

Graphic Lasso: Estimation Error

Jiixin Hu

March 6, 2021

1 Estimation Error

The precision model is stated as

$$\mathbb{E}[S^k] = \Omega^k = \sum_{l=1}^r u_{kl} \Theta^l, \quad k \in [K].$$

Consider the following penalized optimization problem

$$\max_{\mathbf{U}, \Theta^l} \mathcal{L}_S(\mathbf{U}, \Theta^l) = - \sum_{k=1}^K \text{tr}(S^k \Omega^k) + \log \det(\Omega^k) + \lambda \|\Omega^k\|,$$

where \mathbf{U} is a membership matrix, and $\{\Theta^l\}$ are irreducible and invertible.

Notations.

1. $I'_l = \{k : u'_{kl} \neq 0\}$ is the index set for the l -th group based on the membership \mathbf{U}' .
2. δ be the minimal gap between Θ^l . That is

$$\min_{k, l \in [r]} \|\Theta^l - \Theta^k\|_F^2 = \delta^2.$$

3. Let $l(\mathbf{U}, \Theta^l)$ be the population-based loss function. That is

$$l(\mathbf{U}, \Theta^l) = \mathbb{E}_S[\mathcal{L}_S(\mathbf{U}, \Theta^l)] = - \sum_{k=1}^K \text{tr}(\Sigma^k \Omega^k) + \log \det(\Omega^k) - \lambda \sum_{k=1}^K \|\Omega^k\|_1.$$

4. Given the membership \mathbf{U}' , let $\hat{\Theta}^l(\mathbf{U}') = \arg \max_{\Theta^l} \mathcal{L}_S(\mathbf{U}', \Theta)$. Particularly, for each $l \in [r]$, we have

$$\hat{\Theta}^l(\mathbf{U}') = \arg \max_{\Theta} - \sum_{k \in I'_l} \langle S^k, \Theta \rangle + |I'_l| \log \det(\Theta) - \lambda |I'_l| \|\Theta\|_1,$$

5. Given the membership \mathbf{U}' , let $\tilde{\Theta}^l(\mathbf{U}') = \arg \max_{\mathbf{U}', \Theta^l} \mathcal{L}_S(\mathbf{U}', \Theta)$. Particularly, for each $l \in [r]$, we have

$$\tilde{\Theta}^l(\mathbf{U}') = \arg \max_{\Theta} - \sum_{k \in I'_l} \langle \Sigma^k, \Theta \rangle + |I'_l| \log \det(\Theta) - \lambda |I'_l| \|\Theta\|_1.$$

6. Define functions

$$F(\mathbf{U}') = \mathcal{L}_S(\mathbf{U}', \hat{\Theta}^l(\mathbf{U}')), \quad G(\mathbf{U}') = l(\mathbf{U}', \tilde{\Theta}^l(\mathbf{U}')).$$

Upper bound of lambda —> required by clustering accuracy —> low lambda is good for accuracy, whereas high lambda is bad (think about why).

Lower bound of lambda —> required by selection consistency (the second results in Guo et al)—> high lambda is good for selection accuracy, whereas low lambda is bad (think about why.)

7. τ be the maximal singular value of the true precision matrix, i.e., $\tau = \max_{l \in [r]} \varphi_{\max}(\Theta^l)$.

Lemma 1 (Estimation error). *Given a membership \mathbf{U}' , assume $\lambda \leq \mathcal{O}(n^{-1/2})$. With high probability, we have the following probability* use Taylor expansion around $(\lambda, t) = (0, 0)$ to show

$$p(t) = \mathbb{P}(|F(\mathbf{U}') - G(\mathbf{U}')| \geq t) \leq C_1 \exp[-C_2 n a(\lambda, t)^2],$$

where $a(\lambda, t) = \frac{-(2\lambda+1) + \sqrt{(2\lambda+1)^2 - 4(2\lambda^2 - t/Kp^2\tau^2)}}{2}$, C_1, C_2 are two constants, and $p(t) \rightarrow 0$ as $t \rightarrow \infty$.

Proof. With given membership \mathbf{U}' , we have estimations $\hat{\Theta}^l(\mathbf{U}')$ and $\tilde{\Theta}^l(\mathbf{U}')$, which we use $\hat{\Theta}^l$ and $\tilde{\Theta}^l$ refer to them for simplicity, respectively. By the definition, we have

$$\begin{aligned} |F(\mathbf{U}') - G(\mathbf{U}')| &= |\mathcal{L}_S(\mathbf{U}', \hat{\Theta}^l) - l(\mathbf{U}', \tilde{\Theta}^l)| \\ &\leq \sum_{l=1}^r |f^l(\hat{\Theta}^l) - g^l(\tilde{\Theta}^l)|, \end{aligned}$$

Combine with your earlier self-consistency results. What conclusion do we have?

where

$$f^l(\Theta) = - \sum_{k \in I'_l} \langle S^k, \Theta \rangle + |I'_l| \log \det(\Theta) - \lambda |I'_l| \|\Theta\|_1,$$

and

$$g^l(\Theta) = - \sum_{k \in I'_l} \langle \Sigma^k, \Theta \rangle + |I'_l| \log \det(\Theta) - \lambda |I'_l| \|\Theta\|_1.$$

Note that the functions $f^l(\cdot)$ and $g^l(\cdot)$ for $l \in [r]$ depends on the membership \mathbf{U}' , and $\hat{\Theta}^l, \tilde{\Theta}^l$ are unique maximizers for $f^l(\Theta), g^l(\Theta)$, respectively.

Next, for an arbitrary $l \in [r]$, we try to find the upper bound for $|f^l(\hat{\Theta}^l) - g^l(\tilde{\Theta}^l)|$. For simplicity, we use $f, g, \hat{\Theta}, \tilde{\Theta}$ denote $f^l, g^l, \hat{\Theta}^l$ and $\tilde{\Theta}^l$. Consider a new estimation $\check{\Theta}$ such that

$$\check{\Theta} = \arg \max_{\Theta} - \sum_{k \in I'_l} \langle \Sigma^k, \Theta \rangle + |I'_l| \log \det(\Theta).$$

By a straight calculation, we have the closed form of $\check{\Theta}$, which is equal to

$$\check{\Theta} = \left(\frac{\sum_{k \in I'_l} \Sigma^k}{|I'_l|} \right)^{-1}.$$

Then, we have

$$\begin{aligned} |f(\hat{\Theta}) - g(\tilde{\Theta})| &\leq |f(\hat{\Theta}) - f(\check{\Theta})| + |f(\check{\Theta}) - g(\check{\Theta})| + |g(\check{\Theta}) - g(\tilde{\Theta})| \\ &= M_1 + M_2 + M_3. \end{aligned}$$

1. **For M_1** , we have

$$f(\hat{\Theta}) - f(\check{\Theta}) = \sum_{k \in I'_l} \langle S^k, \check{\Theta} - \hat{\Theta} \rangle + |I'_l| \left(\log \det(\hat{\Theta}) - \log \det(\check{\Theta}) \right) - \lambda |I'_l| \left(\|\hat{\Theta}\|_1 - \|\check{\Theta}\|_1 \right).$$

Define $\Delta_1 = \hat{\Theta} - \check{\Theta}$ and consider the function $m(t) = \log \det(\check{\Theta} + t\Delta_1)$. By Taylor expansion, we have

$$\begin{aligned} \log \det(\hat{\Theta}) - \log \det(\check{\Theta}) &= m(1) - m(0) \\ &= \langle \check{\Theta}^{-1}, \Delta_1 \rangle - \text{vec}(\Delta_1)^T \int_0^1 (1-v)(\check{\Theta} + v\Delta_1)^{-1} \otimes (\check{\Theta} + v\Delta_1)^{-1} dv \text{vec}(\Delta_1) \\ &\leq \langle \check{\Theta}^{-1}, \Delta_1 \rangle - \frac{1}{4\tau^2} \|\Delta_1\|_F^2, \end{aligned}$$

where the first inequality follows by the proof of Theorem 1 in A.J. Rothman et al. (inequality (18)). Note that $f(\hat{\Theta}) - f(\check{\Theta}) \geq 0$, we have

$$\begin{aligned} |f(\hat{\Theta}) - f(\check{\Theta})| &\leq \sum_{k \in I'_l} \langle S^k - \Sigma^k, \Delta_1 \rangle - \frac{1}{4\tau^2} |I'_l| \|\Delta_1\|_F^2 + \lambda |I'_l| \|\Delta_1\|_1 \\ &\leq |I'_l| \max_{(i,j), k \in I'_l} |S_{ij}^k - \Sigma_{ij}^k| \|\Delta_1\|_1 - \frac{1}{4\tau^2} |I'_l| \|\Delta_1\|_F^2 + \lambda |I'_l| \|\Delta_1\|_1 \\ &\leq |I'_l| \left(-\frac{1}{4\tau^2} \|\Delta_1\|_F^2 + (\lambda + \max_{(i,j), k \in I'_l} |S_{ij}^k - \Sigma_{ij}^k|) p \|\Delta_1\|_F \right), \\ &\leq |I'_l| \tau^2 p^2 (\lambda + \max_{(i,j), k \in I'_l} |S_{ij}^k - \Sigma_{ij}^k|)^2 \end{aligned}$$

where the third inequality follows by the fact the $\|\Delta\|_1 \leq p \|\Delta\|_F$, and the last inequality follows by the property of quadratic function.

2. **For** M_2 , we have

$$\begin{aligned} |f(\check{\Theta}) - g(\check{\Theta})| &= \left| \sum_{k \in I'_l} \langle S^k - \Sigma^k, \check{\Theta} \rangle \right| \\ &\leq |I'_l| \left\| S^k - \Sigma^k \right\|_2 \|\check{\Theta}\|_2 \\ &\leq p^2 \tau^2 |I'_l| \max_{(i,j), k \in I'_l} |S_{ij}^k - \Sigma_{ij}^k|. \end{aligned}$$

3. **For** M_3 , we have

$$g(\check{\Theta}) - g(\tilde{\Theta}) = \sum_{k \in I'_l} \langle \Sigma^k, \tilde{\Theta} - \check{\Theta} \rangle + |I'_l| \left(\log \det(\check{\Theta}) - \log \det(\tilde{\Theta}) \right) - \lambda |I'_l| (\|\check{\Theta}\|_1 - \|\tilde{\Theta}\|_1).$$

Let $\Delta_2 = \tilde{\Theta} - \check{\Theta}$. By Taylor Expansion and similar procedures for M_1 , we have

$$\log \det(\tilde{\Theta}) - \log \det(\check{\Theta}) \leq \langle \check{\Theta}^{-1}, \Delta_2 \rangle - \frac{1}{4\tau^2} \|\Delta_2\|_F^2.$$

Then, we have

$$\begin{aligned} g(\check{\Theta}) - g(\tilde{\Theta}) &\geq \sum_{k \in I'_l} \langle \Sigma^k, \Delta_2 \rangle - |I'_l| (\langle \check{\Theta}^{-1}, \Delta_2 \rangle - \frac{1}{4\tau^2} \|\Delta_2\|_F^2) - \lambda |I'_l| \|\Delta_2\|_1 \\ &= \frac{1}{4\tau^2} |I'_l| \|\Delta_2\|_F^2 - \lambda |I'_l| \|\Delta_2\|_1. \end{aligned}$$

Since $g(\check{\Theta}) - g(\tilde{\Theta}) \leq 0$, we have

$$\begin{aligned} |g(\check{\Theta}) - g(\tilde{\Theta})| &\leq -\frac{1}{4\tau^2} |I'_l| \|\Delta_2\|_F^2 + \lambda |I'_l| \|\Delta_2\|_1 \\ &\leq -\frac{1}{4\tau^2} |I'_l| \|\Delta_2\|_F^2 + \lambda |I'_l| p \|\Delta_2\|_F \\ &\leq \tau^2 \lambda^2 p^2 |I'_l| \end{aligned}$$

Therefore, we have the upper bound

$$\begin{aligned} |f(\hat{\Theta}) - g(\tilde{\Theta})| &\leq M_1 + M_2 + M_3 \\ &\leq |I'_l| p^2 \tau^2 \left[(\lambda + \max_{(i,j), k \in I'_l} |S_{ij}^k - \Sigma_{ij}^k|)^2 + \max_{(i,j), k \in I'_l} |S_{ij}^k - \Sigma_{ij}^k| + \lambda^2 \right], \end{aligned}$$

and thus we have

$$\begin{aligned} |F(\mathbf{U}') - G(\mathbf{U}')| &\leq \sum_{l=1}^r |f^l(\hat{\Theta}^l) - g^l(\tilde{\Theta}^l)| \\ &\leq K p^2 \tau^2 \left[(\lambda + \max_{(i,j), k \in K} |S_{ij}^k - \Sigma_{ij}^k|)^2 + \max_{(i,j), k \in K} |S_{ij}^k - \Sigma_{ij}^k| + \lambda^2 \right]. \end{aligned}$$

Intuitively, if λ tends to 0, the error only related to the gap between population and sample $\max_{(i,j), k \in K} |S_{ij}^k - \Sigma_{ij}^k|$.

Last, we obtain the probability

$$\begin{aligned} \mathbb{P}(|F(\mathbf{U}') - G(\mathbf{U}')| \geq t) &\leq \mathbb{P}\left((\lambda + \max_{(i,j), k \in K} |S_{ij}^k - \Sigma_{ij}^k|)^2 + \max_{(i,j), k \in K} |S_{ij}^k - \Sigma_{ij}^k| + \lambda^2 \geq \frac{t}{K p^2 \tau^2}\right) \\ &= \mathbb{P}\left(\max_{(i,j), k \in K} |S_{ij}^k - \Sigma_{ij}^k|^2 + (2\lambda + 1) \max_{(i,j), k \in K} |S_{ij}^k - \Sigma_{ij}^k| + 2\lambda^2 - \frac{t}{K p^2 \tau^2} \geq 0\right) \\ &= \mathbb{P}\left(\max_{(i,j), k \in K} |S_{ij}^k - \Sigma_{ij}^k| \geq \frac{-(2\lambda + 1) + \sqrt{(2\lambda + 1)^2 - 4(2\lambda^2 - t/K p^2 \tau^2)}}{2}\right). \end{aligned}$$

Let $a(\lambda, t) = \frac{-(2\lambda+1) + \sqrt{(2\lambda+1)^2 - 4(2\lambda^2 - t/K p^2 \tau^2)}}{2}$. Note that $\lim_{\lambda \rightarrow 0} a(\lambda, t) = \frac{-1 + \sqrt{1 + 4t/K p^2 \tau^2}}{2}$ is an increasing function along with t . By the Lemma 2, we have

$$p(t) = \mathbb{P}(|F(\mathbf{U}') - G(\mathbf{U}')| \geq t) \geq \mathbb{P}\left(\max_{(i,j), k \in K} |S_{ij}^k - \Sigma_{ij}^k| \geq a(\lambda, t)\right) \leq C_1 \exp\{-C_2 n a(\lambda, t)^2\}.$$

To ensure $p(t)$ decreases as $n \rightarrow \infty$, we need $\lambda \leq \mathcal{O}(n^{-1/2})$ since $a(\lambda, t) = \mathcal{O}(\lambda)$. □

Remark 1. In non-penalized case, it is easy to measure the distance between $\hat{\Theta}$ and $\tilde{\Theta}$, since both of them have closed form and can be represented by the gap between sample and population $|S - \Sigma|$ and the properties of the true Θ . But in our case, both $\hat{\Theta}$ and $\tilde{\Theta}$ do not have closed form and thus it is hard to describe $\|\hat{\Theta} - \tilde{\Theta}\|$. Therefore, I introduce a new estimation $\check{\Theta}$ which is the estimate without the penalty and is a combination of true precision matrices. Then, taking the advantage of the optimization properties of $\hat{\Theta}, \tilde{\Theta}$, we find the bound.

Lemma 2. *Let $Z_i \sim_{i.i.d.} \mathcal{N}(0, \Sigma)$ and $\varphi_{\max}(\Sigma) \leq \tau < \infty$. Let $\Sigma = \llbracket \Sigma_{ij} \rrbracket$, then*

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

where c_1, c_2, δ depends on τ only.