Graphic Lasso: Single layer consistency proof sketch

Jiaxin Hu

December 29, 2020

- Single layer consistency proof sketch

Now consider the single layer estimation problem. Let S be the sample covariance matrix. The objective function of the estimation is

$$Q(\Omega) = \mathrm{tr}(S\Omega) - \log|\Omega| + \lambda \sum_{j \neq j'} |\omega_{j,j'}|^{1/2}. \tag{1}$$

on the similar order as ll matrix ll_F

on the order of ~q

Let Ω_0 be the true precision matrix, and Σ_0 be the true covariance matrix, where $\Sigma = \Omega_0^{-1}$. Let $\hat{\Omega}$ be the local minimizer of (1). Let $T = \{(j,j') : j \neq j', \omega_{j,j'} \neq 0\}$, and $q = |T|$. We assume following assumptions.

1. There exist two constants τ_1, τ_2 such that $0 < \tau_1 < \phi_{\min}(\Omega_0) \leq \phi_{\max}(\Omega_0) < \tau_2 < \infty$, for all $p \geq 1, k = 1, ..., K$, where $\phi_{\min}(\cdot), \phi_{\max(\cdot)}$ denote the minimal and maximal eigenvalues, respectively.

2. There exists a constant $\tau_3 > 0$ such that $\min_{(j,j') \in T} |\omega_{0,j,j'}| \geq \tau_3$.

We want to prove the following theorem.

intuitively, optimal lambda is set such that the two terms in (1) have si order

Theorem 0.1 (Consistency)**.** *Suppose the above assumptions hold,* $\frac{(p+q)\log p}{n} = o(1)$*, and there exist two positive constants* Λ_1, Λ_2 *such that* $\Lambda_1 \left\{\frac{\log p}{n}\right\}^{1/2} \leq \lambda \leq \Lambda_2 \left\{\frac{(1+p/q)\log p}{n}\right\}^{1/2}$*. There exists a local minimizer of* (1) *such that*

Intuition: roughly degree of freedom = # diagonals + # non-zero off-diagonals

$$\left\|\hat{\Omega} - \Omega_0\right\|_F = O_p\left[\left\{\frac{(p+q)\log p}{n}\right\}^{1/2}\right].$$

n^(1/2) is to be expected based on CLT.

1 Follow the proof in Guo's paper

Key idea: Let $\Delta = \Omega - \Omega_0 = [\![\delta_{j,j'}]\!]$ and $G(\Delta) = Q(\Omega_0 + \Delta) - Q(\Omega_0) = Q(\Omega) - Q(\Omega_0)$. If $G(\Delta)$ has a local minimizer in a set $\mathcal{A}$, then $Q(\Omega)$ has a local minimizer Ω^* such that the corresponding Δ^* falls in $\mathcal{A}$. We can construct some $\mathcal{A}$ implies the consistency. In another word, it is sufficient to prove the existence of the local minimizer of $G(\Delta)$ in the neighborhood of $\Delta = 0$.

1.1 Proof

> c > 0, where c is a constant independent of n, p, q

Proof. Note that $G(\Delta)$ is a continuous function and $G(0) = 0$. Let $r_n = \left\{\frac{(p+q)\log p}{n}\right\}^{1/2}$ and $\mathcal{A} = \{\|\Delta\|_F \le M r_n\}$ for some positive constant M. Then $\mathcal{A}$ is a closed bounded convex set. To prove the existence of the local minimizer inside $\mathcal{A}$, by extreme value theorem, it is sufficient to show that $G(\Delta) > 0$ with probability tending to 1 for all Δ in the boundary $\partial\mathcal{A} = \{\|\Delta\|_F = M r_n\}$. Thus, the local minimizer Ω^* satisfies $\|\Omega^* - \Omega_0\|_F = O_p(r_n)$.

The following proves that $G(\Delta) > 0$ with probability tending to 1.

We rewrite the function G.

$$\begin{aligned} G(\Delta) &= \operatorname{tr}\{S\Omega_0 + S\Delta\} - \log|\Omega_0 + \Delta| - \operatorname{tr}\{S\Omega_0\} + \log|\Omega_0| \\ &\quad + \lambda \sum_{(j,j')\in T^c} |\delta_{j,j'}|^{1/2} + \lambda \sum_{(j,j')\in T} \left(|\omega_{0,j,j'} + \delta_{j,j'}|^{1/2} - |\omega_{0,j,j'}|^{1/2}\right). \end{aligned} \tag{2}$$

Consider the function $f(t) = \log|\Omega_0 + t\Delta|$. By Taylor expansion with integral form remainder, we have

$$f(t) - f(0) = \frac{\partial}{\partial t} f(t)|_{t=0} t + \int_0^t \frac{\partial^2}{2\partial t^2} f(t)|_{t=v}(t-v)dv,$$

where

$$\frac{\partial}{\partial t} f(t)|_{t=0} = \frac{\partial}{\partial t}|\Omega_0 + t\Delta| \frac{1}{|\Omega_0 + t\Delta|} = \operatorname{tr}\left((\Omega_0 + t\Delta)^{-1}\Delta\right) = \operatorname{tr}(\Sigma_0\Delta),$$

$$\frac{\partial^2}{\partial t^2} f(t)|_{t=v} = \frac{\partial}{\partial t}\operatorname{tr}\left((\Omega_0 + t\Delta)^{-1}\Delta\right)|_{t=v} = (\tilde{\Delta})^T(\Omega_0 + v\Delta)^{-1} \otimes (\Omega_0 + v\Delta)^{-1}\tilde{\Delta},$$

$\tilde{\Delta} \in \mathbb{R}^{p^2}$ is the vectorization of Δ, and $\otimes$ is the Kronecker product of two matrices. Plug the Taylor expansion of $f(1)$ at $t = 0$ into the equation (2). Now, we decompose G by four parts

$$G(\Delta) = I_1 + I_2 + I_3 + I_4, \tag{3}$$

where

$$\begin{aligned} I_1 &= \operatorname{tr}\left((S - \Sigma_0)\Delta\right), \\ I_2 &= (\tilde{\Delta})^T \int_0^1 (1-v)(\Omega_0 + v\Delta)^{-1} \otimes (\Omega_0 + v\Delta)^{-1} dv\tilde{\Delta}, \\ I_3 &= \lambda \sum_{(j,j')\in T^c} |\delta_{j,j'}|^{1/2}, \\ I_4 &= \lambda \sum_{(j,j')\in T} \left(|\omega_{0,j,j'} + \delta_{j,j'}|^{1/2} - |\omega_{0,j,j'}|^{1/2}\right). \end{aligned}$$

Let Δ^+ be the diagonal matrix with the same diagonal of Δ, $\Delta^- = \Delta - \Delta^+$, and Δ_T denote the matrix Δ with all elements outside the index set T replaced by 0. Note that $|\cdot|_1$ is not the L_1 for matrix but the L_1 for vector, i.e., $|\Delta|_1 = \sum_{ij}|\delta_{ij}|$.

By Guo et al. and Rothman et al., with probability tending to 1, we have the bound

$$|I_1| \le I_{1,1} + I_{1,2},$$

where

$$I_{1,1} = C_1\sqrt{\frac{\log p}{n}}|\Delta_T^-|_1 + C_2\sqrt{\frac{p\log p}{n}}\left\|\Delta^+\right\|_F,$$
$$I_{1,2} = C_1\sqrt{\frac{\log p}{n}}|\Delta_{T^c}^-|_1,$$

and C_1, C_2 are two positive constants. By applying the bound $|\Delta_T^-|_1 \le q^{1/2}\left\|\Delta_T^-\right\|_F$, we know that

$$I_{1,1} \le (C_1+C_2)\sqrt{\frac{(p+q)\log p}{n}}\left\|\Delta\right\|_F \le (C_1+C_2)\sqrt{\frac{(p+q)\log p}{n}}Mr_n = M(C_1+C_2)\frac{(p+q)\log p}{n}.$$

Rewrite I_3, and we notice that for r_n small enough we have $I_3 \ge \lambda|\Delta_{T^c}^-|_1$. Then,

$$I_3 - I_{1,2} \ge \left(\lambda - C_1\sqrt{\frac{\log p}{n}}\right)|\Delta_{T^c}^-|_1 \ge (\Lambda_1 - C_1)\sqrt{\frac{\log p}{n}}|\Delta_{T^c}^-|_1 \ge 0,$$

where the second inequality follows the assumption that $\lambda \ge \Lambda_1\sqrt{\frac{\log p}{n}}$ and Λ_1 is large enough.

Next, by Guo et al. and Rothman et al., I_2 can be lower bounded as following.

$$I_2 \ge \left\|\Delta\right\|_F^2 \phi_{\min}\left(\int_0^1 (1-v)(\Omega_0+v\Delta)^{-1}\otimes(\Omega_0+v\Delta)^{-1}dv\right) \ge \frac{1}{4\tau_2^2}\left\|\Delta\right\|_F^2.$$

The specific steps for the second inequality are in Rothman et al..

Now, we consider the term I_4. By triangle inequality we have

$$\begin{aligned}
|I_4| &\le \lambda\sum_{(j,j')\in T}\left||\omega_{0,j,j'}+\delta_{j,j'}|^{1/2} - |\omega_{0,j,j'}|^{1/2}\right| \\
&= \lambda\sum_{(j,j')\in T}\left|\frac{|\omega_{0,j,j'}+\delta_{j,j'}| - |\omega_{0,j,j'}|}{|\omega_{0,j,j'}+\delta_{j,j'}|^{1/2}+|\omega_{0,j,j'}|^{1/2}}\right| \\
&\le \frac{\lambda}{\tau_3^{1/2}}|\Delta_T|_1 \\
&\le \frac{M\Lambda_2}{\tau_3^{1/2}}\frac{(p+q)\log p}{n},
\end{aligned}$$

where the last inequality follows the bound $|\Delta_T^-|_1 \le q^{1/2}\left\|\Delta_T^-\right\|_F$ and the assumption that $\lambda \le \Lambda_2\sqrt{\frac{(1+p/q)\log p}{n}}$.

Back to (3), for $\Delta \in \partial\mathcal{A}$, we have

$$\begin{aligned}
G(\Delta) &\ge -I_{1,1} - I_{1,2} + I_2 + I_3 - |I_4| \\
&\ge I_2 - I_{1,1} - I_4 \\
&\ge \frac{M^2(p+q)\log p}{n}\left(\frac{1}{4\tau_2^2} - \frac{C_1+C_2-\Lambda_2/\tau_3^{1/2}}{M}\right).
\end{aligned}$$

For M large enough, we have $G(\Delta)$ for all $\Delta \in \partial\mathcal{A}$. Then, the proof completes.

$\square$

1.2 Extension to multiple layers

It is quite easy to extend the proof for single layer to multiple layers following Guo's proof. The objective function (1) changes to

$$Q(\{\Omega^k\}_{k=1}^K) = \sum_{k=1}^K \left\{ \mathrm{tr}(S^k\Omega^k) - \log|\Omega^k| \right\} + \lambda \sum_{(j,j')} \left(\sum_{k=1}^K |\omega_{j,j'}^k| \right)^{1/2}.$$

First, replace $\mathcal{A}$ by $\mathcal{A} = \left\{ \sum_{k=1}^K \left\| \Delta^k \right\|_F \leq Mr_n \right\}$. We still decompose the function $G(\{\Delta^k\}_{k=1}^K)$ by four parts, i.e.,

$$G(\{\Delta^k\}_{k=1}^K) = I_1 + I_2 + I_3 + I_4,$$

where

$$\begin{aligned}
I_1 &= \sum_{k=1}^K \mathrm{tr}\left((S - \Sigma_0)\Delta^k \right), \\
I_2 &= \sum_{k=1}^K (\tilde{\Delta}^k)^T \int_0^1 (1-v)(\Omega_0^k + v\Delta^k)^{-1} \otimes (\Omega_0^k + v\Delta^k)^{-1} dv \tilde{\Delta}^k, \\
I_3 &= \lambda \sum_{(j,j') \in T^c} \left(\sum_{k=1}^K |\delta_{j,j'}^k| \right)^{1/2}, \\
I_4 &= \lambda \sum_{(j,j') \in T} \left(\left(\sum_{k=1}^K |\omega_{0,j,j'}^k + \delta_{j,j'}^k| \right)^{1/2} - \left(\sum_{k=1}^K |\omega_{0,j,j'}^k| \right)^{1/2} \right).
\end{aligned}$$

Define $I_{1,1}, I_{1,2}$ similarly as single layer case. By simple modification, we have

$$I_{1,1} \leq \sum_{k=1}^K (C_1 + C_2) \sqrt{\frac{(p+q)\log p}{n}} \left\| \Delta^k \right\|_F \leq M(C_1 + C_2) \frac{(p+q)\log p}{n}.$$

Similarly, we have

$$I_2 \geq \sum_{k=1}^K \frac{1}{4\tau_2^2} \left\| \Delta^k \right\|_F^2 \geq \frac{M^2}{4\tau_2^2} \frac{(p+q)\log p}{n},$$

for all $\Delta \in \partial\mathcal{A}$. Note that for r_n small enough, we have $I_3 \geq \lambda \sum_{k=1}^K |\Delta_{T^c}^{k,-}|_1$. Then,

$$I_3 - I_{1,2} \geq \sum_{k=1}^K (\Lambda_1 - C_1) \sqrt{\frac{\log p}{n}} |\Delta_{T^c}^{k,-}|_1,$$

for Λ_1 large enough. Last, for I_4,

$$
\begin{aligned}
|I_4| &\leq \lambda \sum_{(j,j')\in T} \left| \frac{\left(\sum_{k=1}^K |\omega_{0,j,j'}^k + \delta_{j,j'}^k|\right) - \left(\sum_{k=1}^K |\omega_{0,j,j'}^k|\right)}{\left(\sum_{k=1}^K |\omega_{0,j,j'}^k + \delta_{j,j'}^k|\right)^{1/2} + \left(\sum_{k=1}^K |\omega_{0,j,j'}^k|\right)^{1/2}} \right| \\
&\leq \frac{\lambda}{\tau_3^{1/2}} \sum_{(j,j')\in T} \sum_{k=1}^K |\delta_{j,j'}^k| \\
&\leq \frac{\lambda}{\tau_3^{1/2}} \sum_{k=1}^K |\Delta_T^k|_1 \\
&\leq \frac{M\Lambda_2}{\tau_3^{1/2}} \frac{(p+q)\log p}{n}.
\end{aligned}
$$

Therefore, we still have

$$
G(\Delta) \geq I_2 - I_{1,2} - |I_4| \geq \frac{M^2(p+q)\log p}{n} \left(\frac{1}{4\tau_2^2} - \frac{C_1 + C_2 - \Lambda_2/\tau_3^{1/2}}{M} \right) > 0,
$$

for all $\Delta \in \partial\mathcal{A}$ and M large enough.