Graphic Lasso: Scaled membership (Simple Case)

Jiaxin Hu

April 1, 2021

1 Thoughts

1. In Note 0323, I decomposed the original difference of the likelihood into 5 terms $H_1, ..., H_5$, and I tried to use the following inequality to show the MLE estimate is near to the true parameters.
$$0 \geq G(\hat{u}, \hat{\Theta}) \geq H_1 + H_5 - H_2 - |H_3| + H_4.$$
However, from G to $H_1, ..., H_5$, there are a lot of inequalities. I think this may be the reason why I can not show $\hat{u} \to u$ and $\hat{\Theta} \to \Theta$.

   Therefore, in the following new proof, I would like to use the original G and show that $\hat{u}\hat{\Theta} \to u\Theta$ and further $\hat{u} \to u, \hat{\Theta} \to \Theta$.

   In the discrete case, we have $\sum_{al} D_{al} \|\Delta_{al}\|_F \to 0$, where D_{al} is the entries of confusion matrix and $\Delta_{al} = \hat{\Theta}^l - \Theta^a$. Then, we know that
$$D_{al} \|\Delta_{al}\| + D_{a'l} \|\Delta_{a'l}\| \geq \min\{D_{al}, D_{a'l}\} \left\| \Theta^a - \Theta^{a'} \right\| \geq \min\{D_{al}, D_{a'l}\}\delta,$$
where δ is the minimal gap between Θ^l. Thus, for each a, there is only one l such that D_{al} does not tend to 0, i.e., with proper permutation, all the off-diagonal elements in the confusion matrix tends to 0.

   In our case, $\sum_{k=1}^K \left\| \hat{u}_k \hat{\Theta} - u_k \Theta \right\|$ is an analogy of $\sum_{al} D_{al} \|\Delta_{al}\|_F$ in the continuous case. Since we do not have minimal gap here and $\hat{\Theta}, \Theta$ are positive definite, I think similar techniques can be applied to our case from the angle of u_k. See Step I for details.

2. The constraint $\|u\|_F^2 = K$ is crucial since we need $u_k \geq a > 0$ and the norm of u grows along with K.

2 Simple case

Consider the model in which K categories share the same precision matrix structure with different magnitude. The optimization problem is stated below:

$$\begin{aligned}
\min_{\{u,\Theta\}} \quad & \mathcal{L}(u,\Theta) = \sum_{k=1}^{K} \langle S^k, \Omega^k \rangle - \log\det(\Omega^k), \\
s.t. \quad & \Omega^k = u_k\Theta, \quad k = 1, ..., K, \\
& u_k \geq a, \|u\|_F^2 = K, \quad a > 0, \\
& \Theta \text{ is positive definite with, and } \tau_1 < \varphi_{\min}(\Theta) \leq \varphi_{\max}(\Theta) < \tau_2, \tau_1, \tau_2 > 0
\end{aligned}$$

Lemma 1 (Precision matrix Accuracy)**.** *Let $\{u,\Theta\}$ denote the true parameters. Consider a estimation $\{\hat{u},\hat{\Theta}\}$ such that $\mathcal{L}(\hat{u},\hat{\Theta}) \geq \mathcal{L}(u,\Theta)$. With probability tends to 1 as $n \to \infty$, we have the accuracy*

$$\sum_{k=1}^{K} \left\|\hat{\Omega}^k - \Omega^k\right\|_F = \sum_{k=1}^{K} \left\|\hat{u}_k\hat{\Theta} - u_k\Theta\right\|_F \leq 16\tau_2^2 K^{3/2} C\sqrt{\frac{p^2\log p}{n}}$$

Remark 1. In the accuracy rate, the order of K is $\mathcal{O}(K^{3/2})$. We can consider the factor $\mathcal{O}(\sqrt{K})$ is from the estimation of Θ and K is from the estimation of u_K. I think this conclusion follows the intuition. By previous result Note 0113, the accuracy rate to estimation common precision matrix is of order $\mathcal{O}(\sqrt{K})$. For continuous values u_k, the accuracy rate to estimate K different continuous variables is about $\mathcal{O}(K)$. In our result, $\sum_{k=1}^{K} \left\|\hat{u}_k\hat{\Theta} - u_k\Theta\right\|_F$ can be considered as a multiplication $(\hat{u}\hat{\Theta})$ of the error from estimating Θ and u_k.

Proof. We prove the accuracy rate by two steps.

Step I: Show that $\hat{u} \to u$ and $\hat{\Theta} \to \Theta$.

First, we define

$$\begin{aligned}
G(\hat{u},\hat{\Theta}) &= \mathcal{L}(\hat{u},\hat{\Theta}) - \mathcal{L}(u,\Theta) \\
&= \sum_{k=1}^{K} \langle S^k, \hat{u}_k\hat{\Theta} \rangle - \langle S^k, u_k\Theta \rangle - \log\det(\hat{u}_k\hat{\Theta}) + \log\det(u_k\Theta).
\end{aligned}$$

Let $\Delta_k = \hat{u}_k\Theta - u_k\Theta$. By Taylor expansion, we have

$$\begin{aligned}
-\log\det(\hat{u}_k\hat{\Theta}) + \log\det(u_k\Theta) &\geq -\langle (u_k\Theta)^{-1}, \Delta_k \rangle + \frac{1}{2u_k^2\tau_2^2 + \|\Delta_k\|_F^2} \|\Delta_k\|_F^2, \\
&\geq -\langle u_k^{-1}\Sigma^{-1}, \Delta_k \rangle + \frac{1}{2u_k^2\tau_2^2 + \|\Delta_k\|_F^2} \|\Delta_k\|_F^2. \qquad (1)
\end{aligned}$$

Plugging the inequality (1) into G, we have

$$G(\hat{u},\hat{\Theta}) \geq \sum_{k=1}^{K} \langle S^k - u_k^{-1}\Sigma, \Delta_k \rangle + \frac{1}{2K\tau_2^2 + (\sum_{k=1}^{K} \|\Delta_k\|_F)^2} \sum_{k=1}^{K} \|\Delta_k\|_F^2. \qquad (2)$$

Let $X_1^k, ..., X_n^k \sim_{i.i.d.} \mathcal{N}(0, \Sigma/u_k)$. We know that

$$S_{jl}^k = \frac{1}{n}\sum_{i=1}^{n} \left[X_{ij}^k X_{jl}^k - X_{.j}^k X_{.l}^k\right].$$

Since $X^k_{.j}, X^k_{.l} \to 0$ almost sure when $n \to \infty$, we have

$$|S^k_{jl} - \Sigma_{jl}/u_k| = |\frac{1}{n} X^k_{ij} X^k_{jl} - \Sigma_{jl}/u_k| \le C\sqrt{\frac{\log p}{n}}, \tag{3}$$

with high probability. Therefore, by the assumption $\mathcal{L}(\hat{u}, \hat{\Theta}) \ge \mathcal{L}(u, \Theta)$, we have

$$0 \ge G(\hat{u}, \hat{\Theta}) \ge \frac{1}{2K\tau_2^2 + (\sum_{k=1}^K \|\Delta_k\|_F)^2} \sum_{k=1}^K \|\Delta_k\|_F^2 - C\sqrt{\frac{\log p}{n}} \sum_{k=1}^K \|\Delta_k\|, \tag{4}$$

which implies that

$$C\sqrt{\frac{\log p}{n}} K \left[2K\tau_2^2 + (\sum_{k=1}^K \|\Delta_k\|_F)^2\right] - \sum_{k=1}^K \|\Delta_k\|_F \ge 0.$$

Note that $\sqrt{\frac{\log p}{n}} \to 0$ as $n \to \infty$. We need

$$\sum_{k=1}^K \|\Delta_k\|_F = \sum_{k=1}^K \left\|\hat{u}_k \hat{\Theta} - u_k \Theta\right\|_F \to 0, \quad n \to \infty.$$

Since $\|\Delta_k\|_F \ge 0$, we also have

$$\|\Delta_k\|_F = \left\|\hat{u}_k \hat{\Theta} - u_k \Theta\right\|_F \to 0, \quad n \to \infty, \quad \text{for all} \quad k \in [K]$$

and thus

$$\left\|\hat{u}_k \hat{\Theta} - u_k \Theta\right\|_F / u_k \to 0, \quad \text{for all} \quad k \in [K], \quad \text{and} \quad \sum_{k=1}^K \left\|\hat{u}_k \hat{\Theta} - u_k \Theta\right\|_F / u_k \to 0.$$

For arbitrary $k, k' \in [K]$, note that

$$\left\|\hat{u}_k \hat{\Theta} - u_k \Theta\right\|_F / u_k + \left\|\hat{u}_{k'} \hat{\Theta} - u_{k'} \Theta\right\|_F / u_{k'} \ge \left\|\left(\hat{u}_k/u_k - \hat{u}_{k'}/u_{k'}\right) \hat{\Theta}\right\|_F \to 0,$$

which implies for any pair (k, k'), we need

$$\frac{\hat{u}_k}{u_k} - \frac{\hat{u}_{k'}}{u_{k'}} \to 0, \quad \text{and thus} \quad \hat{u} \to cu,$$

for some constant c. By the assumption that $\|\hat{u}\|_F = \|u\|_F = K$, the constant $c = 1$ and therefore we obtain that $\hat{u} \to u$ as $n \to \infty$. On the other hand, given $\hat{u} \to u$, we also have

$$\|\Delta_k\|_F = \left\|u_k(\hat{\Theta} - \Theta) + (\hat{u}_k - u_k)\hat{\Theta}\right\|_F \to 0, \quad \text{for all} \quad k \in [K],$$

which implies that $\left\|\hat{\Theta} - \Theta\right\|_F \to 0$.

Sanity Check: Let $S^k = u_k^{-1}\Sigma$.

The inequality (2) becomes,

$$0 \geq G(\hat{u}, \hat{\Theta}) \geq \frac{1}{2K\tau_2^2 + (\sum_{k=1}^K \|\Delta_k\|_F)^2} \sum_{k=1}^K \|\Delta_k\|_F^2 \,,$$

which requires $\sum_{k=1}^K \|\Delta_k\|_F^2 \to 0$, otherwise, the right hand side tends to a positive constant as $n \to \infty$. Then, following the above steps from $\sum_{k=1}^K \|\Delta_k\|_F^2 \to 0$ to $\hat{u}_k \to u_k$ and $\hat{\Theta} \to \Theta$, we obtain the conclusion that MLE is near the true parameters.

Step II: Sharpen the accuracy rate.

Note that accuracy rate bound from inequality (4) is sub-optimal since it does not use the common structure of the precision matrix. Therefore, back to the inequality (2) of G.

$$\begin{aligned}
G(\hat{u}, \hat{\Theta}) &\geq \sum_{k=1}^K \langle S^k - u_k^{-1}\Sigma, \Delta_k \rangle + \sum_{k=1}^K \frac{1}{2u_k^2\tau_2^2 + (\sum_{k=1}^K \|\Delta_k\|_F)^2} \|\Delta_k\|_F^2 \,, \\
&\geq \sum_{k=1}^K \langle \left[u_k S^k - \Sigma \right], \Delta_k / u_k \rangle + \frac{1}{4\tau_2^2} \sum_{k=1}^K \|\Delta_k / u_k\|_F^2 \,, \\
&= I_1 + I_2.
\end{aligned}$$

where the second inequality follows by the conclusion in Step I, and I_1, I_2 denote the two terms respectively. Let $\Delta = \hat{\Theta} - \Theta$. Note that

$$\Delta_k / u_k = \hat{u}_k / u_k \hat{\Theta} - \Theta = \Delta + (\hat{u}_k / u_k - 1)\,\hat{\Theta}. \tag{5}$$

For I_1, by the decomposition (5), we have

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

By similar process to obtain the inequality (3), we have

$$\max_{(i.j)} | \sum_{k=1}^K \left[u_k S_{jl}^k - \Sigma_{jl} \right] | \leq \sqrt{K} C \sqrt{\frac{\log p}{n}},$$

with high probability. Therefore, we have

$$|I_1| \leq \sqrt{K} C \sqrt{\frac{p^2 \log p}{n}} \left[\|\Delta\|_F + \max_{k \in [K]} | \, (\hat{u}_k / u_k - 1) \, | \left\| \hat{\Theta} \right\|_F \right]. \tag{6}$$

For I_2, note that for n large enough,

$$\begin{aligned}
\|\Delta_k / u_k\|_F &= \|\Delta\|_F + \max_{k \in [K]} | \, (\hat{u}_k / u_k - 1) \, | \left\| \hat{\Theta} \right\|_F \\
&\qquad + \left\| \Delta + | \, (\hat{u}_k / u_k - 1) \, | \hat{\Theta} \right\|_F - \left(\|\Delta\|_F + \max_{k \in [K]} | \, (\hat{u}_k / u_k - 1) \, | \left\| \hat{\Theta} \right\|_F \right) \\
&\geq \frac{1}{2} \left[\|\Delta\|_F + \max_{k \in [K]} | \, (\hat{u}_k / u_k - 1) \, | \left\| \hat{\Theta} \right\|_F \right],
\end{aligned}$$

where the inequality follows the fact that both $\|\Delta\|_F, \max_{k\in[K]} |(\hat{u}_k/u_k - 1)| \to 0$ as $n \to \infty$. This inequality makes sense since $\|A+B\|_F^2$ are near to $\|A\|_F^2 + \|B\|_F^2$ when all the entries in A, B are close to 0. Therefore, we have

$$\begin{aligned} I_2 &\geq \frac{1}{16\tau_2^2} \sum_{k=1}^K \left[\|\Delta\|_F + \max_{k\in[K]} |(\hat{u}_k/u_k - 1)| \left\|\hat{\Theta}\right\|_F \right]^2 \\ &= \frac{1}{16\tau_2^2} K \left[\|\Delta\|_F + \max_{k\in[K]} |(\hat{u}_k/u_k - 1)| \left\|\hat{\Theta}\right\|_F \right]^2. \end{aligned} \tag{7}$$

Combining the inequality (6), (7) with the assumption that $G(\hat{u}, \hat{\Theta}) \leq 0$, we have

$$\begin{aligned} 0 &\geq I_2 - |I_1| \\ &\geq \frac{1}{16\tau_2^2} K \left[\|\Delta\|_F + \max_{k\in[K]} |(\hat{u}_k/u_k - 1)| \left\|\hat{\Theta}\right\|_F \right]^2 \\ &\quad - \sqrt{K} C \sqrt{\frac{p^2 \log p}{n}} \left[\|\Delta\|_F + \max_{k\in[K]} |(\hat{u}_k/u_k - 1)| \left\|\hat{\Theta}\right\|_F \right], \end{aligned}$$

which implies that

$$K \left[\|\Delta\|_F + \max_{k\in[K]} |(\hat{u}_k/u_k - 1)| \left\|\hat{\Theta}\right\|_F \right] \leq 16\tau_2^2 \sqrt{K} C \sqrt{\frac{p^2 \log p}{n}}.$$

Last, note that

$$\sum_{k=1}^K \|\Delta_k/u_k\|_F \leq K \|\Delta\|_F + \sum_{k=1}^K (\hat{u}_k/u_k - 1) | \left\|\hat{\Theta}\right\|_F \leq K \left[\|\Delta\|_F + \max_{k\in[K]} |(\hat{u}_k/u_k - 1)| \left\|\hat{\Theta}\right\|_F \right],$$

and

$$\sum_{k=1}^K \|\Delta_k\|_F \leq \max_{k\in[K]} u_K \sum_{k=1}^K \|\Delta_k/u_k\|_F \leq \sqrt{K} \sum_{k=1}^K \|\Delta_k/u_k\|_F,$$

where the second inequality follows by the fact that $\max_{k\in[K]} u_k \leq \sqrt{K}$.

Finally, we have the accuracy rate

$$\sum_{k=1}^K \|\Delta_k\|_F \leq 16\tau_2^2 K C \sqrt{\frac{p^2 \log p}{n}}.$$

□

Remark 2. An intermediate conclusion is that

$$\sum_{k=1}^K \|\Delta_k/u_k\|_F \leq 16\tau_2^2 \sqrt{K} C \sqrt{\frac{p^2 \log p}{n}}.$$

Note that

$$\sum_{k=1}^K \Delta_k/u_k = \sum_{k=1}^K \Delta + \sum_{k=1}^K (\hat{u}_k/u_k - 1)\hat{\Theta},$$

where $\Delta = \hat{\Theta} - \Theta$. Since $\sum_{k=1}^{K}(\hat{u}_k/u_k - 1) \to 0$, we can see that the accuracy rate of $\sum_{k=1}^{K} \|\Delta\|_F$ is of order $\mathcal{O}(\sqrt{K})$ and thus $\|\Delta\|_F = \mathcal{O}(1/\sqrt{K})$, which is consistent with discrete case. Notice that

$$\sum_{k=1}^{K} \Delta_k = \sum_{k=1}^{K} u_k \Delta + \sum_{k=1}^{K} (\hat{u}_k - u_k)\hat{\Theta}.$$

An extreme case is that $u_1 = u_2 = ...u_{k-1} = a, u_k = \sqrt{K(1-a^2)}$. Then, the accuracy

$$\sum_{k=1}^{K} \|\Delta_k\| \approx \sum_{k=1}^{K} u_k \|\Delta\| = a(K-1)\|\Delta\| + \sqrt{K(1-a^2)}\|\Delta\| \approx \mathcal{O}(\sqrt{K}) + \mathcal{O}(1).$$