Graphic Lasso: Scaled membership with intercept

Jiaxin Hu

April 8, 2021

1 Identifiability

Consider the model

$$\Omega^k = \Theta^0 + \sum_{l=1}^{r} u_{kl}\Theta^l, \quad k \in [K]. \tag{1}$$

Let $U = [\![u_{kl}]\!] \in \mathbb{R}^{K\times r}$ be the membership matrix and u_l denote the l-th column of U. Let $I_l = \{k : u_{kl} \neq 0\}$ for $l \in [r]$ and $I_0 = \{k : u_{kl} = 0, l \in [r]\}$.

Lemma 1 (Identifiability of scaled membership model with intercept)**.** *Suppose the parameter (U, Θ^l) satisfies the following condition.*

1. *$\Theta^0, \Theta^1, ..., \Theta^l$ are positive definite with bounded singular values, i.e., $0 < \tau_1 \leq \min_{l=0,1,...r} \varphi_{\min}(\Theta^l) \leq \max_{l=0,1,...r} \varphi_{\max}(\Theta^l) \leq \tau_2 < \infty$.*

2. *$\Theta^l, l = 0, 1, ..., r$ are irreducible in the sense that $\Theta^l \neq C\Theta^{l'}$ for any pair l, l' and for any constant C.*

3. *The columns of U are non-overlap, with $\|u_l\|_F = 1$.*

4. *For all $l \in [r]$, we have $\sum_{k=1}^{K} u_{kl} = 0$.*

Then, the parameters in model (1) are identifiable.

Intuition

Consider the linear regression model

$$\mathbb{E}[Y] = X\beta, \tag{2}$$

where $Y \in \mathbb{R}^K$, $X \in \mathbb{R}^{K\times(r+1)}$, and $\beta = (\beta_0, \beta_1, ..., \beta_r)^T$. The linear regression model (2) is a scalar analogy of model (1). Let $Y_k = \Omega^k, k = 1, ..., K$, $X = [1_K, U]$, and $\beta_i = \Theta^i, i = 0, 1, ..., r$. Then, we obtain the model (1).

Hence, we may get some intuitions of the identifiability problem from the simple model (2).

1. By the textbook, when X is fixed, the necessary and sufficient condition to identify β is that X has full rank. Thus, we rule out the case that $|I_0| = 0$ and $u_{kl} = u_{k'l}$, for all $k, k' \in I_l$ and $l \in [r]$.

2. Unlike the linear regression model in textbook, X and β are both unknown, and we also want to know the identifiability of X. Suppose there exist another pair of parameters $\tilde{X}, \tilde{\beta}$ such that $X\beta = \tilde{X}\tilde{\beta}$. Then, we must have $C(X) = C(\tilde{X})$.

Though the first column of $X, \tilde{X}$ are both 1_K, it is possible $X \neq \tilde{X}$ and $C(X) = C(\tilde{X})$. A simple example is that $X = [1_K, e_1]$ and $\tilde{X} = [1_K, 1_K + e_1]$. In previous note, the old condition 4 only guarantees the full rankness. Therefore, we need stronger condition to identify the unique X from the unique column space $C(X)$.

The new condition 4 requires the matrix X to be an orthogonal matrix. The following proof will show that the orthogonality and non-overlapping is sufficient to identify X from $C(X)$.

Proof. Suppose $\{\tilde{U}, \tilde{\Theta}^l\}$ also satisfy the model (1). By condition 4, we have

$$\begin{aligned}
\sum_{k=1}^{K} \Omega^k &= K\Theta^0 + \sum_{k=1}^{K}\sum_{l=1}^{r} u_{kl}\Theta^l = K\Theta^0 \\
&= K\tilde{\Theta}^0 + \sum_{k=1}^{K}\sum_{l=1}^{r} \tilde{u}_{kl}\tilde{\Theta}^l = K\tilde{\Theta}^0,
\end{aligned}$$

which implies that $\Theta^0 = \tilde{\Theta}^0$.

1. Suppose $I_l \neq \tilde{I}_l$ for some $l = 0, 1, ..., r$. Then, there exist a pair $k, k' \in I_l$ but $k \in \tilde{I}_l$ and $k' \in \tilde{I}_l'$. That is

$$u_{kl}\Theta^l = \tilde{u}_{kl}\tilde{\Theta}^l, \quad u_{k'l}\Theta^l = \tilde{u}_{k'l'}\tilde{\Theta}^{l'},$$

   which implies that $\tilde{\Theta}^l = C\tilde{\Theta}^{l'}$ for some constant C. This contradicts to the condition 2.

2. Suppose $I_l = \tilde{I}_l$ for all $l = 0, 1, ..., r$. Then, for all k, l we have

$$u_{kl}\Theta^l = \tilde{u}_{kl}\tilde{\Theta}^{l'},$$

   which implies $\Theta^l = C\tilde{\Theta}^l$ for some constant C. Thus, we have $u_{kl}C = \tilde{u}_{kl}$. By condition 3, note that $\|u_l\|_F = C\|\tilde{u}_l\| = 1$. We have $C = 1$. Therefore, we have $U = \tilde{U}$ and $\Theta^l = \tilde{\Theta}^l$.

□

Necessity

However, condition 4 is not a necessary condition for identifiable. Here is the counterexample.

Example 1. Consider the case $K = 6, r = 2$ with following membership matrix

$$[1_K, U] = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & 0 \\ 1 & \frac{\sqrt{2}}{2} & 0 \\ 1 & 0 & \frac{1}{2} \\ 1 & 0 & \frac{1}{2} \\ 1 & 0 & \frac{\sqrt{2}}{2} \end{bmatrix},$$

and $\Theta^0, \Theta^1, \Theta^2$ satisfy the condition 1,2. Suppose there exist another set of parameters $\{\tilde{U}, \tilde{\Theta}^l\}$ lead to the same model.

1. Suppose $I_l \neq \tilde{I}_l$ for some $l = 0, 1, 2$.

(a) Suppose there exist $k \in I_1$ and $k' \in I_2$ such that $k, k' \in \tilde{I}_0$. We have

$$\Theta^0 + u_{k1}\Theta^1 = \Theta^0 + u_{k'2}\Theta^2 = \tilde{\Theta}^0,$$

which contradicts to the condition 2 that $\Theta^1 \neq C\Theta^2$.

(b) Without the loss of generality, suppose $|I_1 \cap \tilde{I}_0| > 0$ and $|I_2 \cap \tilde{I}_0| = 0$. Then, for $k \in I_1 \cap \tilde{I}_0$, we have

$$\Theta^0 - \tilde{\Theta}^0 = -u_{k1}\Theta^1. \tag{3}$$

Note that $u_{21} \neq u_{31}$. Thus, there also exists a k' such that

$$\Theta^0 - \tilde{\Theta}^0 = \tilde{u}_{k'1}\tilde{\Theta}^1 - u_{k'1}\Theta^1,$$

which implies that $\tilde{\Theta}^1 = \frac{u_{k'1} - u_{k1}}{\tilde{u}_{k'1}}\Theta^1$. If $I_2 = \tilde{I}_2$, then by $k = 5, k' = 6$, we have $\Theta^0 - \tilde{\Theta}^0 = c\Theta^2$ for some constant c, and this contradicts to (3). If $I_2 \neq \tilde{I}_2$, then there exists a $k \in I_2 \cap \tilde{I}_1$ and

$$\Theta^0 + u_{k2}\Theta^2 = \tilde{\Theta}^0 + \tilde{u}_{k1}\tilde{\Theta}^1,$$

which contradictions to the condition $\Theta^2 \neq C\Theta^1$.

(c) Suppose $|\tilde{I}_0| = 0$. Without the loss of generality, let $|\tilde{I}_1| \geq |I_1|$ and thus there exists a $k \in I_2 \cap \tilde{I}_1$ and $k', k'' \in I_1 \cap \tilde{I}_1$ with $\tilde{u}_{k'1} \neq u_{k''1}$. Then, we have

$$\begin{aligned}\Theta^0 - \tilde{\Theta}^0 &= \tilde{u}_{k1}\tilde{\Theta}^1 - u_{k2}\tilde{\Theta}^2 \\ &= \tilde{u}_{k'1}\tilde{\Theta}^1 - u_{k'1}\Theta^1 \\ &= \tilde{u}_{k''1}\tilde{\Theta}^1 - u_{k''1}\Theta^1,\end{aligned}$$

which implies that $\tilde{\Theta}^1 = c\Theta^1$ and $\tilde{\Theta}^1 = c_1\Theta^1 - c_2\Theta^2$ for some constants c, c_1, c_2. This contradicts to the condition 2.

2. Suppose $I_l = \tilde{I}_l, l = 1, 2$. By $k = 2, 3$ and $k = 5, 6$, we know that $\tilde{\Theta}^1 = c_1\Theta^1$ and $\tilde{\Theta}^2 = c_2\Theta^2$ for some constant c_1, c_2. By the assumption that $\|u_l\| = \|\tilde{u}_l\| = 1$, we obtain that $c_1 = c_2 = 1$.

2 Conjectures

Here is the counterexample. Given that $\Theta^1 - \Theta^2$ is still positive definite, we have

$$\Omega = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} \Theta^0 \\ \Theta^1 \\ \Theta^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \Theta^0 + \Theta^2 \\ \Theta^1 - \Theta^2 \\ \Theta^2 \end{bmatrix}$$

The right hand side still satisfies the non-overlapping assumptions. My conjecture is that the identifiability holds if at least two subspaces are non-parallel with intercept. In the following example, we can not do the same transformation as above.

$$\Omega = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} \Theta^0 \\ \Theta^1 \\ \Theta^2 \\ \Theta^3 \end{bmatrix}.$$

However, the conjecture that at least two subspaces are non-parallel with intercept is still not enough for identifiability. Consider another example

$$\Omega = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} \Theta^0 \\ \Theta^1 \\ \Theta^2 \end{bmatrix}.$$

Let $\tilde{\Theta}^1 = \Theta^1 - \Theta^2$ and $\tilde{\Theta}^2 = 2\Theta^2 - 3\Theta^2$. Then we have

$$\begin{aligned} \Theta^0 + \Theta^1 &= \tilde{\Theta}^0 + c_1\tilde{\Theta}^1 \\ \Theta^0 + 2\Theta^1 &= \tilde{\Theta}^0 + c_2\tilde{\Theta}^2 \\ \Theta^0 + \Theta^2 &= \tilde{\Theta}^0 + c_3\tilde{\Theta}^1 \\ \Theta^0 + 3\Theta^2 &= \tilde{\Theta}^0 + c_4\tilde{\Theta}^2, \end{aligned}$$

and

$$4\Theta^0 - 4\tilde{\Theta}^0 = (c_1 + c_3)(\Theta^1 - \Theta^2) + (c_2 + c_4)(2\Theta^1 - 3\Theta^2) - 3\Theta^! - 4\Theta^2.$$

Solving the equations, we have the solutions

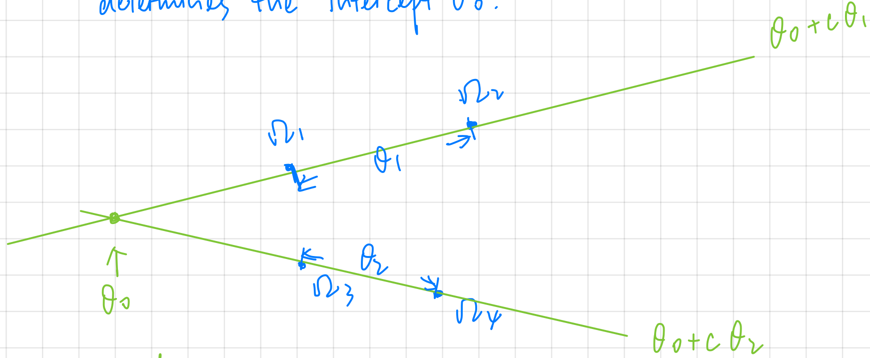
$$c_1 = -3, \quad c_2 = -1, \quad c_3 = -4, \quad c_4 = -2, \quad \tilde{\Theta}^0 = \Theta^0 + 4\Theta^1 - 3\Theta^2.$$

Therefore, we have

$$\Omega = \begin{bmatrix} 1 & -3 & 0 \\ 1 & 0 & -1 \\ 1 & -4 & 0 \\ 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} \Theta^0 + 4\Theta^1 - 3\Theta^2 \\ \Theta^1 - \Theta^2 \\ 2\Theta^1 - 3\Theta^2 \end{bmatrix}.$$

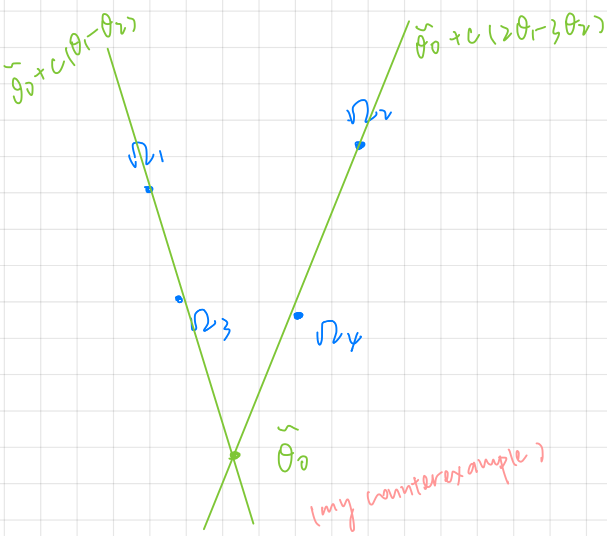
Thus, the parameters are not identifiable. See the following figure for graphic illustration.

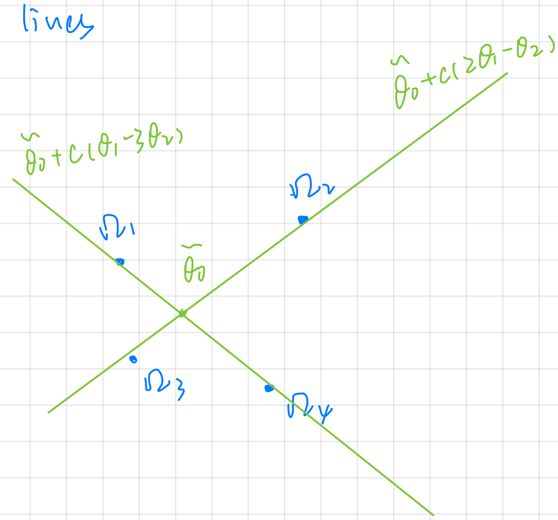
Claim: The intersection of $\{\theta_0 + c\theta_1, c \in \mathbb{R}\}$ and $\{\theta_0 + c\theta_2, c \in \mathbb{R}\}$ determines the intercept θ_0.



Two non-parallel subspaces uniquely determine the intersection.

However, 4 different points lead to different lines





Therefore, at least two subspaces not align with intercept is not enough for identifiability.