Graphic Lasso: Common precision matrix

Jiaxin Hu

January 13, 2021

1 Consistency

Suppose K categories share the same precision matrix Θ_0. Consider the constrained optimization problem

$$\begin{aligned} &\min_{\Theta} \sum_{k=1}^{K} \operatorname{tr}(S^k\Theta) - \log|\Theta| \\ &s.t. \quad \|\Theta\|_0 \leq b_0, \end{aligned}$$

where S^k is the sample covariance matrix for k-th category with sample size n, $\|\cdot\|_0$ is the number of non-zero elements in the matrix.

Before the theorem, here is a useful lemma for the proof.

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

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

where c_1, c_2, δ depends on τ only.

Proof. See Lemma 1 of Rothman et.al. □

Theorem 1.1. *Let Θ_0 be the true precision matrix. Suppose $0 < \tau_1 < \phi_{min}(\Theta_0) \leq \phi_{max}(\Theta_0) < \tau_2 < \infty$, where τ_1, τ_2 are positive constants. For the estimate such that $\sum_{k=1}^{K} tr(S^k\hat{\Theta}_0) - \log|\hat{\Theta}_0| \leq \sum_{k=1}^{K} tr(S^k\Theta_0) - \log|\Theta_0|$, we have the following accuracy bound with probability tending to 1.*

$$\left\|\hat{\Theta}_0 - \Theta_0\right\|_F \leq K^{-1/2}\left(C_1\sqrt{\frac{b_0\log p}{n}} + C_2\sqrt{\frac{p\log p}{n}}\right).$$

Proof. Let $\Delta = \hat{\Theta}_0 - \Theta_0$. Define the function

$$G(\Delta) = \frac{1}{K}\sum_{k=1}^{K} \operatorname{tr}(S^k(\Theta_0+\Delta)) - \operatorname{tr}(S^k\Theta_0) - \log|\Theta_0+\Delta| + \log|\Theta_0| = I_1 + I_2.$$

By Taylor Expansion, we have

$$I_1 = \operatorname{tr}\Big(\Big(\frac{1}{K}\sum_{k=1}^{K} S^k - \Sigma\Big)\Delta\Big), \quad I_2 = (\tilde{\Delta})^T \int_0^1 (1-v)(\Theta_0+v\Delta)^{-1} \otimes (\Theta_0+v\Delta)^{-1} dv\tilde{\Delta},$$

where $\tilde{\Delta} = \text{vec}(\Delta)$, and Σ is the true covariance matrix.

Let $\bar{S} = \frac{1}{K}\sum_{k=1}^K S^k$. Let $X_1^k, ..., X_n^k \sim_{i.i.d.} \mathcal{N}_p(0, \Sigma)$ denote the sample for k-th category. Consider the entry of $\bar{S}$.

$$\begin{aligned}
\bar{S}_{jk} &= \frac{1}{K}\sum_{m=1}^K \frac{1}{n}\sum_{i=1}^n (X_{ij}^m - X_{.j}^m)(X_{ik}^m - X_{.k}^m) \\
&= \frac{1}{nK}\sum_{i=1}^n \sum_{m=1}^K \left(X_{ij}^m X_{ik}^m - X_{.j}^m X_{.k}^m\right),
\end{aligned}$$

where $X_{.j}^m = \frac{1}{n}\sum_i X_{ij}^m$. By Lemma (1), we have

$$\left|\frac{1}{nK}\sum_{i=1}^n \sum_{m=1}^K X_{ij}^m X_{ik}^m - \Sigma_{jk}\right| \leq C\sqrt{\frac{\log p}{nK}},$$

by letting $n = nK$ and $\nu = \sqrt{\frac{\log p}{nK}}$, with probability tending to 1 as $p \to \infty$. Also, by SLLN, $X_{.j}^m \to_{a.s.} 0$ as $n \to \infty$ for $j = 1, ..., p, m = 1, ..., K$. Then, we have

$$\max_{jk} |\bar{S}_{jk} - \Sigma_{jk}| \leq C\sqrt{\frac{\log p}{nK}},$$

with probability tending to 1 for some constant C.

Back to $|I_1|$. We obtain the upper bound

$$\begin{aligned}
|I_1| &\leq |\sum_{i\neq j}(\bar{S}_{ij} - \Sigma_{ij})\Delta_{ij}| + |\sum_{i=1}^p(\bar{S}_{ii} - \Sigma_{ii})\Delta_{ii}| \\
&\leq C\sqrt{\frac{\log p}{nK}}|\Delta^-|_1 + \left[\sum_{i=1}^p(\bar{S}_{ii} - \Sigma_{ii})^2\right]^{1/2} \left\|\Delta^+\right\|_F \\
&\leq C\sqrt{\frac{\log p}{nK}}|\Delta^-|_1 + C_2\sqrt{\frac{p\log p}{nK}} \left\|\Delta^+\right\|_F,
\end{aligned}$$

where C_2 is a positive constants. Further, let $T = \{(i,j)|\Theta_{0,ij} \neq 0\}$, and we have $|\Delta^-|_1 = |\Delta_T^-|_1 + |\Delta_{T^c}^-|_1$. Note that $\left\|\Delta_T^-\right\|_0 \leq b_0$ and $\left\|\Delta_{T^c}^-\right\|_0 \leq b_0$. Thus, we have $|\Delta_T^-|_1 \leq \sqrt{b_0}\left\|\Delta\right\|_F$ and $|\Delta_{T^c}^-|_1 \leq \sqrt{b_0}\left\|\Delta\right\|_F$. Therefore, we obtain the upper bound

$$|I_1| \leq C_1\sqrt{\frac{b_0\log p}{nK}}\left\|\Delta\right\|_F + C_2\sqrt{\frac{p\log p}{nK}}\left\|\Delta\right\|_F. \tag{1}$$

By Rothman et.al, we also have

$$I_2 \geq \frac{1}{4\tau_2^2}\left\|\Delta\right\|_F^2. \tag{2}$$

Since the estimate satisfies $\sum_{k=1}^K \text{tr}(S^k\hat{\Theta}_0) - \log|\hat{\Theta}_0| \leq \sum_{k=1}^K \text{tr}(S^k\Theta_0) - \log|\Theta_0|$, we have $G(\Delta) \leq 0$. Then, we need $I_2 \leq |I_1|$. Combining the upper bound (1) and lower bound (2), we obtain the accuracy rate

$$\frac{1}{4\tau_2^2}\left\|\Delta\right\|_F^2 \leq C_1\sqrt{\frac{b_0\log p}{nK}}\left\|\Delta\right\|_F + C_2\sqrt{\frac{p\log p}{nK}}\left\|\Delta\right\|_F,$$

which implies that with probability tending to 1 we have

$$\|\Delta\|_F = \left\|\hat{\Theta}_0 - \Theta\right\|_F \le K^{-1/2}\left(C_1\sqrt{\frac{b_0\log p}{n}} + C_2\sqrt{\frac{p\log p}{n}}\right).$$

$\square$

2 New comparison results

Let $Q\left(\{\Omega^k\}\right) = \sum_{k=1}^K \mathrm{tr}(S^k\Omega^k) - \log|\Omega^k|$, $q^k = |\Omega^k|_0$, and $q = \max_k q^k$.

	Penalized	
	L_0	L_1
Problem	$\min_{\{\Omega^k\}} Q(\{\Omega^k\}) + \lambda\sum_{k=1}^K |\Omega^k|_0.$	$\min_{\{\Omega^k\}} Q(\{\Omega^k\}) + \lambda\sum_{k=1}^K |\Omega^k|_1.$
Accuracy	**A1:** Suppose $\lambda \ge \Lambda_1\left(\frac{\log p}{n}\right)^{1/2}\sum_{k=1}^K\left\|\Delta^k\right\|_F$ for Λ_1 large enough, we have ground truth $\sum_{k=1}^K\|\Delta\|_F \le C'K\frac{F(p,q,K)}{n^{1/2}}$, where $F(p,q,K) = CKq\sqrt{\log p} + C_1\sqrt{q\log p} + C_2\sqrt{p\log p}$	**A2:** Suppose $\lambda \ge \Lambda_1\left(\frac{\log p}{n}\right)^{1/2}$, we have $\sum_{k=1}^K\|\Delta\|_F \le C'K\frac{F(p,q)}{n^{1/2}}$, where $F(p,q) = C\sqrt{q\log p} + C_2\sqrt{p\log p}$.

~~lambda |\Delta|_1 - |Delta|_1>0~~

lambda |\Delta|_0 - |delta|_1 >0

	Constrained	
	L_0	L_1
Problem	$\min_{\{\Omega^k\}} Q(\{\Omega^k\}) \quad s.t. |\Omega^k|_0 \le s.$	$\min_{\{\Omega^k\}} Q(\{\Omega^k\}) \quad s.t. |\Omega^k|_1 \le c^k.$
Accuracy	**A3:** We have $\sum_{k=1}^K\|\Delta\|_F \le C'K\frac{F(p,q)}{n^{1/2}}$, where $F(p,q) = C_1\sqrt{s\log p} + C_2\sqrt{p\log p}$	**A4:** Suppose $c^k = |\Omega^k|_1$. We have $\sum_{k=1}^K\|\Delta\|_F \le C'K\frac{F(p,q)}{n^{1/2}}$, where $F(p,q) = C_1\sqrt{q\log p} + C_2\sqrt{p\log p}$.

Summary

The dependence on (p, n) are at the same rate now. The penalized L_0 has a slightly worse dependence on the sparsity q. Note that the condition for λ in penalized L_0 and the condition c^k in constrained L_1 may be improved.