41500: High-Dimensional Statistics	**Winter 2020**

Precision Matrix Estimation

Lecturer: Chao Gao *Scribe: Xinze Li*

1 Precision Matrix Estimator

1.1 Problem Formulation

First suppose that $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma)$ and that $\Omega = \Sigma^{-1}$ is the precision matrix. We also assume the parameter space $\mathcal{G}(p, s, m)$ as

$$\mathcal{G}(p,s,m) = \left\{ \Omega : M^{-1} \leq \lambda_{\min}(\Omega) \leq \lambda_{\max}(\Omega) \leq M, \max_i \sum_{j=1}^{p} \mathbb{1}(\Omega_{ij} \neq 0) \leq s \right\} \tag{1}$$

Suppose $x_1 \in \mathbb{R}^{p_1}$, $x_2 \in \mathbb{R}^{p_2}$, and $p_1 + p_2 = p$

$$X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right) = \mathcal{N}\left(0, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}^{-1}\right) \tag{2}$$

And so we have the conditional probability distribution as follows

$$x_1 | x_2 \sim \mathcal{N}\left(-\Sigma_{12}\Sigma_{22}^{-1}x_2, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right) \tag{3}$$

Now let $S = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ the Schur complement. From matrix theory, we know

$$\begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1} = \begin{pmatrix} S^{-1} & -S^{-1}\Sigma_{11}\Sigma_{22}^{-1} \\ * & * \end{pmatrix} \tag{4}$$

Thus we also have

$$x_1 | x_2 \sim \mathcal{N}\left(-\Omega_{11}^{-1}\Omega_{12}x_2, \Omega_{11}^{-1}\right) \tag{5}$$

We denote $x_{-i} = x_{[p]\setminus i} \in \mathbb{R}^{p-1}$, and so

$$x_1 | x_{-1} \sim \mathcal{N}\left(-\Omega_{11}^{-1}\Omega_{1,-1}x_{-1}, \Omega_{11}^{-1}\right) \tag{6}$$

1.2 Idea

Now the idea is to apply Lasso to estimate each column of the precision matrix Ω (See the article [Meinshausen and Bühlmann, 2006]). Now recall that in Lasso, we assume that $y = X\beta + \sigma z$, and $z \sim \mathcal{N}(0, I_n)$, $X \in \mathbb{R}^{n\times p}$. The goal is to recover both β and θ. Suppose $X = (X_1, \cdots, X_n)^\top$, and that $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma)$. If we also assume that $\frac{s\log p}{n}$ is small and

$$M^{-1} \le \lambda_{\min}(\Sigma) \le \lambda_{\max}(\Sigma) \le M, \quad \sigma^2 \le M, \quad \beta \in \Theta(p, s) \tag{7}$$

And that $\widehat{\beta}$ is the Lasso estimator

$$\widehat{\beta} = \arg\min_{\beta} \|y - X\beta\|^2 + \lambda \|\beta\|_1 \tag{8}$$

If we choose

$$\lambda = C\sqrt{n\log p} \tag{9}$$

We then have

$$\begin{cases} \frac{1}{n}\left\|X\left(\widehat{\beta}-\beta\right)\right\|^2 \lesssim \frac{s\log p}{n\kappa^2} \\ \left\|\widehat{\beta}-\beta\right\|^2 \lesssim \frac{s\log p}{n\kappa^4} \end{cases} \tag{10}$$

And if we have the above assumption, we could actually obtain $\kappa \gtrsim 1$. So we have

$$\begin{cases} \frac{1}{n}\left\|X\left(\widehat{\beta}-\beta\right)\right\|^2 \lesssim \frac{s\log p}{n} \\ \left\|\widehat{\beta}-\beta\right\|^2 \lesssim \frac{s\log p}{n} \end{cases} \tag{11}$$

Now using the cone condition

$$\left\|\widehat{\beta}-\beta\right\|_1 = \|\Delta_S\|_1 + \|\Delta_{S^\complement}\|_1 \le 4\|\Delta_S\|_1 \le 4\sqrt{s}\,\|\Delta\| \le 4\sqrt{s}\left\|\widehat{\beta}-\beta\right\| \tag{12}$$

And thus

$$\left\|\widehat{\beta}-\beta\right\|_1 \lesssim s\sqrt{\frac{\log p}{n}} \tag{13}$$

Now from 6, we know that if we write

$$x_1 = \begin{pmatrix} X_{11} \\ | \\ X_{n1} \end{pmatrix} \in \mathbb{R}^{n\times 1}, \quad x_{-1} = \begin{pmatrix} X_{1,-1} \\ | \\ X_{n,-1} \end{pmatrix} \in \mathbb{R}^{n\times (p-1)} \tag{14}$$

then we have

$$x_1 = -\Omega_{11}^{-1} x_{-1}^\top \Omega_{-1,1} + \Omega_{11}^{-\frac{1}{2}} z, \quad z \sim \mathcal{N}(0, I_n), \quad \Omega_{11} \in \mathbb{R}, \quad \Omega_{-1,1} \in \mathbb{R}^{(p-1)\times 1} \tag{15}$$

And so we could do the following Lasso estimator

$$\begin{cases} \widehat{\beta} = \arg\min_{\beta\in\mathbb{R}^{p-1}} \left\|x_1 - x_{-1}^\top\beta\right\|^2 + \lambda\|\beta\|_1 \\ \widehat{\sigma}^2 = \frac{1}{n}\left\|x_1 - x_{-1}^\top\beta\right\|^2 \end{cases} \tag{16}$$

Now we are ready to give the estimate of the first column of the precision matrix

$$\begin{cases} \widehat{\Omega}_{11} = \dfrac{1}{\widehat{\sigma}^2} \\ \widehat{\Omega}_{-1,1} = \dfrac{\widehat{\beta}}{\widehat{\sigma}^2} \end{cases} \tag{17}$$

1.3 Analysis

To do the analysis, we should first check that all the assumptions are satisfied. From 14, we know that $X_{1,-1}, \cdots, X_{n,-1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Omega_{-1,-1}^{-1})$, and here $\sigma^2 = \Omega_{11}^{-1}$, $\beta = -\Omega_{11}^{-1}\Omega_{-1,1}$. Now since we assume in 1, we know that

$$M^{-1} \leq \lambda_{\min}(\Omega) \leq \lambda_{\min}(\Omega_{-1,-1}) \leq \lambda_{\max}(\Omega_{-1,-1}) \leq \lambda_{\max}(\Omega) \leq M \tag{18}$$

But $\lambda_{\max}\left(\Omega_{-1,-1}^{-1}\right) = 1/\lambda_{\min}(\Omega_{-1,-1})$, and $\lambda_{\min}\left(\Omega_{-1,-1}^{-1}\right) = 1/\lambda_{\max}(\Omega_{-1,-1})$. Thus

$$M^{-1} \leq \lambda_{\min}(\Omega_{-1,-1}^{-1}) \leq \lambda_{\max}(\Omega_{-1,-1}^{-1}) \leq M \tag{19}$$

By the same logic

$$M^{-1} \leq \sigma^2 \leq M \tag{20}$$

Also since $\sum \mathbb{1}(\Omega_{1i} \neq 0) \leq s$, and Ω is symmetric, we have $\beta \in \Theta(p-1, s)$. Thus the assumptions 7 are met. Now we are ready to do our analysis. First let's analyze $|\widehat{\sigma}^2 - \sigma^2|$,

$$\begin{aligned} \left|\widehat{\sigma}^2 - \sigma^2\right| &= \left|\frac{1}{n}\left\|X\left(\beta - \widehat{\beta}\right) + \sigma z\right\| - \sigma^2\right| \\ &\leq \frac{1}{n}\left\|X(\beta - \widehat{\beta})\right\|^2 + \sigma^2\left|\frac{\|z\|^2}{n} - 1\right| + \frac{2\sigma}{n}\left|z^\top X(\beta - \widehat{\beta})\right| \\ &\lesssim \frac{s\log p}{n} + \sqrt{\frac{\log p}{n}} + \frac{2\sigma}{n}\left\|X^\top z\right\|_\infty \cdot \left\|\beta - \widehat{\beta}\right\|_1 \\ &\lesssim \frac{s\log p}{n} + \sqrt{\frac{\log p}{n}} + \frac{2\sigma}{n}\sqrt{n\log p} \cdot s\sqrt{\frac{\log p}{n}} \\ &\lesssim \frac{s\log p}{n} + \sqrt{\frac{\log p}{n}} \end{aligned} \tag{21}$$

The second inequality is Lasso's property 11, concentration of chi-square and Holder's inequality. The third inequality is union bound and Lasso's property 11. Thus, w.h.p. the following holds

$$\left|\widehat{\sigma}^2 - \sigma^2\right| \lesssim \frac{s\log p}{n} + \sqrt{\frac{\log p}{n}} \tag{22}$$

And now we could analyze the error of the precision matrix estimator 17

$$
\begin{aligned}
\left|\widehat{\Omega}_{11}-\Omega_{11}\right| &= \left|\frac{1}{\sigma^2}-\frac{1}{\widehat{\sigma}^2}\right| \\
&= \frac{|\widehat{\sigma}^2-\sigma^2|}{\sigma^2\widehat{\sigma}^2} \\
&\lesssim \left|\widehat{\sigma}^2-\sigma^2\right| \\
&\lesssim \frac{s\log p}{n}+\sqrt{\frac{\log p}{n}}
\end{aligned} \tag{23}
$$

The first inequality is because $\sigma^2 \geq M^{-1}$. Also, $\widehat{\sigma}^2$ and σ^2 are close due to 22. Thus, $\sigma^2\widehat{\sigma}^2 \gtrsim 1$.

$$
\begin{aligned}
\left\|\widehat{\Omega}_{-1,1}-\Omega_{-1,1}\right\|_1 &= \left\|\frac{\widehat{\beta}}{\widehat{\sigma}^2}-\frac{\beta}{\sigma^2}\right\| \\
&\leq \left\|\frac{\widehat{\beta}}{\widehat{\sigma}^2}-\frac{\beta}{\widehat{\sigma}^2}\right\|_1+\left\|\frac{\beta}{\widehat{\sigma}^2}-\frac{\beta}{\sigma^2}\right\|_1 \\
&= \frac{1}{\widehat{\sigma}^2}\left\|\widehat{\beta}-\beta\right\|_1+\|\beta\|_1\left|\frac{1}{\widehat{\sigma}^2}-\frac{1}{\sigma}^2\right| \\
&\lesssim s\sqrt{\frac{\log p}{n}}+\sqrt{s}\cdot\left(\frac{s\log p}{n}+\sqrt{\frac{\log p}{n}}\right) \\
&\lesssim s\sqrt{\frac{\log p}{n}}
\end{aligned} \tag{24}
$$

Note that we used the estimate $\|\beta\|_1 \lesssim \sqrt{s}$ in the second inequality. This is valid because

$$
\begin{aligned}
\|\beta\|_1 &\leq \sqrt{s}\,\|\beta\| \\
&= \sqrt{s}\left\|\sigma^2\Omega_{-1,1}\right\| \\
&\lesssim \sqrt{s}\sigma^2\,\|\Omega\| \\
&= \sqrt{s}\sigma^2\lambda_{\max} \\
&\lesssim \sqrt{s}
\end{aligned} \tag{25}
$$

Thus we have

$$
\mathbf{Pr}\left(\left\|\widehat{\Omega}_{*1}-\Omega_{*1}\right\|_1 > Cs\sqrt{\log p/n}\right) \leq p^{-10} \tag{26}
$$

Using union bound, we have

$$
\mathbf{Pr}\left(\max_{1\leq j\leq p}\left\|\widehat{\Omega}_{*j}-\Omega_{*j}\right\|_1 > Cs\sqrt{\log p/n}\right) \leq p^{-9} \tag{27}
$$

Thus the following holds w.h.p.

$$
\left\|\widehat{\Omega}-\Omega\right\|_{op}^2 \leq \left\|\widehat{\Omega}-\Omega\right\|_1^2 \lesssim \frac{s^2\log p}{n} \tag{28}
$$

which is also the optimal rate (See [Cai et al., 2012])

$$\inf_{\widehat{\Omega}} \sup_{\Omega \in \mathcal{G}(p,s,m)} \mathbf{E} \left\| \widehat{\Omega} - \Omega \right\|_{op}^2 \gtrsim \frac{s^2 \log p}{n} \tag{29}$$

References

[Cai et al., 2012] Cai, T. T., Liu, W., and Zhou, H. H. (2012). Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation.

[Meinshausen and Bühlmann, 2006] Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34(3):14361462.