Graphic Lasso: Possible Accuracy

Jiaxin Hu

January 5, 2021

1 Optimization with $1/2$ norm

Let $Q(\Omega) = \mathrm{tr}(S\Omega) - \log|\Omega|$. Consider the primal minimization problem

$$\min_{\Omega = [\![\omega_{j,j'}]\!]} Q(\Omega),$$
$$s.t. \quad \sum_{j \neq j'} |\omega_{j,j'}|^{1/2} \leq C.$$

For simplicity, let $|\Omega|^{1/2} = \sum_{j\neq j'} |\omega_{j,j'}|^{1/2}$, T denote the set of indices of non-zero off-diagonal elements, 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 $\min_{(j,j')\in T} |\omega_{0,j,j'}| \geq \tau_3$.

Theorem 1.1 (Consistency (Preliminary))**.** *Suppose two assumptions hold and C is a positive constant. Let Ω denote the true precision matrix. There exists a local minimizer $\hat{\Omega}$ such that $Q(\hat{\Omega}) \leq Q(\Omega)$ and $|\hat{\Omega}|^{1/2} \leq C$, and the following accuracy bound holds with probability tending to 1.*

$$\left\|\hat{\Omega} - \Omega\right\|_F = O_p\left[\left\{\frac{(p+q)\log p}{n}\right\}^{1/4}\right].$$

Proof follows tensor paper

Proof. Consider the following decomposition

$$G(\Delta) = \mathrm{tr}(S(\Omega + \Delta)) - \mathrm{tr}(\Omega) - \log|\Omega + \Delta| + \log|\Omega| = I_1 + I_2,$$

where

$$I_1 = \mathrm{tr}((S - \Sigma)\Delta), \quad I_2 = (\tilde{\Delta})^T \int_0^1 (1 - v)(\Omega + v\Delta)^{-1} \otimes (\Omega + v\Delta)^{-1} dv \tilde{\Delta}.$$

Suppose $\hat{\Omega} = \Omega + \Delta$ has larger or equal likelihood value than the true precision matrix Ω. Then, we have $G(\Delta) \leq 0$, i.e.,

$$I_2 \leq -I_1 \leq |I_1|. \tag{1}$$

Note that

$$|I_1| \le C_1 \left(\frac{\log p}{n}\right)^{1/2} \left(|\Delta_T^-|_1 + |\Delta_{T^c}^-|_1\right) + C_2 \left(\frac{p \log p}{n}\right)^{1/2} \left\|\Delta^+\right\|_F, \quad I_2 \ge \frac{1}{4\tau_2^2} \left\|\Delta\right\|_F^2,$$

$|\Delta_T^-|_1 \le q^{1/2} \left\|\Delta\right\|_F$, and $|\Delta_{T^c}^-|_1 \le C$. To satisfy the inequality (1), we have

$$\frac{1}{4\tau_2^2} \left\|\Delta\right\|_F^2 \le (C_1 + C_2) \left(\frac{(p+q)\log p}{n}\right)^{1/2} \left\|\Delta\right\|_F + C_1 \left(\frac{(p+q)\log p}{n}\right)^{1/2} C. \tag{2}$$

Consider the equation

$$0 = -\frac{1}{4\tau_2^2} x^2 + (C_1 + C_2) \left(\frac{(p+q)\log p}{n}\right)^{1/2} x + C_1 \left(\frac{(p+q)\log p}{n}\right)^{1/2} C. \tag{3}$$

The solutions to the equation (3) are

$$\begin{aligned} x^* &= 2\tau_2^2 \left\{ (C_1 + C_2) \left(\frac{(p+q)\log p}{n}\right)^{1/2} \pm \sqrt{(C_1 + C_2)^2 \left(\frac{(p+q)\log p}{n}\right) + C_1 C \left(\frac{(p+q)\log p}{n}\right)^{1/2} / \tau_2^2} \right\} \\ &= \mathcal{O}\left[\left(\frac{(p+q)\log p}{n}\right)^{1/4}\right], \end{aligned} \tag{4}$$

where the second equality follows by the fact that the term $\sqrt{C_1 C \left(\frac{(p+q)\log p}{n}\right)^{1/2} / \tau_2^2}$ dominates the solution. Therefore, to satisfy the inequality (2), we have

$$\left\|\hat{\Omega} - \Omega\right\|_F = \left\|\Delta\right\|_F = \mathcal{O}\left[\left(\frac{(p+q)\log p}{n}\right)^{1/4}\right].$$

□

Proof follows Guo's paper

Proof. Let $\mathcal{A} = \left\{ \left\|\Delta\right\|_F \le M \left(\frac{(p+q)\log p}{n}\right)^{1/4}, |\Omega + \Delta|^{1/2} \le C \right\}$. Define $G(\Delta)$, I_1, and I_2 same as above proof. We know that

$$\begin{aligned} G(\Delta) &\ge I_2 - |I_1| \\ &\ge \frac{1}{4\tau_2^2} \left\|\Delta\right\|_F^2 - \left(\frac{(p+q)\log p}{n}\right)^{1/2} \left\|\Delta\right\|_F - C_1 \left(\frac{(p+q)\log p}{n}\right)^{1/2} C. \end{aligned}$$

By the solution (4), we have $G(\Delta) > 0$ for all $\Delta \in \partial\mathcal{A}$ with M large enough. Therefore, there exists a local minimizer inside $\mathcal{A}$ and thus $\left\|\hat{\Omega} - \Omega\right\|_F = \mathcal{O}\left[\left(\frac{(p+q)\log p}{n}\right)^{1/4}\right]$. □

2 Different constrains

1. We change the constrain as
$$\frac{|\Omega+\Delta|_1}{\|\Omega+\Delta\|_F} < C.$$
Then we have
$$|\Delta^-_{T^c}|_1 < C\,\|\Omega+\Delta\|_F\,.$$
However, the relationship between $\|\Omega+\Delta\|_F$ and $\|\Delta\|_F$ is uncertain. We only have
$$\|\Omega+\Delta\|_F \le \|\Delta\|_F + \|\Omega\|_F = \|\Delta\|_F + C'.$$
The solutions x^* are still dominated by the $\left(\frac{(p+q)\log p}{n}\right)^{1/4}$ term.

2. We change the constrain as
$$|\Omega+\Delta|_0 < s.$$
Note that $|\Omega+\Delta|_0 = |\Omega_T+\Delta_T|_0 + |\Delta_{T^c}|_0 < s$. We have
$$|\Delta^-_{T^c}|_1 < |\Delta^-_{T^c}|_0\,\|\Delta\|_{\max} \le s\,\|\Delta\|_F\,. \tag{5}$$
Plugging (5) into the inequality (1), we have
$$\frac{1}{4\tau_2^2}\|\Delta\|_F^2 \le (C_1+C_2)\left(\frac{(p+q)\log p}{n}\right)^{1/2}\|\Delta\|_F + C_1\left(\frac{\log p}{n}\right)^{1/2} s\,\|\Delta\|_F\,,$$
and thus we have
$$\left\|\hat\Omega-\Omega\right\|_F = \|\Delta\|_F = \mathcal{O}\left[\left(\frac{(p+q)\log p}{n}\right)^{1/2}\right]. \tag{6}$$
Remark 1. The above accuracy (6) holds when q is fixed. Since the true Ω should satisfy the constrain, we need $|\Omega|_0 = q < s$. Therefore, when q is not a fixed number, let $s = Mq$ for some constant M. The accuracy is of order $\mathcal{O}\left\{q\left(\frac{\log p}{n}\right)^{1/2}\right\}$.

3 Optimization without constrain

Consider the problem
$$\min_{\Omega=[\![\omega_{j,j'}]\!]} Q(\Omega).$$

Theorem 3.1. *Suppose two assumptions hold. For estimation $\hat\Omega$ such that $Q(\hat\Omega) \le Q(\Omega)$, we have following accuracy bound with probability tending to 1.*
$$\left\|\hat\Omega-\Omega\right\|_F = O_p\left[p\left\{\frac{\log p}{n}\right\}^{1/2}\right].$$

Proof. Define $G(\Delta)$, I_1, and I_2 same as above proofs. Unlike above proofs, we have

$$
\begin{aligned}
|I_1| &\leq C_1 \left(\frac{\log p}{n}\right)^{1/2} |\Delta|_1 + C_2 \left(\frac{p \log p}{n}\right)^{1/2} \left\|\Delta^+\right\|_F \\
&\leq \left\{ C_1 p \left(\frac{\log p}{n}\right)^{1/2} + C_2 \left(\frac{p \log p}{n}\right)^{1/2} \right\} \|\Delta\|_F ,
\end{aligned}
$$

where the second inequality follows by the fact that $|\Delta|_1 \leq p \|\Delta\|_F$. Then, to let $G(\Delta) \neq 0$, we need $I_2 \leq |I_1|$, i.e.,

$$
\frac{1}{4\tau_2^2} \|\Delta\|_F^2 \leq C' p \left(\frac{\log p}{n}\right)^{1/2} \|\Delta\|_F ,
$$

which implies that

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

□

Remark 2. This result makes sense. The degree of freedom for the dense model is p^2 while for sparse model is $p + q$. Compared with the accuracy with constrain, the optimization without constrain corresponds to the dense model, and the $p + q$ part in the accuracy is replaced by p^2.

4 Summary

In general, we have $|\Delta^-_{T^c}|_1 \leq \sqrt{p^2 - q} \, \|\Delta\|_F$. Thus we still have a $n^{-1/2}$ accuracy rate, $\mathcal{O}\left\{ F(p,q) \left(\frac{\log p}{n}\right)^{1/2} \right\}$, under any constrain at a cost of an additional factor of $F(p,q)$. Under the no constrain case and the L_1 norm penalty, the factor $F(p,q) = p$ while under the L_0 norm the factor $F(p,q) = q$. Therefore, L_0 norm has the best accuracy, in case of growing (p,n) and fixed q. This result intuitively makes sense because L_0 norm controls the number of non-zero entries directly while L_1 norm only controls the sum of absolute value of the entries. The L_1 norm has weaker control on sparsity compared with L_0 norm.