq \max\{\varphi_{\max}(\Theta_0), \varphi_{\max}(\Theta_1)\} < \tau_2, \tau_1, \tau_2 > 0$;
2. $\Theta_0 \neq C\Theta_1$ for all $C \in \mathbb{R}$;
3. $\|u\|_F = 1$ and $\sum_{k=1}^K u_k = 0$.

Consider the MLE $\{\hat{\Theta}_0, \hat{\Theta}_1, \hat{u}\}$ which is the solution (local minimizer) to the following optimization program.

$$\begin{aligned} \min_{\{U,\Theta\}} \quad & \mathcal{L}(U,\Theta) = \sum_{k=1}^K \langle S^k, \Omega^k \rangle - \log\det(\Omega^k), \\ s.t. \quad & \Omega^k = \Theta_0 + u_k\Theta_1, \quad k \in [K], \\ & \{\Theta_0, \Theta_1, u\} \text{ satisfy the identifiability conditions.} \end{aligned}$$

Let $\Delta_k = \hat{\Omega}_k - \Omega_k = \hat{\Theta}_0 + \hat{u}_k\hat{\Theta}_1 - \Theta_0 - u_k\Theta_1$.

Goal: Our goal is to find the accuracy rate for the Frobenius norm $\sum_{k=1}^K \|\Delta_k\|_F$.

1.2 Equivalent matrix decomposition problem

Rewrite the model (1) into the matrix form. We have the equivalent model

$$\Omega = \Theta U^T,$$

where

$$\Omega = [\mathrm{vec}(\Omega_1), ..., \mathrm{vec}(\Omega_K)]_{p^2 \times K}, \quad \Theta = [\mathrm{vec}(\Theta_0), \mathrm{vec}(\Theta_1)]_{p^2 \times 2}, \quad U = \begin{bmatrix} 1 & u_1 \\ \vdots & \vdots \\ 1 & u_K \end{bmatrix}_{K \times 2}.$$

Consider the same MLE $\{\hat{\Theta}_0, \hat{\Theta}_1, \hat{u}\}$ estimate. Note that

$$\left\|\Omega - \hat{\Omega}\right\|_F = \sqrt{\sum_{k=1}^{K} \|\Delta_k\|_F^2}.$$

By Cauchy-Schwartz, we have $\sum_{k=1}^{K} \|\Delta_k\|_F^2 \geq \frac{1}{K}\left(\sum_{k=1}^{K} \|\Delta_k\|_F\right)^2$. Then, we know that

$$\sum_{k=1}^{K} \|\Delta_k\|_F \leq \sqrt{K}\left\|\Omega - \hat{\Omega}\right\|_F.$$

Goal: Our goal is to find the accuracy rate for the Frobenius norm $\sqrt{K}\left\|\Omega - \hat{\Omega}\right\|_F$.

1.3 What we already know

how does their result compare to d.f.?
Consistent to our earlier conjecture?

From Soloveychik's paper, we know the following things.

1. Let $\Sigma = [\mathrm{vec}(\Omega_1^{-1}), ..., \mathrm{vec}(\Omega_k^{-1})]$. We already know the accuracy rate of $\left\|\Sigma - \tilde{\Sigma}\right\|_F$, where $\tilde{\Sigma}$ is the least-square estimate of Σ under the assumptions

   $$\frac{\sigma_1(\Sigma)}{\sigma_2(\Sigma)} \leq \kappa\sqrt{K}, \quad \sigma_2(\Sigma) \geq \epsilon > 0.$$

   Their results:
   || Sigma - \hat Sigma ||^2_F <= O(r(K + p^2)/n),

   when K << p^2, the result consistent to our earlier conjecture.

2. Let $S = [\mathrm{vec}(S_1), ..., \mathrm{vec}(S_K)]$ and $R = S - \Sigma$. We know the bound for the residual

   $$\|R\|_2 \leq \max_k \sigma_{\max}(\Sigma_k)\left(\sqrt{\frac{2}{n}} + C\sqrt{\frac{p^2}{nK}}\right).$$

   This lemma (Lemma 6 in the paper) may help the previous conjectured lemma 1.

Two intermediate steps:
1. Use their objective function, but define Sigma = [vec(Omega_1), ..., vec(Omega_k)] (without inverse
). What is the accuracy?

1.4 Problems

1. Soloveychik's paper estimates the covariance matrices while we estimate the precision matrices. Though our precision matrices and covariance matrices both are structured, we can not apply the covariance estimation results to precision matrices. Hence, we can not apply the Theorem 1, Lemma 3,4,5 under our model.

   #2 is easy to handle. We can combine accuracy for ||\hat U - O U|| (if we had proved) and identifiability to obtain l_{2,\infty} norm error bound.

2. Soloveychik's paper seems do not worry about the identifiability. In the proof of Lemma 5, they use bounds for $\left\|\hat{U} - U\right\|_F$, $\left\|\hat{W} - W\right\|_F$, where $U, \hat{U}, W, \hat{W}$ are SVD components. Though the SVD in the paper requires the descent order of the singular values, the paper does not consider the rotation issues when there are singular values with multiple singular vectors.

3. Should we stick to the previous proof? Or should we give up previous proof and try to find similar proof in Soloveychik's paper?

Perhaps combine both? or even more? Ultimate goal is to prove our theorem.
We should try whatever path works. (However, we never know which works if

2 Conjectures

Lemma 1 (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* $c_k \geq 0, k \in [K]$*, we have*

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

for some v *small enough.*