arXiv:2303.01326v1 [math.ST] 2 Mar 2023

融合图形套索在多个稀疏精度矩阵统计推断中的应用

张秋燕[a]、白志东[b]、李凌睿[a]、杨虎[c,*]

[a]首都经济贸易大学统计学院，北京[b]东北师范大学*KLASMOE*与数学与统计学院，吉林长春[c]中央财经大学信息学院，北京

摘要

本文使用融合图形套索（FGL）方法，同时从多个总体中估计多个精度矩阵。FGL模型中的套索惩罚项限制了精度矩阵的稀疏性，而对来自不同组的两个精度矩阵施加适度惩罚，则限制了多个组之间的相似结构。在高维情形下，本文为FGL估计量提供了预言机不等式，这是建立中心极限定律所必需的。本文不仅关注精度矩阵的点估计，还研究了对多个精度矩阵元素线性组合的假设检验。受Jankov$\acute{a}$ 和van de Geer [关于高维逆协方差估计的置信区间研究的启发，《电子统计学杂志》**9**(1) (2015) 1205-1229.], 他们研究了去偏技术，以获得具有已知分布的新相合估计量，从而实施统计推断；本文将统计推断问题扩展到多个总体，并提出去偏FGL估计量。本文给出了去偏FGL估计量相应的渐近性质。模拟研究表明，所提出的检验在高维情形下表现良好。

关键词： 图形套索、高维数据分析、假设检验

1. 引言

无向图模型是表示数据网络结构的常用工具，已广泛应用于机器学习、遗传学和金融等多个领域。设$\mathbf{x} = (\mathbf{x}^1, ..., \mathbf{x}^p)^T$为一个均值向量为 μ、协方差为 Σ_0（Σ_0为正定）的 p 维正态随机向量，其精度矩阵（或浓度矩阵）定义为协方差矩阵的逆矩阵，即$\Theta_0 := \Sigma_0^{-1}$。图模型通过精度矩阵中的非零元素来描述随机变量之间的条件依赖关系。若 $\Theta_{0ij} \neq 0$，在给定所有其他变量的条件下，$\mathbf{x}^i$与$\mathbf{x}^j, i, j = 1, ..., p$彼此依赖。与此同时，精度矩阵中的零元素对应于给定其他变量后条件独立的变量对。因此，图模型与精度矩阵密切

*通讯作者电子邮箱： hu.yang@cufe.edu.cn（杨虎）

相关。过去几年中，精度矩阵的估计与检验已成为一个快速发展的研究方向。

设$\mathbf{x}_1, ..., \mathbf{x}_n$是来自总体$\mathbf{x}$的独立同分布（i.i.d.）观测序列，$\mathbf{X}_{p\times n} := (\mathbf{x}_1, ..., \mathbf{x}_n)$。精度矩阵的一个自然估计量是样本协方差矩阵 $\widehat{\Sigma}$的逆，其中 $\widehat{\Sigma} = \frac{1}{n}\mathbf{X}^T\mathbf{X}$。一方面，在高维情形下，Johnstone[10] 指出，对于 $\Sigma = \mathbf{I}$，样本协方差矩阵的特征值不会收敛到总体协方差矩阵相应的特征值。因此，当维度 p 与样本量 n相当时，该估计量便会失效。另一方面，在 $p > n - 1$ 情形下，样本协方差矩阵是奇异的。使用 $\widehat{\Sigma}_n^{-1}$估计 Θ_0会因此产生不可忽略的误差。此外，高维精度矩阵具有稀疏性（即许多元素为零或近似为零）这一假设至关重要，因为零元素意味着条件独立结构，而这正是图模型中我们最关注的内容。一般而言， $\widehat{\Sigma}^{-1}$ 并不具有稀疏结构。如何在高维情形下估计稀疏精度矩阵，是一个棘手的问题。

近年来，人们提出了多种方法来估计高维情形下的精度矩阵，其中带稀疏促进惩罚项的图模型适用于获得稀疏估计量。Yuan 和 Lin [16] 通过对浓度矩阵的各元素施加 l_1 （套索惩罚），提出了一种最大行列式算法，以获得 Θ_0的估计量。该估计量的收敛性结论是在 p 固定的假设下推导的。Friedman 等人[5] 利用坐标下降过程提出了一种求解图形 Lasso 估计量的算法，速度非常快，即使 $p > n$也是如此。Rothman 等人 [13]研究了一种稀疏置换不变协方差估计量，并建立了其在弗罗贝尼乌斯范数下的收敛速率；当数据维度 p 和样本量 n 均允许增长时，该速率明确取决于真实浓度矩阵的稀疏程度。关于图模型惩罚似然方法的更多理论细节，请参见 Fan 等人[4], 、Ravikumar 等人 [11], 、Xue 和 Zou [14],以及 Yuan 等人[17]。

上述方法着重于估计单个图模型，但与针对多个图模型分别进行估计相比，对于来自多个图模型的数据，联合估计量能更好地恢复真实图；这些图模型具有相似的结构，但并不完全相同。Guo 等人 [6] 研究了具有层次结构假设的精度矩阵的联合估计。Liu 和 Lee 提出了一种联合估计量，假设多个精度矩阵可分解为两个分量之和。Danaher 等人 [3] 提出了融合图形 Lasso，其惩罚项要求不同组之间的精度矩阵具有相似结构。假设$\mathbf{X}_{p\times n_k}^{[k]} := (\mathbf{x}_1^{[k]}, ..., \mathbf{x}_{n_k}^{[k]})$是样本矩阵，且$\mathbf{x}_i^{[k]} \in R^p (i = 1, ..., n_k)$是从均值为 $\mu^{[k]}$、协方差为$\Sigma_0^{[k]}$ 的分布中独立同分布抽取的样本，其中 $k = 1, ..., K$，不失一般性，我们假设 $\mu^{[k]} = 0$。为简化记号，我们省略$\mathbf{X}_{p\times n_k}^{[k]}$的下标，并将样本矩阵记为$\mathbf{X}^{[k]}$。总体精度矩阵定义为总体协方差矩阵的逆，即 $\Theta_0^{[k]} = (\Sigma_0^{[k]})^{-1}$。精度矩阵的估计量

$\{\Theta_0^{[k]}\}$ 通过最小化负惩罚对数似然进行研究

$$\{\widehat{\Theta}^{[k]}\} = \arg\min_{\{\Theta^{[k]}\in\mathcal{S}^{++}\}} \sum_k \{tr(\widehat{\Sigma}^{[k]}\Theta^{[k]}) - \log\det(\Theta^{[k]})\} + \mathbf{P}(\{\Theta^{[k]}\}), \tag{1}$$

其中$\mathbf{P}(\{\Theta^{[k]}\})$表示惩罚函数，$\{\widehat{\Theta}^{[k]}\}$ 是式 (1) 的极小化解，我们在对称正定矩阵集合 $\mathcal{S}^{++}$上进行优化。融合图形套索（FGL）是采用融合套索惩罚求解优化问题 (1) 得到的解

$$\mathbf{P}(\{\Theta^{[k]}\}) = \lambda\sum_{k=1}^{K}||(\Theta^{[k]})^{-}||_1 + \rho\sum_{k<k'}||(\Theta^{[k]} - \Theta^{[k']})^{-}||_1, \tag{2}$$

其中，λ和 ρ 是非负正则化参数，$(\Theta^{[k]})^{-}$ 表示将$(\Theta^{[k]})$的对角元素设为零后得到的矩阵，$||\cdot||_1$ 表示向量或矩阵的 l_1 范数。由于我们最关注不同变量之间的条件独立性，因此合理的做法是对 $\Theta^{[k]}$的非对角元素进行约束。注意，式（2）中的第一项是经典的套索惩罚，它会随着 λ 增大而将系数向0收缩，从而确保发现模型的稀疏估计量 $\{\widehat{\Theta}^{[k]}\}$ 。对$(\Theta^{[k]}-\Theta^{[k']})^{-}$ 的惩罚表明，$\widehat{\Theta}^{[1]}$、..... $\widehat{\Theta}^{[K]}$ 的元素在不同类别间具有相似的网络结构。

联合图模型的估计方法在很大程度上依赖于惩罚估计。惩罚项会使估计值偏向假设的结构，从而令精度矩阵的假设检验更具挑战性。近期已有研究针对图模型中的低维参数开展统计推断（Jankov$\acute{a}$ 和van de Geer [7]；Jankov$\acute{a}$ 和van de Geer [8]；Ren 等人 [12]；Yu 等人 [15]），这些研究基于 l_1惩罚估计量。Jankov$\acute{a}$ 和van de Geer [7] 提出了一种去偏技术，以获得具有已知分布的新相合估计量。然而，这些方法仅适用于推断单个图参数的情形。相比之下，利用跨组惩罚所得估计量进行推断的方法研究较少。多个图模型的统计推断是一个值得未来研究的有趣领域。受Jankov$\acute{a}$ 和van de Geer [7], 的启发，我们不仅基于共同变动数据给出了多个精度矩阵的FGL估计量，还检验了这些精度矩阵元素的线性组合。所提出方法的核心基于去偏技术，并根据所提出的中心极限定理，在高维情形下对精度矩阵进行统计推断。

本文其余部分安排如下。第 2 节给出带有 FGL 惩罚项及其加权版本的多个估计量的预言机不等式。本节还讨论了对多个精度矩阵相应元素的线性组合进行假设检验。基于去偏技术，第 2 节还推导了所提出的多个总体统计量的中心极限定理。第 3 节报告模拟结果。所有技术细节均置于附录中。

2. 主要结果

全文中我们采用以下记号。对于矩阵 $A = (a_{ij})_{i,j=1}^{p}$，我们将 $(A)_{ij}$称为其 (i, j) 元素；为简化记号，也将其 (i, j) 元素记作 A_{ij}。我们用 $|A|$ 表示 A的行列式，矩阵 A 的迹记为 $tr(A)$。令 $A^+ = diag(A)$表示与 A具有相同对角线元素的对角矩阵，$A^- = A - A^+$。$||A||_F^2 = \sum_{i,j} a_{ij}^2$ 表示弗罗贝尼乌斯范数（亦称矩阵二范数）。我们用记号 $||A||_\infty = \max_{i,j} |a_{ij}|$ 表示矩阵 A的上确界范数，并用$|||A|||_1 := \max_j \sum_i |a_{ij}|$ 表示其 l_1-算子范数。

如果对于某个常数 $c < \infty$，$f(n) \leq cg(n)$，我们记作 $f(n) = \mathcal{O}(g(n))$；如果对于某个常数 $c' > 0$，$f(n) \geq c'g(n)$，则记作 $f(n) = \Omega(g(n))$。记号 $f(n) \asymp g(n)$ 表示 $f(n) = \mathcal{O}(g(n))$，且 $f(n) = \Omega(g(n))$。在常见的高维情形下，维度 p 允许趋于无穷。维度与样本量相比，可以相当、大得多或小得多。本文中，我们将样本量设为 $n_1 \asymp ... \asymp n_K \asymp n$，并令 $n^* = n_1 + ... + n_K$ 趋于无穷。此外，为了简化记号，我们假设 $n_1 = ... = n_K = n$。

2.1. 预言机不等式

为了获得 FGL 模型的多个估计量的预言机不等式，我们引入一些与真实精度矩阵各元素的稀疏性假设相关的记号。令

$$S_k := \{(i, j) : \Theta_{0ij}^{[k]} \neq 0, i \neq j\}, \tag{3}$$

其中，$\Theta_{0ij}^{[k]}$ 是 $\Theta_0^{[k]}$ 的（i，j）元素，$s_k = |S_k|$ 是 S_k的基数。我们采用真实精度矩阵的特征值有界性，以及 Jankovᡠ 和 Van De Geer [7]提出的某些尾部条件。

条件 1（有界特征值）。存在普适常数 L，使得对于 k，

$$0 < L < \Lambda_{\min}(\Theta_0^{[k]}) \leq \Lambda_{\max}(\Theta_0^{[k]}) < 1/L < \infty,$$

其中，$\Lambda_{\min}$ 和 $\Lambda_{\max}$ 分别表示矩阵的最小特征值和最大特征值。

条件 2（次高斯向量条件）。观测值 $\mathbf{x}_i^{[k]}$、$i = 1$、.....、n_k分别是各组中的一致次高斯向量。

我们针对 $K = 2$情形提出了 FGL lasso 的预言机不等式。

定理1. 假设条件*1*和条件*2*成立，对于 $k = 1,2$，满足 $2(\rho + \lambda_0) \leq \lambda \leq c/8L$ 的调节参数 λ，以及 $\frac{8\lambda^2(s_1+s_2)}{c} + \frac{4p\lambda_0^2}{c} \leq \lambda_0 c/2L)$。在集合 $\{\max_k ||\widehat{\Sigma}^{[k]} - \Sigma_0^{[k]}||_\infty \leq \lambda_0\}, k = 1,2$ 上，有

$$c\sum_{k=1}^{2} ||\widehat{\Theta}^{[k]} - \Theta_0^{[k]}||_F^2 + \lambda\sum_{k=1}^{2} ||(\widehat{\Theta}^{[k]} - \Theta_0^{[k]})^-||_1 \leq \frac{8\lambda^2(s_1 + s_2)}{c} + \frac{4p\lambda_0^2}{c}, \tag{4}$$

and

$$\sum_{k=1}^{2}|||\widehat{\Theta}^{[k]}-\Theta_0^{[k]}|||_1\leq\frac{4\lambda(8s_1+8s_2+p)}{c},\tag{5}$$

其中 $c=1/(8L^2)$。

注 1. 根据该不等式，我们必须选择 λ，使得 $\lambda p\to 0$ 随着 $n\to\infty$ 趋于.....，以确保一致性，而次高斯随机向量不满足这一条件。因此，条件 $\lambda p\to 0$ 排除了 $p\gg n$情形。

FGL 没有考虑到变量通常具有不同的尺度。因此，我们考虑加权 FGL。采用加权 FGL 惩罚项的优化问题 (1) 的极小化解

$$\mathbf{P}(\{\Theta^{[k]}\})=\lambda\sum_k\sum_{i\neq j}\widehat{W}_{ii}^{[k]}\widehat{W}_{jj}^{[k]}|\Theta_{ij}^{[k]}|+\rho\sum_{k<k'}\sum_{i\neq j}|\widehat{W}_{ii}^{[k]}\widehat{W}_{jj}^{[k]}\Theta_{ij}^{[k]}-\widehat{W}_{ii}^{[k']}\widehat{W}_{jj}^{[k']}\Theta_{ij}^{[k']}|\tag{6}$$

记为 $\{\widehat{\Theta}_w^{[k]}\}$，其中 $\widehat{W}^{[k]}=\left[diag(\widehat{\Sigma}^{[k]})\right]^{\frac{1}{2}}$。此外，总体相关矩阵记为$R_0^{[k]}$，样本相关矩阵记为

$$\widehat{R}^{[k]}=(\widehat{W}^{[k]})^{-1}\widehat{\Sigma}^{[k]}(\widehat{W}^{[k]})^{-1}.\tag{7}$$

如果我们用 $\widehat{R}^{[k]}$ 替换 $\widehat{\Sigma}^{[k]}$，则.....的极小化解

$$\arg\min_{\{\Theta^{[k]}\in\mathcal{S}^{++}\}}\sum_k\{tr(\widehat{R}^{[k]}\Theta^{[k]})-\log\det(\Theta^{[k]})\}+\mathbf{P}(\{\Theta^{[k]}\})\tag{8}$$

采用 FGL 惩罚项 (2) 时记为 $\{\widehat{\Theta}_R^{[k]}\}$，这相当于利用归一化数据估计该参数。于是,

$$\widehat{\Theta}_R^{[k]}=\widehat{W}^{[k]}\widehat{\Theta}_w^{[k]}\widehat{W}^{[k]},\tag{9}$$

这基本上意味着，$\widehat{\Theta}_R^{[k]}$ 是 $\Theta_{R0}^{[k]}$ 的估计量：$=\left(R_0^{[k]}\right)^{-1}$。

定理 2. 在定理*1*的条件下，在集合 $\{\max_k||\widehat{R}^{[k]}-R_0^{[k]}||_\infty\leq\lambda_0\}$, $k=1,2$ 上，有

$$c\sum_{k=1}^{2}||\widehat{\Theta}_R^{[k]}-\Theta_{R0}^{[k]}||_F^2+\lambda\sum_{k=1}^{2}||(\widehat{\Theta}_R^{[k]}-\Theta_{R0}^{[k]})^-||_1\leq\frac{8\lambda^2(s_1+s_2)}{c},\tag{10}$$

$$\sum_{k=1}^{2}|||\widehat{\Theta}_R^{[k]}-\Theta_{R0}^{[k]}|||_1\leq\frac{32\lambda(s_1+s_2)}{c},\tag{11}$$

and

$$\sum_{k=1}^{2}|||\widehat{\Theta}_w^{[k]}-\Theta_0^{[k]}|||_1\leq\frac{32\lambda(s_1+s_2)}{c}.\tag{12}$$

自然可以将这一结论推广到 $K > 2$ FGL模型。对于 $k = 1, ..., K$ 以及 $K > 2$情形，我们得到以下定理。

定理 3（多重FGL模型）。假设条件*1*和条件*2*成立，对于 $K > 2, 2\left(\binom{K(K-1)}{2}\rho + \lambda_0\right) \leq \lambda \leq c/8L$ 和 $\frac{8\lambda^2 \sum_{k=1}^{K} s_k}{c} + \frac{2Kp\lambda_0^2}{c} \leq \lambda_0/2L$，在集合 $\{\max_k ||\widehat{\Sigma}^{[k]} - \Sigma_0^{[k]}||_\infty \leq \lambda_0\}, k = 1, ..., K$ 上，有

$$c\sum_{k=1}^{K} ||\widehat{\Theta}^{[k]} - \Theta_0^{[k]}||_F^2 + \lambda \sum_{k=1}^{K} ||(\widehat{\Theta}^{[k]} - \Theta_0^{[k]})^-||_1 \leq \frac{8\lambda^2 \sum_{k=1}^{K} s_k}{c} + \frac{2Kp\lambda_0^2}{c} \tag{13}$$

and

$$\sum_{k=1}^{K} |||\widehat{\Theta}^{[k]} - \Theta_0^{[k]}|||_1 \leq \frac{2K\lambda\left(8\sum_{k=1}^{K} s_k + \frac{Kp}{2}\right)}{c}. \tag{14}$$

定理 4（加权版本的多重FGL模型）。在定理 *3* 的条件下，在集合$\{\max_k ||\widehat{R}^{[k]} - R_0^{[k]}||_\infty \leq \lambda_0\}$, $k = 1,2$上，有

$$c\sum_{k=1}^{K} ||\widehat{\Theta}_R^{[k]} - \Theta_{R0}^{[k]}||_F^2 + \lambda \sum_{k=1}^{K} ||(\widehat{\Theta}_R^{[k]} - \Theta_{R0}^{[k]})^-||_1 \leq \frac{8\lambda^2 \sum_{k=1}^{K} s_k}{c}, \tag{15}$$

$$\sum_{k=1}^{K} |||\widehat{\Theta}_R^{[k]} - \Theta_{R0}^{[k]}|||_1 \leq \frac{16K\lambda \sum_{k=1}^{K} s_k}{c}, \tag{16}$$

and

$$\sum_{k=1}^{K} |||\widehat{\Theta}_w^{[k]} - \Theta_0^{[k]}|||_1 \frac{16K\lambda \sum_{k=1}^{K} s_k}{c}. \tag{17}$$

2.2 渐近性质

我们不仅关注多个精度矩阵的点估计，还关注两个组中精度矩阵元素线性组合的假设检验。我们可能希望检验两个组中精度矩阵的元素是否相等:

$$H_0 : \Theta_{0ij}^{[1]} = \Theta_{0ij}^{[2]} \quad vs. \quad H_1 : \Theta_{0ij}^{[1]} \neq \Theta_{0ij}^{[2]}. \tag{18}$$

为检验假设（18），我们旨在获得基于去偏技术的估计量的置信区间，该技术用于消除与惩罚项相关的偏差。去偏估计量定义为 $\widehat{\Theta}_d^{[k]} = 2\widehat{\Theta}^{[k]} - \widehat{\Theta}^{[k]}\widehat{\Sigma}^{[k]}\widehat{\Theta}^{[k]}$。去偏估计量与真实值之间的差异可分解为以下两部分:

$$\widehat{\Theta}_d^{[k]} - \Theta_0^{[k]} = \Xi^{[k]} + \Upsilon^{[k]}, \tag{19}$$

其中

$$\Xi^{[k]} = -\Theta_0^{[k]}(\widehat{\Sigma}^{[k]} - \Sigma_0^{[k]})\Theta_0^{[k]}, \tag{20}$$

$$\Upsilon^{[k]} = -(\widehat{\Theta}^{[k]} - \Theta_0^{[k]})(\widehat{\Sigma}^{[k]} - \Sigma_0^{[k]})\Theta_0^{[k]} - (\widehat{\Theta}^{[k]} - \Theta_0^{[k]})(\widehat{\Sigma}^{[k]}\widehat{\Theta}^{[k]} - \mathbf{I}_p). \tag{21}$$

在兼容性条件下，Jankov´a 和 van de Geer [9] 提出，$\widehat{\Theta}_d^{[k]} - \Theta_0^{[k]}$ 的（i, j）元素具有渐近正态性，且 $\sqrt{n}||\Upsilon^{[k]}||_\infty$ 依概率收敛于零。因此，为检验假设（18），我们构造检验统计量

$$T_{ij} := \left(\widehat{\Theta}_d^{[1]} - \widehat{\Theta}_d^{[2]}\right)_{ij} = \left[2\widehat{\Theta}^{[1]} - \widehat{\Theta}^{[1]}\widehat{\Sigma}^{[1]}\widehat{\Theta}^{[1]} - (2\widehat{\Theta}^{[2]} - \widehat{\Theta}^{[2]}\widehat{\Sigma}^{[2]}\widehat{\Theta}^{[2]})\right]_{ij} \tag{22}$$

使用去偏估计量。

对于 $K = 2$，令

$$s = \max\{s_1, s_2\}, \quad d = \max\{d_1, d_2\}, \tag{23}$$

其中

$$d_k = \max_{j=1,\ldots,p} |D_j^{[k]}|, \quad D_j^{[k]} = \{(i, j) : \Theta_{0ij}^{[k]} \neq 0, i \neq j\}. \tag{24}$$

接下来，我们为 T_{ij}建立中心极限定理。

定理5. 假设条件 *1*、*2*，以及 $\lambda \asymp \rho \asymp \sqrt{\log\, p/n}$ 和 $\overline{(p+s)}\sqrt{d} = o(\sqrt{n}/\log\, p)$，则有

$$\widehat{\Theta}_d^{[1]} - \widehat{\Theta}_d^{[2]} - (\Theta_0^{[1]} - \Theta_0^{[2]}) = \Xi^{[1]} - \Xi^{[2]} + rem, \tag{25}$$

其中

$$||rem||_\infty = ||\Upsilon^{[1]} - \Upsilon^{[2]}||_\infty = o_p(1/\sqrt{n}), \tag{26}$$

且 o_p 表示依概率收敛。此外，

$$\sqrt{n}\big[T_{ij} - \mathbf{\Theta}_{0ij}\big] \to_D N(0, \sigma_{ij}^2), \tag{27}$$

其中 $\mathbf{\Theta}_{0ij} = (\Theta_0^{[1]} - \Theta_0^{[2]})_{ij}$。

为完成检验过程，我们使用相合估计量 $\hat{\sigma}_{ij}^2 = (\widehat{\Theta}^{[1]})_{ii}(\widehat{\Theta}^{[1]})_{jj} + (\widehat{\Theta}^{[1]})_{ij}^2 + (\widehat{\Theta}^{[2]})_{ii}(\widehat{\Theta}^{[2]})_{jj} + (\widehat{\Theta}^{[2]})_{ij}^2$来处理定理5。定理5为获取该检验统计量的p值和临界值提供了一种实用且高效的方法。在原假设下，我们观察到 $\Theta_{0ij}^{[1]} - \Theta_{0ij}^{[2]} = 0$。对于显著性水平为 α 的检验，若 $|\sqrt{n}T_{ij}/\hat{\sigma}_{ij}^2| > \xi_{\alpha/2}$，则拒绝 H_0，其中 ξ_α 是标准正态分布的上分位数。

定理5所需的稀疏性条件比定理1中相应的预言机型不等式更强。根据$(p+s)\sqrt{d}$的收敛速率，定理5适用于 $p \ll n$ 情形。对于$p \gg n$，我们给出以下定理。

定理 6. 假设条件 *1*、*2*，以及 $\lambda \asymp \rho \asymp \sqrt{\log\ p/n}$ 和 $\overline{s\sqrt{d} = o}(\sqrt{n}/\overline{\log}\ p)$，对于 $p \ll n$情形，等式 *(32)* 在 $\widehat{\Theta}_w^{[k]}$下成立，其中

$$||rem||_\infty = o_p(1/\sqrt{n}). \tag{28}$$

此外，

$$\sqrt{n}\big[T_{wij} - \mathbf{\Theta}_{0ij}\big] \to_D N(0, \sigma_{ij}^2), \tag{29}$$

其中 $T_{wij} = (2\widehat{\Theta}_w^{[1]} - \widehat{\Theta}_w^{[1]}\widehat{\Sigma}^{[1]}\widehat{\Theta}_w^{[1]})_{ij} - (2\widehat{\Theta}_w^{[2]} - \widehat{\Theta}_w^{[2]}\widehat{\Sigma}^{[2]}\widehat{\Theta}_w^{[2]})_{ij}$。

与 Brownlees 等人 [2]不同，我们无需对 Σ 施加所谓的不可表示条件，即可推导出我们估计量的理论性质。

此外，对于多样本精度矩阵假设问题，我们可能希望检验一个线性假设检验问题：

$$H_0 : a_1\Theta_{0ij}^{[1]} + ... + a_K\Theta_{0ij}^{[K]} = 0 \quad vs. \quad H_1 : \text{not } H_0, \tag{30}$$

其中 $a_1, ..., a_K$ 为已知常数。与双样本情形类似，我们提出了如下检验统计量：

$$a_1\widehat{\Theta}_{dij}^{[1]} + ... + a_K\widehat{\Theta}_{dij}^{[K]}. \tag{31}$$

对于 $K > 2$ 多重情形，我们假设 $s = \max\{s_1, ..., s_K\}$ 且 $d = \max\{d_1, ..., d_K\}$。因此，我们在以下推论中证明了所提出统计量的渐近正态性，即推论 1。

推论 1。 在定理*5*的假设下，有

$$f\big(\widehat{\Theta}_d^{[1]}, ..., \widehat{\Theta}_d^{[K]}\big) - f\big(\Theta_0^{[1]}, ..., \Theta_0^{[K]}\big) = f\big(\Xi^{[1]}, ..., \Xi^{[K]}\big) + rem, \tag{32}$$

$$||rem||_\infty = ||f\big(\Upsilon^{[1]}, ..., \Upsilon^{[K]}\big)||_\infty = o_p(1/\sqrt{n}), \tag{33}$$

其中 $f\ (x_1、...、\ x_K)\ = a_1x_1 + ... + a_Kx_K$。此外，

$$\sqrt{n}\big[T_{ij} - \mathbf{\Theta}_{0ij}\big] \to_D N(0, \sigma_{ij}^2), \tag{34}$$

其中 $T_{ij} = f\left(\widehat{\Theta}_{dij}^{[1]}, ..., \widehat{\Theta}_{dij}^{[K]}\right)$ 和 $\mathbf{\Theta}_{0ij} = f\left(\Theta_{0ij}^{[1]}, ..., \Theta_{0ij}^{[K]}\right)$。

推论 1 中的渐近方差 σ_{ij} 未知，因此，为构造置信区间，我们使用相合估计量

$$\hat{\sigma}_{ij}^2 = f_v\big(\big[(\widehat{\Theta}^{[1]})_{ii}(\widehat{\Theta}^{[1]})_{jj} + (\widehat{\Theta}^{[1]})_{ij}^2\big], ..., \big[(\widehat{\Theta}^{[K]})_{ii}(\widehat{\Theta}^{[K]})_{jj} + (\widehat{\Theta}^{[K]})_{ij}^2\big]\big), \tag{35}$$

where $f_v(x_1、...、\ x_K) = a_1^2x_1 + ... + a_K^2x_K$。此外，提出了如下加权版本 .

推论 2. 在满足定理 *6* 的假设时，*(33)* 中的残差项以速率 $1/\sqrt{n}$ 依概率收敛；将 $\widehat{\Theta}^{[k]}$ 替换为通过求解加权 *FGL* 优化问题得到的 $\widehat{\Theta}_w^{[k]}$ 后，*(34)* 中的中心极限定理仍成立。

3. 数值研究

我们开展了模拟实验，以评估所提出的去偏 FGL 检验的性能。我们考虑了稀疏图模型，并从多元正态分布 $N(0_p, (\Theta_0^{[k]})^{-1})$ 中生成随机样本，其总体协方差矩阵定义为总体精度矩阵的逆矩阵。

为求解具有特定惩罚项的图形套索问题，我们采用交替方向乘子法（ADMM）算法，因为该算法保证收敛至全局最优解。详情请参阅 Boyd 等人 [1] 和 Danaher 等人 [3]。若需客观地选择调优参数 λ 和 ρ，可使用赤池信息准则（AIC）、贝叶斯信息准则或交叉验证法的近似值来选择调优参数。以下模拟采用 AIC 方法，λ 和 ρ 均从 0.05 递增至 0.3，步长为 0.0086，其中步长由 $(0.3-0.05)/(30-1)$ 得出。

此外，所有报告的模拟结果均基于 500 次模拟，名义显著性水平为 0.05，且维度设为 100。

3.1. 检验的波动

我们通过模拟数据展示了针对双样本问题 (18) 进行检验时的理论渐近正态性结果，并在原假设下令精度矩阵相等，即 $\Theta_0^{[1]} = \Theta_0^{[2]}$。

令 G 为一个对角元素为 0、$\tilde{\alpha}$ 百分比的非对角元素为 1 的 $p \times p$ 对称图矩阵，令 U 为一个元素由均匀分布在 $\tilde{}$ $\tilde{}$ 区间 $(0,1)$ 上独立同分布生成的 $p \times p$ 矩阵，即 $U(0,1)$，我们将对称矩阵 $\boldsymbol{\Theta}$ 的元素记为 θ_{ij}。对于 $i > j$，

$$\tilde{\theta}_{ij} = \frac{g_{ij}u_{ij} + g_{ji}u_{ji}}{2} - \mathbf{1}_{\{\frac{g_{ij}u_{ij}+g_{ji}u_{ji}}{2} < 0.5\}}, \tag{36}$$

其中，g_{ij} 和 u_{ij} 分别是 G 和 U 的 (i,j) 元素，$\mathbf{1}_{\{\cdot\}}$ 是示性函数。对于 $i < j$，我们令 $\tilde{}$ $\tilde{}$ $\tilde{}$ $\theta_{ij} = \theta_{ji}$。矩阵 $\boldsymbol{\Theta}$ 的对角元素为零。然后，精度矩阵生成为

$$\Theta_0^{[k]} = \widetilde{\boldsymbol{\Theta}} + \left(|\Lambda_{\min}(\widetilde{\boldsymbol{\Theta}})| + 0.1\right)\mathbf{I}_p. \tag{37}$$

这表明生成的矩阵是对称正定的。为了使非零元素远离 0 并生成稀疏矩阵，我们从非零元素中减去 1。此外，精度矩阵的生成过程表明，$\tilde{\alpha}$ 是控制稀疏性的参数。当 $\tilde{\alpha} = 1$时，会生成稠密矩阵。众所周知，矩阵的稀疏性不仅要求非零元素数量较少，还要求非零元素的绝对值较大。参数 $\tilde{\alpha}$ 从稀疏元素的数量这一角度控制稀疏性。

我们分别考察了极稀疏和稠密精度矩阵情形下，在 $(p, n) = (100, 200)$ 和 $(p, n) = (100, 400)$ 设置中 $\sqrt{n}T_{ij}/\hat{}\sigma_{ij}$ 的波动。对于极稀疏精度矩阵情形，我们设定参数 $\tilde{\alpha} = 0.01$；对于稠密情形，则使用 $\tilde{\alpha} = 1$。

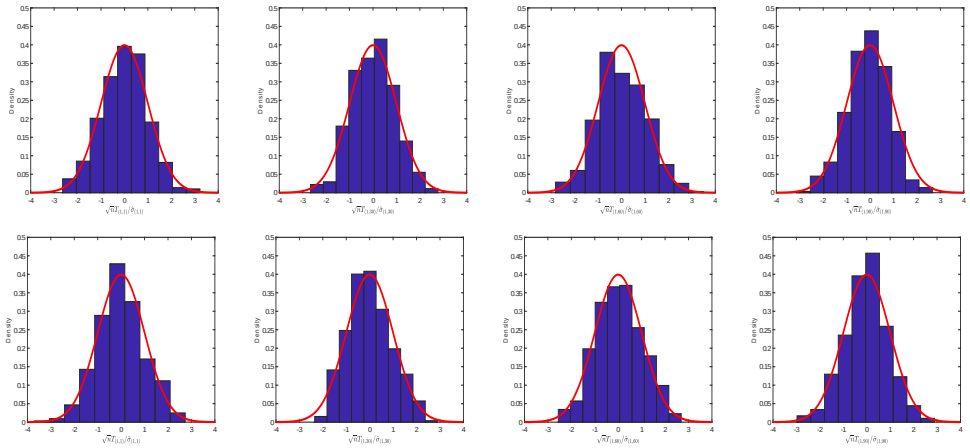


图1： $\tilde{\alpha} = 0.01$的 $\sqrt{n}T_{ij}^{-}/\hat{\sigma}_{ij}$ 直方图。这里， $T_{(i,j)} = T_{ij}$ 和 $\hat{\sigma}_{(i,j)} = \hat{\sigma}_{ij}$。第一行四个图的设置为$(p, n) = (100, 200)$，其中$(i, j) \in \{(1, 1), (1, 30), (1, 60), (1, 90)\}$。第二行四个图的样本量和维度设为$(p, n) = (100, 400)$。

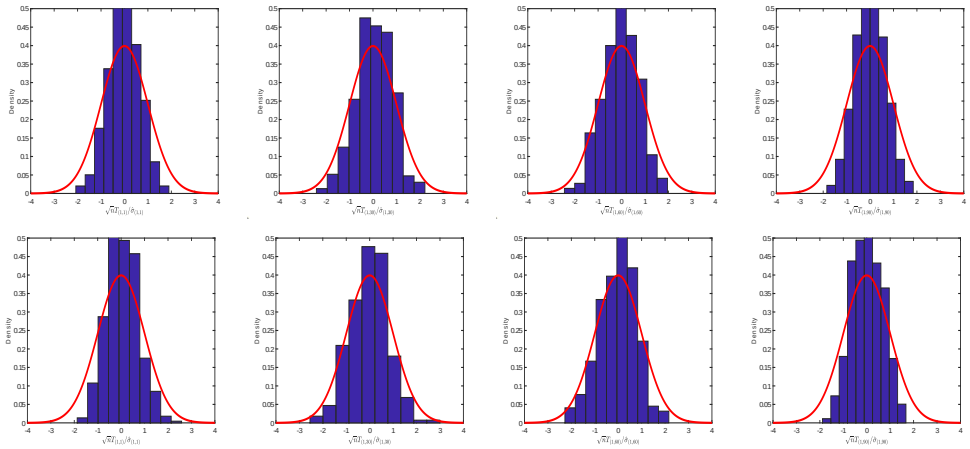


图2： $\tilde{\alpha} = 1$的 $\sqrt{n}T_{ij}/\hat{\sigma}_{ij}$ 直方图。这里， $T_{(i,j)} = T_{ij}$ 和 $\hat{\sigma}_{(i,j)} = \hat{\sigma}_{ij}$。第一行四个图的设置为$(p, n) = (100, 200)$，其中$(i, j) \in \{(1, 1), (1, 30), (1, 60), (1, 90)\}$。第二行四个图的样本量和维度设为$(p, n) = (100, 400)$。

我们模拟了极稀疏情形下的波动情况，如图1所示；稠密情形下的波动情况如图2所示。模拟中间歇性地选取索引（i，j）。事实上，中心极限定理提供了检验精度矩阵线性组合中任意元素的方法。从理论上讲，我们可以检验 $\mathbf{\Theta}_0$ 的任意索引（i，j）元素的真实值是否为零。

3.2. 平均覆盖概率

我们展示该检验方法在 $K = 2$ 情形下检验以下假设时的表现。

- **等值零假设**。 检验假设（18）；
- **线性零假设**。 检验线性零假设 H_0： $a_1\Theta_{0ij}^{[1]} + a_2\Theta_{0ij}^{[2]} = 0$，即 H_0： $\Theta_{0ij}^{[2]} = -\frac{a_1}{a_2}\Theta_{0ij}^{[1]}$。不失一般性，我们选择由（37）生成的 $-\frac{a_1}{a_2} = \underline{0.5}$ 和 $\Theta_{0ij}^{[1]}$。

从全局角度来看，我们采用了平均覆盖率，Jankov$\acute{a}$ 和 van deGeer [7]也考虑了这一指标。令

$$I_{ij} := \left[T_{ij} - 1.96\frac{\sigma_{ij}}{\sqrt{n}}, T_{ij} + 1.96\frac{\sigma_{ij}}{\sqrt{n}}\right] \tag{38}$$

为 $\mathbf{\Theta}_{0ij}$的 95%渐近置信区间，我们用估计量 $\hat{}\sigma_{ij}$ 替换σ_{ij} ，从而得到经验版本。置信区间 (38) 覆盖真实值的频率定义为 $\hat{}\,\vartheta_{ij}$。然后，集合 A上的平均覆盖率记为

$$Avgcov_A = \frac{1}{|A|}\sum_{(i,j)\in A}\hat{\vartheta}_{ij}. \tag{39}$$

S 表示 $\Theta_{0ij}^{[1]}$的非零元素集合。容易验证， $S = S_1 = S_2$ ，这是因为在等值零假设和线性零假设情形下， $\Theta_{0ij}^{[1]}$和 $\Theta_{0ij}^{[2]}$ 具有相同的稀疏结构。因此，对于不同的零假设，我们模拟了 S 及其补集 S^c 上的平均覆盖率。稀疏参数为 $\tilde{\alpha} = 0.1$,0.5, 和 0.9。

表1中的部分结果符合我们的预期。然而，我们不否认模拟结果会受到随机性的影响。此外，所提出的方法基于估计与假设检验的结合，因此误差会累积。模拟结果为实践提供了指导。

3.3. 多重*FGL*情形

对于多重FGL情形，我们考察了在检验以下假设时， $K = 3$ 情形下统计量 T_{ij} 的波动情况。

- **三样本线性零假设**。检验假设 H_0： $\Theta_{0ij}^{[3]} = -\frac{a_1}{a_3}\Theta_{0ij}^{[1]} - \frac{a_2}{a_3}\Theta_{0ij}^{[2]}$，其中$-\frac{a_1}{a_3} = 0.6$和 $-\frac{a_2}{a_3} = \underline{0.9}$均由 U(0，1) 生成。 $\Theta_{0ij}^{[1]}$和 $\Theta_{0ij}^{[2]}$分别由参数为 0.01 和 0.1 的 (37) 生成。

表1： $K = 2$ 情形下估计的平均覆盖概率。

$\tilde{\alpha}$	n	等值零假设		线性零假设	
		S	S^c	S	S^c
0.1	200	0.9886	0.9875	0.9101	0.9824
	400	0.9885	0.9867	0.8607	0.9762
0.5	200	0.9880	0.9878	0.9384	0.9745
	400	0.9870	0.9868	0.8820	0.9647
0.9	200	0.9901	0.9899	0.9509	0.9751
	400	0.9889	0.9890	0.9091	0.9639

我们将 $-\frac{a_1}{a_3}$和 $-\frac{a_2}{a_3}$设为正数，因为假设检验的设定应保证 $\{\Theta_{0ij}^{[k]}\}_{k=1}^{3}$ 为对称正定矩阵。此外，对于三样本线性零假设， S 表示 $\Theta_{0ij}^{[1]} + \frac{a_2}{a_3}\Theta_{0ij}^{[2]}$的非零元素集合。维度和样本量分别为 $(p, n) = (100, 200)$ 和 $(p, n) = (100, 400)$。所提出统计量 T_{ij} 在

$$(i, j) \in \{(1, 1), (1, 10), (1, 20), (1, 30)\}$$

精度矩阵的位置如图3所示。

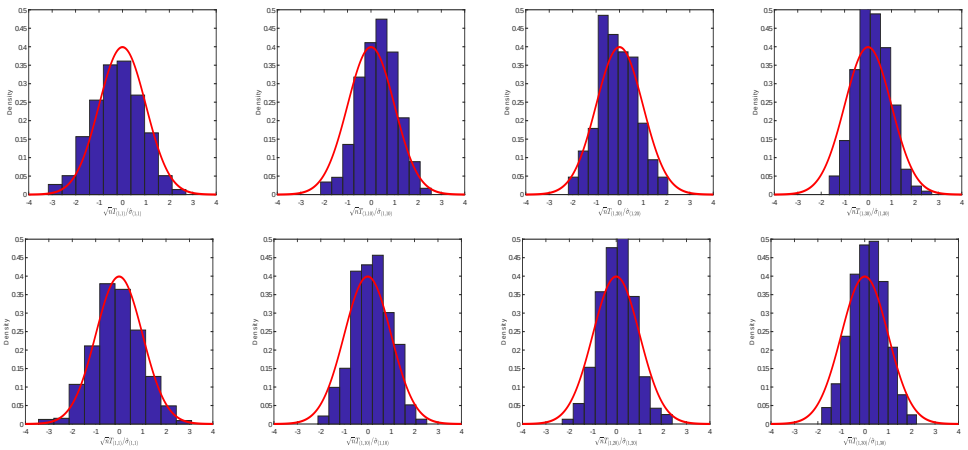


图3： $\sqrt{n}T_{ij}/\hat{\sigma}_{ij}$ 在 $\tilde{\alpha} \equiv 1$条件下的直方图。这里， $T_{(i,j)} = T_{ij}$和$\hat{\sigma}_{(i,j)} = \hat{\sigma}_{ij}$。设置为 $(p, n) = (100, 200)$，其中 $(i, j) \in \{(1, 1), (1, 10), (1, 20), (1, 30)\}$ 表示第一行中的四个图。第二行中四个图的样本量和维度设为 $(p, n) = (100, 400)$。

致谢

Q.Y. Zhang 部分获得了国家自然科学基金项目 12201430、11971097，以及首都经济贸易大学北京市属高校基本科研业务费项目 XRZ2021044 的资助。Z.D. Bai

获得了国家自然科学基金项目 12171198、12271536 和福建省科技发展基金项目 20210101147JC 的资助。H. Yang 获得了中国国家社会科学基金项目 22FGLB056 的资助。

参考文献

参考文献

[1] Boyd, S., Boyd, S. P., Vandenberghe, L., 2004. 凸优化。剑桥大学出版社。[2] Brownlees, C., Nualart, E., Sun, Y. C., 2018. 已实现网络。《应用计量经济学杂志》33 (7), 986–1006.[3] Danaher, P., Wang, P., Witten, D. M., 2014. 用于跨多个类别进行逆协方差估计的联合图形套索。皇家统计学会杂志：B辑（统计方法学）76 (2), 373–397.[4] Fan, J. Q., Feng, Y., Wu, Y. C., 2009. 通过自适应套索和SCAD惩罚进行网络探索。《应用统计年鉴》3 (2), 521–541.[5] Friedman, J., Hastie, T., Tibshirani, R., 2008. 使用图形套索进行稀疏逆协方差估计。生物统计学9 (3), 432–441.[6] Guo, J., Levina, E., Michailidis, G., Zhu, J., 2011. 多个图模型的联合估计。《生物统计学》98 (1), 1–15.[7] Jankov´a, J., van de Geer, S., 2015. 高维逆协方差估计的置信区间。《电子统计学期刊》9 (1), 1205–1229.[8] Jankov´a, J., van de Geer, S., 2017. 高维精度矩阵估计中的诚实置信区域与最优性。Test 26 (1), 143–162.[9] Jankov´a, J., van de Geer, S., 2018. 高维图模型中的推断。网址：http://arxiv.org/abs/arXiv:1801.08512[10] Johnstone, I. M., 2001. 主成分分析中最大特征值的分布。《统计学年鉴》29 (2), 295–327.[11] Ravikumar, P., Wainwright, M. J., Raskutti, G., Yu, B., 2011. 通过最小化 l_1-惩罚的对数行列式散度进行高维协方差估计。《电子统计学期刊》5, 935–980.[12] Ren, Z., Sun, T., Zhang, C.-H., Zhou, H. H., 2015. 大型高斯图模型估计中的渐近正态性与最优性。《统计学年鉴》43 (3), 991–1026。

[13] Rothman, A. J., Bickel, P. J., Levina, E., Zhu, J., 2008. 稀疏置换不变协方差估计。《电子统计学期刊》2, 494–515。

[14] Xue, L. Z.、Zou, H.，2012。高维非高斯图模型的正则化秩估计。《统计学年鉴》40（5），2541–2571。

[15] Yu, M.、Gupta, V.、Kolar, M.，2020。基于广义得分匹配的成对图模型同时推断。《机器学习研究期刊》21（91），1–51。

[16] Yuan, M.、Lin, Y.，2007。高斯图模型中的模型选择与估计。《生物统计学》94（1），19–35。

[17] Yuan, Y. P.、Shen, X. T.、Pan, W.、Wang, Z. Z.，2019。用于重构有向无环高斯图的约束似然。《生物统计学》106（1），109–125。

附录

A. 定理证明

A.1. 定理*1*的证明

To p为了证明定理1，我们需要使用 Jankov´a 和 Van de Geer [9], 的一个引理，该引理如下所示。

引理7。 令 $f(\Delta)\colon = tr(\Delta\Sigma_0) - [\log\det(\Delta+\Theta_0) - \log\det(\Theta_0)]$。假设 $1/L \le \lambda_{min}(\Theta_0) \le \lambda_{max}(\Theta_0) \le L$ 对某个常数 $L \ge 1$成立。那么对于所有满足 $||\Delta||_F \le 1/(2L)$ 的 Δ， $f(\Delta)$ 定义良好，并且

$$f(\Delta) \ge \frac{1}{2(L+1/(2L))^2}||\Delta||_F^2. \tag{40}$$

为简化记号，我们分别用 $\widehat{\Sigma}^{[k]}$、 $\Sigma_0^{[k]}$ 、 $\widehat{\Theta}^{[k]}$、 $\Theta_0^{[k]}$ 替换 $\widehat{\Sigma}_k$、 Σ_{0k}、 $\widehat{\Theta}_k$、 Θ_{0k}。

$\widehat{\ }$ $\widetilde{\ }$ $\widehat{\ }$**证明1**。 注意，Θ_k 是 $k = 1,2$的融合图形 *Lasso* 的最小值。令 $\Theta_k = \alpha_k\Theta_k + (1-\alpha_k)\Theta_{0k}$，且令 $\alpha_k = \frac{M}{M+||\widehat{\Theta}_k - \Theta_{0k}||_F}$。根据 $\widetilde{\Theta}_k$的定义以及损失函数的凸性

$F_n(\Theta_1, \Theta_2) = tr(\{v7 \quad \widehat{\Sigma}_1) - $对数行列式$(\Theta_1) + tr(\Theta_2\ \widehat{\Sigma}_2) - \log\det(\Theta_2) + \lambda||\Theta_1^-||_1 + \lambda||\Theta_2^-||_1 + \rho||\Theta_1^- - \Theta_2^-||_1,$

我们得到

$$F_n(\widetilde{\Theta}_1, \widetilde{\Theta}_2) \le F_n(\Theta_{01}, \Theta_{02}). \tag{41}$$

也就是说

$$\begin{aligned} &\sum_{k=1}^{2}\left\{tr(\widetilde{\Theta}_k - \Theta_{0k})\widehat{\Sigma}_k - \left(\log\det(\widetilde{\Theta}_k) - \log\det(\Theta_{0k})\right) + \lambda||\widetilde{\Theta}_k^-||_1\right\} + \rho||\widetilde{\Theta}_1^- - \widetilde{\Theta}_2^-||_1 \\ \le\ & \lambda||\Theta_{01}^-||_1 + \lambda||\Theta_{02}^-||_1 + \rho||\Theta_{01}^- - \Theta_{02}^-||_1. \end{aligned} \tag{42}$$

~令 $\Delta_k = \Theta_k - \Theta_{0k}$，并且

$$f(\Delta_k) := tr(\Delta_k \Sigma_{0k}) - \Big[\log\det(\Delta_k + \Theta_{0k}) - \log\det(\Theta_{0k})\Big],$$

减去 ^ ^在不等式（42）两边同时作用 $tr(\Delta_1(\Sigma_1 - \Sigma_{01})) + tr(\Delta_2(\Sigma_2 - \Sigma_{02}))$，得到 t

$$\begin{aligned} &f(\Delta_1) + f(\Delta_2) + \lambda||\widetilde{\Theta}_1^-||_1 + \lambda||\widetilde{\Theta}_2^-||_1 + \rho||\widetilde{\Theta}_1^- - \widetilde{\Theta}_2^-||_1 \\ \leq &- tr(\Delta_1(\widehat{\Sigma}_1 - \Sigma_{01})) - tr(\Delta_2(\widehat{\Sigma}_2 - \Sigma_{02})) + \lambda||\Theta_{01}^-||_1 + \lambda||\Theta_{02}^-||_1 + \rho||\Theta_{01}^- - \Theta_{02}^-||_1. \end{aligned} \tag{43}$$

^对于 $tr(\Delta_k(\Sigma_k - \Sigma_{0k}))$ 项，我们有

$$\begin{aligned} |tr(\Delta_k(\widehat{\Sigma}_k - \Sigma_{0k}))| &= |G(\Delta_k \circ (\widehat{\Sigma}_k - \Sigma_{0k}))| \\ &\leq |G(\Delta_k^- \circ (\widehat{\Sigma}_k^- - \Sigma_{0k}^-))| + |G(\Delta_k^+ \circ (\widehat{\Sigma}_k^+ - \Sigma_{0k}^+))|, \end{aligned} \tag{44}$$

其中，函数 $G(M)$ 对矩阵 M的所有元素求和，$\circ$是*Hadamard*乘积。^根据柯西–施瓦茨不等式，在集合 $\{$最大$_k\{||\Sigma_k - \Sigma_{0k}||_\infty\} \leq \lambda_0\}$上，

$$\begin{aligned} |G(\Delta_k^- \circ (\widehat{\Sigma}_k^- - \Sigma_{0k}^-))| + |G(\Delta_k^+ \circ (\widehat{\Sigma}_k^+ - \Sigma_{0k}^+))| &\leq ||\widehat{\Sigma}_k^- - \Sigma_{0k}^-||_\infty ||\Delta_k^-||_1 + ||\widehat{\Sigma}_k^+ - \Sigma_{0k}^+||_F ||\Delta_k^+||_F \\ &\leq \lambda_0||\Delta_k^-||_1 + ||\widehat{\Sigma}_k^+ - \Sigma_{0k}^+||_F ||\Delta_k^+||_F. \end{aligned} \tag{45}$$

因此，

$$\begin{aligned} -tr(\Delta_k(\widehat{\Sigma}_k - \Sigma_{0k})) &\leq |tr(\Delta_k(\widehat{\Sigma}_k - \Sigma_{0k}))| \\ &\leq \lambda_0||\Delta_k^-||_1 + ||\widehat{\Sigma}_k^+ - \Sigma_{0k}^+||_F ||\Delta_k^+||_F. \end{aligned} \tag{46}$$

接下来，对于满足条件 $L_k \geq 1$ 的

$$1/L_k \leq \lambda_{min}(\Theta_{0k}) \leq \lambda_{max}(\Theta_{0k}) \leq L_k, \tag{47}$$

~我们选择满足 $1/L \leq 1/L_k$ 和 $L_k \leq L, k = 1,2$ 的 $L > 1$。根据 Δ_k 和 Θ_k 的定义，

我们得到

$$||\Delta_k||_F = \alpha_k||\widehat{\Theta}_k - \Theta_{0k}||_F = \frac{||\widehat{\Theta}_k - \Theta_{0k}||_F}{M + ||\widehat{\Theta}_k - \Theta_{0k}||_F} M, \tag{48}$$

对于任意 $M \in (0, 1/2L]$。因此，$||\Delta_k||_F$ 由 M 界定，即 $||\Delta_k||_F \leq M$。对于 $f(\Delta_k)$ 项，根据引理*7*，我们有

$$f(\Delta_k) \geq c||\widetilde{\Theta}_k - \Theta_{0k}||_F^2, \tag{49}$$

w这里 $c = \frac{1}{2(L+1/(2L))^2}$。特别地，我们选择 $c = 1/(8L^2)$，且不等式 *(49)* 仍然成立 s.

利用界（*46*）和（*49*），不等式（*43*）变为

$$\begin{aligned} &c||\widetilde{\Theta}_1 - \Theta_{01}||_F^2 + c||\widetilde{\Theta}_2 - \Theta_{02}||_F^2 + \lambda||\widetilde{\Theta}_1^-||_1 + \lambda||\widetilde{\Theta}_2^-||_1 + \rho||\widetilde{\Theta}_1^- - \widetilde{\Theta}_2^-||_1 \\ \leq &\lambda_0||\Delta_1^-||_1 + \lambda_0||\Delta_2^-||_1 + ||\widehat{\Sigma}_1^+ - \Sigma_{01}^+||_F ||\Delta_1^+||_F + ||\widehat{\Sigma}_2^+ - \Sigma_{02}^+||_F ||\Delta_2^+||_F \\ &+ \lambda||\Theta_{01}^-||_1 + \lambda||\Theta_{02}^-||_1 + \rho||\Theta_{01}^- - \Theta_{02}^-||_1. \end{aligned} \tag{50}$$

将不等式（*50*）中的若干项移项并合并，得到以下不等式

$$\begin{aligned}
&c||\widetilde{\Theta}_1-\Theta_{01}||_F^2+c||\widetilde{\Theta}_2-\Theta_{02}||_F^2\\
&+\lambda\Big\{||\widetilde{\Theta}_1^-||_1-||\Theta_{01}^-||_1+||\widetilde{\Theta}_2^-||_1-||\Theta_{02}^-||_1\Big\}\\
\leq&\lambda_0\Big\{||\widetilde{\Theta}_1^--\Theta_{01}^-||_1+||\widetilde{\Theta}_2^--\Theta_{02}^-||_1\Big\}+\rho\Big\{||\Theta_{01}^--\Theta_{02}^-||_1-||\widetilde{\Theta}_1^--\widetilde{\Theta}_2^-||_1\Big\}\\
&+||\widehat{\Sigma}_1^+-\Sigma_{01}^+||_F||\widetilde{\Theta}_1^+-\Theta_{01}^+||_F+||\widehat{\Sigma}_2^+-\Sigma_{02}^+||_F||\widetilde{\Theta}_2^+-\Theta_{02}^+||_F.
\end{aligned}\tag{51}$$

接下来需要证明三个不等式：

$$||\widetilde{\Theta}_k^-||_1-||\Theta_{0k}^-||_1\geq||\Delta_{kS_k^c}^-||_1-||\Delta_{kS_k}^-||_1,\tag{52}$$

$$||\widetilde{\Theta}_k^--\Theta_{0k}^-||_1\leq||\Delta_{kS_k^c}^-||_1+||\Delta_{kS_k}^-||_1,\tag{53}$$

$$||\Theta_{01}^--\Theta_{02}^-||_1-||\widetilde{\Theta}_1^--\widetilde{\Theta}_2^-||_1\leq||\widetilde{\Theta}_1^--\Theta_{01}^-||_1+||\widetilde{\Theta}_2^--\Theta_{02}^-||_1.\tag{54}$$

因为

$$\begin{aligned}
||\widetilde{\Theta}_k^-||_1=&||\Theta_{0k}^-+\Delta_k^-||_1\\
=&||\Theta_{0kS_k}^-+\Delta_{kS_k}^-||_1+||\Delta_{kS_k^c}^-||_1,
\end{aligned}\tag{55}$$

and

$$||\Theta_{0k}^-||_1=||\Theta_{0kS_k}^-||_1\tag{56}$$

成立。因此，

$$\begin{aligned}
||\widetilde{\Theta}_k^-||_1-||\Theta_{0k}^-||_1=&||\Theta_{0kS_k}^-+\Delta_{kS_k}^-||_1+||\Delta_{kS_k^c}^-||_1-||\Theta_{0kS_k}^-||_1\\
\geq&||\Delta_{kS_k^c}^-||_1-\Big|||\Theta_{0kS_k}^-+\Delta_{kS_k}^-||_1-||\Theta_{0kS_k}^-||_1\Big|\\
\geq&||\Delta_{kS_k^c}^-||_1-||\Delta_{kS_k}^-||_1,
\end{aligned}\tag{57}$$

这就证明了不等式（*52*）。根据三角不等式，我们自然得到

$$\begin{aligned}
||\widetilde{\Theta}_k^--\Theta_{0k}^-||_1=&||\Delta_k^-||_1\\
=&||\Delta_{kS_k^c}^-+\Delta_{kS_k}^-||_1\\
\leq&||\Delta_{kS_k^c}^-||_1+||\Delta_{kS_k}^-||_1.
\end{aligned}\tag{58}$$

因此，不等式（*53*）成立。对于不等式（*54*），我们有

$$\begin{aligned}
||\Theta_{01}^--\Theta_{02}^-||_1-||\widetilde{\Theta}_1^--\widetilde{\Theta}_2^-||_1=&||\Theta_{01}^--\widetilde{\Theta}_1^-+\widetilde{\Theta}_1^--\widetilde{\Theta}_2^-+\widetilde{\Theta}_2^--\Theta_{02}^-||_1-||\widetilde{\Theta}_1^--\widetilde{\Theta}_2^-||_1\\
\leq&||\widetilde{\Theta}_1^--\Theta_{01}^-||_1+||\widetilde{\Theta}_2^--\Theta_{02}^-||_1.
\end{aligned}\tag{59}$$

因此，不等式 *(51)* 得出

$$\begin{aligned}
&c||\widetilde{\Theta}_1-\Theta_{01}||_F^2+c||\widetilde{\Theta}_2-\Theta_{02}||_F^2\\
&+\lambda\Big\{||\Delta_{1S_1^c}^-||_1-||\Delta_{1S_1}^-||_1+||\Delta_{2S_2^c}^-||_1-||\Delta_{2S_2}^-||_1\Big\}\\
\leq&(\rho+\lambda_0)\Big\{||\Delta_{1S_1^c}^-||_1+||\Delta_{1S_1}^-||_1+||\Delta_{2S_2^c}^-||_1+||\Delta_{2S_2}^-||_1\Big\}\\
&+||\widehat{\Sigma}_1^+-\Sigma_{01}^+||_F||\widetilde{\Theta}_1^+-\Theta_{01}^+||_F+||\widehat{\Sigma}_2^+-\Sigma_{02}^+||_F||\widetilde{\Theta}_2^+-\Theta_{02}^+||_F.
\end{aligned}\tag{60}$$

取 $2(\rho+\lambda_0)<\lambda$，我们得出

$$\begin{aligned}
&2c\Big\{||\widetilde{\Theta}_1-\Theta_{01}||_F^2+||\widetilde{\Theta}_2-\Theta_{02}||_F^2\Big\}+\lambda\Big\{||\Delta^-_{1S_1^c}||_1+||\Delta^-_{2S_2^c}||_1\Big\}\\
\leq&3\lambda\Big\{||\Delta^-_{1S_1}||_1+||\Delta^-_{2S_2}||_1\Big\}\\
&+2\Big\{||\widehat{\Sigma}_1^+-\Sigma_{01}^+||_F||\widetilde{\Theta}_1^+-\Theta_{01}^+||_F+||\widehat{\Sigma}_2^+-\Sigma_{02}^+||_F||\widetilde{\Theta}_2^+-\Theta_{02}^+||_F\Big\}.
\end{aligned}\tag{61}$$

根据 Δ_k的定义，我们有

$$\begin{aligned}
||\Delta_k^-||_1=&||\Delta^-_{kS_k}+\Delta^-_{kS_k^c}||_1\\
\leq&||\Delta^-_{kS_k}||_1+||\Delta^-_{kS_k^c}||_1.
\end{aligned}\tag{62}$$

因此我们推得

$$\begin{aligned}
&2c\Big\{||\widetilde{\Theta}_1-\Theta_{01}||_F^2+||\widetilde{\Theta}_2-\Theta_{02}||_F^2\Big\}+\lambda\Big\{||\Delta_1^-||_1+||\Delta_2^-||_1\Big\}\\
\leq&4\lambda\Big\{||\Delta^-_{1S_1}||_1+||\Delta^-_{2S_2}||_1\Big\}\\
&+2\Big\{||\widehat{\Sigma}_1^+-\Sigma_{01}^+||_F||\widetilde{\Theta}_1^+-\Theta_{01}^+||_F+||\widehat{\Sigma}_2^+-\Sigma_{02}^+||_F||\widetilde{\Theta}_2^+-\Theta_{02}^+||_F\Big\}
\end{aligned}\tag{63}$$

成立。由于算术-几何平均不等式，不等式 $||\Delta^-_{kS_k}||_1\leq\sqrt{s_k}||\Delta^-_{kS_k}||_F$成立。因此 —

$$\begin{aligned}
&2c\Big\{||\widetilde{\Theta}_1-\Theta_{01}||_F^2+||\widetilde{\Theta}_2-\Theta_{02}||_F^2\Big\}+\lambda\Big\{||\Delta_1^-||_1+||\Delta_2^-||_1\Big\}\\
\leq&4\lambda\Big\{\sqrt{s_1}||\Delta^-_{1S_1}||_F+\sqrt{s_2}||\Delta^-_{2S_2}||_F\Big\}\\
&+2\Big\{||\widehat{\Sigma}_1^+-\Sigma_{01}^+||_F||\widetilde{\Theta}_1^+-\Theta_{01}^+||_F+||\widehat{\Sigma}_2^+-\Sigma_{02}^+||_F||\widetilde{\Theta}_2^+-\Theta_{02}^+||_F\Big\}.
\end{aligned}\tag{64}$$

利用 $xy\leq\ (x^2+y^2)\ /2$，由不等式（*64*）可得

$$\begin{aligned}
&2c\Big\{||\widetilde{\Theta}_1-\Theta_{01}||_F^2+||\widetilde{\Theta}_2-\Theta_{02}||_F^2\Big\}+\lambda\Big\{||\Delta_1^-||_1+||\Delta_2^-||_1\Big\}\\
\leq&\frac{1}{2}\Big(c||\Delta^-_{1S_1}||_F^2+\frac{16\lambda^2s_1}{c}+c||\Delta^-_{2S_2}||_F^2+\frac{16\lambda^2s_2}{c}\Big)\\
&+\frac{1}{2}\Big(c||\widetilde{\Theta}_1^+-\Theta_{01}^+||_F^2+\frac{4||\widehat{\Sigma}_1^+-\Sigma_{01}^+||_F^2}{c}+c||\widetilde{\Theta}_2^+-\Theta_{02}^+||_F^2+\frac{4||\widehat{\Sigma}_2^+-\Sigma_{02}^+||_F^2}{c}\Big).
\end{aligned}\tag{65}$$

因为

$$\begin{aligned}
c||\widetilde{\Theta}_k^+-\Theta_{0k}^+||_F^2+c||\Delta^-_{kS_k}||_F^2\leq&\Big\{c||\widetilde{\Theta}_k^+-\Theta_{0k}^+||_F^2+c||\Delta_k^-||_F^2\Big\}\\
&+\Big\{c||\Delta^-_{kS_k}||_F^2+c||\Delta^-_{kS_k^c}||_F^2+c||\Delta_k^+||_F^2\Big\}\\
=&2c||\Delta_k||_F^2,
\end{aligned}\tag{66}$$

我们得到

$$\begin{aligned}
&2c\Big\{||\widetilde{\Theta}_1-\Theta_{01}||_F^2+||\widetilde{\Theta}_2-\Theta_{02}||_F^2\Big\}+\lambda\Big\{||\Delta_1^-||_1+||\Delta_2^-||_1\Big\}\\
\leq&c\Big\{||\Delta_1||_F^2+||\Delta_2||_F^2\Big\}+\frac{8\lambda^2(s_1+s_2)}{c}+\frac{2||\widehat{\Sigma}_1^+-\Sigma_{01}^+||_F^2}{c}+\frac{2||\widehat{\Sigma}_2^+-\Sigma_{02}^+||_F^2}{c}.
\end{aligned}\tag{67}$$

因此，

$$\begin{aligned} &c\Big\{||\Delta_1||_F^2+||\Delta_2||_F^2\Big\}+\lambda\Big\{||\Delta_1^-||_1+||\Delta_2^-||_1\Big\} \\ \leq&\frac{8\lambda^2(s_1+s_2)}{c}+\frac{2||\widehat{\Sigma}_1^+-\Sigma_{01}^+||_F^2}{c}+\frac{2||\widehat{\Sigma}_2^+-\Sigma_{02}^+||_F^2}{c}. \end{aligned} \tag{68}$$

根据不等式$||\widehat{\Sigma}_k^+-\Sigma_{0k}^+||_F\leq\sqrt{p}||\widehat{\Sigma}_k^+-\Sigma_{0k}^+||_\infty$，我们有

$$c\Big\{||\Delta_1||_F^2+||\Delta_2||_F^2\Big\}+\lambda\Big\{||\Delta_1^-||_1+||\Delta_2^-||_1\Big\}\leq\frac{8\lambda^2(s_1+s_2)}{c}+\frac{4p\lambda_0^2}{c}. \tag{69}$$

接下来，我们证明，将 Θ_k 替换为 Θ_k后，结论仍然成立。根据条件 n,

$$||\Delta_1||_F^2+||\Delta_2||_F^2\leq\frac{\lambda_0}{2cL}\leq\frac{\lambda}{4cL}\leq\frac{1}{32L^2}. \tag{70}$$

取 $M=1/(2\sqrt{2L})^- <1/2L$，我们有

$$||\Delta_1||_F^2+||\Delta_2||_F^2\leq M^2/4. \tag{71}$$

因此，$||\Delta_k||_F$ 由 $M/2$界定。此外，

$$||\widehat{\Theta}_k-\Theta_{0k}||_F=\frac{M||\Delta_k||_F}{M-||\Delta_k||_F}, \tag{72}$$

这意味着 $||\Theta_k-\Theta_{0k}||_F$是集合 $(0,\ M)$ 上的 $||\Delta_k||_F$的单调递增函数。我们得到 $||\Theta_k-\Theta_{0k}||_F\leq M$。因此，我们可以将 Θ_k 替换为 Θ_k，从而得到不等式 *(69)* 对 Θ_k成立。

根据不等式（*69*），我们得到

$$\begin{aligned} ||\widehat{\Theta}_k-\Theta_{0k}||_F^2\leq&\frac{8\lambda^2(s_1+s_2)}{c^2}+\frac{4p\lambda_0^2}{c^2} \\ \leq&\frac{\lambda^2(8s_1+8s_2+p)}{c^2}, \end{aligned} \tag{73}$$

and

$$\begin{aligned} ||\widehat{\Theta}_k^--\Theta_{0k}^-||_1\leq&\frac{8\lambda(s_1+s_2)}{c}+\frac{4p\lambda_0^2}{\lambda c} \\ \leq&\frac{\lambda(8s_1+8s_2+p)}{c}. \end{aligned} \tag{74}$$

因此，我们得出 $\sum_{k=1}^2|||\widehat{\Theta}_k-\Theta_{0k}|||_1$ 的上界，

$$\begin{aligned} \sum_{k=1}^2|||\widehat{\Theta}_k-\Theta_{0k}|||_1\leq&\sum_{k=1}^2\Big(||\widehat{\Theta}_k^+-\Theta_{0k}^+||_\infty+||\widehat{\Theta}_k^--\Theta_{0k}^-||_1\Big) \\ \leq&\sum_{k=1}^2\Big(||\widehat{\Theta}_k-\Theta_{0k}||_F+||\widehat{\Theta}_k^--\Theta_{0k}^-||_1\Big) \\ \leq&\frac{2\lambda\sqrt{8s_1+8s_2+p}}{c}+\frac{2\lambda(8s_1+8s_2+p)}{c} \\ \leq&\frac{4\lambda(8s_1+8s_2+p)}{c}. \end{aligned} \tag{75}$$

A.2. 定理 *2* 的证明

证明 2. 满足不等式 *(68)* 的极小化解 $(\widehat{\Theta}_R^{[1]}, \widehat{\Theta}_R^{[2]})$，即

$$\begin{aligned} &c\Big\{||\widehat{\Theta}_R^{[1]} - \Theta_{R0}^{[1]}||_F^2 + ||\widehat{\Theta}_R^{[2]} - \Theta_{R0}^{[2]}||_F^2\Big\} + \lambda\Big\{||(\widehat{\Theta}_R^{[1]} - \Theta_{R0}^{[1]})^-||_1 + ||(\widehat{\Theta}_R^{[2]} - \Theta_{R0}^{[2]})^-||_1\Big\} \\ \leq &\frac{8\lambda^2(s_1+s_2)}{c} + \frac{2||(\widehat{R}^{[1]} - R_0^{[1]})^+||_F^2}{c} + \frac{2||(\widehat{R}^{[2]} - R_0^{[2]})^+||_F^2}{c}. \end{aligned} \tag{76}$$

$\widehat{R}^{[k]}$ 和 $R_0^{[k]}$ 的对角元素均为1。因此

$$\begin{aligned} &c\Big\{||\widehat{\Theta}_R^{[1]} - \Theta_{R0}^{[1]}||_F^2 + ||\widehat{\Theta}_R^{[2]} - \Theta_{R0}^{[2]}||_F^2\Big\} + \lambda\Big\{||(\widehat{\Theta}_R^{[1]} - \Theta_{R0}^{[1]})^-||_1 + ||(\widehat{\Theta}_R^{[2]} - \Theta_{R0}^{[2]})^-||_1\Big\} \\ \leq &\frac{8\lambda^2(s_1+s_2)}{c}. \end{aligned} \tag{77}$$

此外，对于 l_1-算子范数的结论，我们得到

$$\begin{aligned} |||\widehat{\Theta}_R^{[1]} - \Theta_{R0}^{[1]}|||_1 + |||\widehat{\Theta}_R^{[2]} - \Theta_{R0}^{[2]}|||_1 \leq &\sum_{k=1}^{2}\Big(||(\widehat{\Theta}_R^{[k]} - \Theta_{R0}^{[k]})^+||_\infty + ||(\widehat{\Theta}_R^{[k]} - \Theta_{R0}^{[k]})^-||_1\Big) \\ \leq &\sum_{k=1}^{2}\Big(||\widehat{\Theta}_R^{[k]} - \Theta_{R0}^{[k]}||_F + ||(\widehat{\Theta}_R^{[k]} - \Theta_{R0}^{[k]})^-||_1\Big) \\ \leq &\frac{32\lambda(s_1+s_2)}{c}. \end{aligned} \tag{78}$$

对于极小值点$(\widehat{\Theta}_w^{[1]}, \widehat{\Theta}_w^{[2]})$，有如下不等式

$$\begin{aligned} |||\widehat{\Theta}_R^{[k]} - \Theta_{R0}^{[k]}|||_1 = &|||\widehat{W}^{[k]}\widehat{\Theta}_w^{[k]}\widehat{W}^{[k]} - W_0^{[k]}\Theta_{w0}^{[k]}W_0^{[k]}|||_1 \\ \leq &||\widehat{W}^{[k]}||_\infty^2|||\widehat{\Theta}_w^{[k]} - \Theta_{w0}^{[k]}|||_1 + ||\widehat{W}^{[k]} - W_0^{[k]}||_\infty|||\Theta_{w0}^{[k]}|||_1||\widehat{W}^{[k]}||_\infty \\ &+ ||W_0^{[k]}||_\infty|||\Theta_{w0}^{[k]}|||_1||\widehat{W}^{[k]} - W_0^{[k]}||_\infty. \end{aligned} \tag{79}$$

为了得出结论，我们有以下事实：

- *The Sub-Gaussian vector with covariance* $\Sigma_0^{[k]}$ *implies that* $\sqrt{n/\log p}||(\widehat{\Sigma}^{[k]} - \Sigma_0^{[k]})||_\infty$ *is bounded in probability.*
- *The eigenvalues of* $\Theta_{w0}^{[k]}$ *are bounded by a constant.*

因此，$|||\widehat{\Theta}_R^{[k]} - \Theta_{R0}^{[k]}|||_1$和 $|||\widehat{\Theta}_w^{[k]} - \Theta_{w0}^{[k]}|||_1$具有相同的边界。

A.3. 定理 *3* 的证明

证明 3. 同理，Θ_k是针对 $k = 1、2、\cdots、K$ 的融合图形 *Lasso* 最小值。令 $\widetilde{\Theta}_k = \alpha_k\widehat{\Theta}_k + (1-\alpha_k)\Theta_{0k}$，且 $\alpha_k = \frac{M}{M+||\widehat{\Theta}_k - \Theta_{0k}||_F}$。表示

$$F_n(\Theta_1, \cdots, \Theta_K) = \sum_{k=1}^{K}\Big\{tr(\Theta_k\widehat{\Sigma}_k) - \log\det(\Theta_k)\Big\} + \lambda\sum_{k=1}^{K}||\Theta_k^-||_1 + \rho\sum_{k<k'}||\Theta_k^- - \Theta_{k'}^-||_1,$$

我们得到

$$F_n(\widetilde{\Theta}_1, \widetilde{\Theta}_2, \cdots, \widetilde{\Theta}_K) \leq F_n(\Theta_{01}, \Theta_{02}, \cdots, \Theta_{0K}). \tag{80}$$

因此，

$$\begin{aligned}&\sum_{k=1}^{K}\Big\{tr(\widetilde{\Theta}_k-\Theta_{0k})\widehat{\Sigma}_k-\Big(\log\det(\widetilde{\Theta}_k)-\log\det(\Theta_{0k})\Big)+\lambda||\widetilde{\Theta}_k^-||_1\Big\}+\rho\sum_{k<k'}||\widetilde{\Theta}_k^--\widetilde{\Theta}_{k'}^-||_1\\ \leq\ &\lambda\sum_{k=1}^{K}||\Theta_{0k}^-||_1+\rho\sum_{k<k'}||\Theta_{0k}^--\Theta_{0k'}^-||_1.\end{aligned}\tag{81}$$

~采用记号 $\Delta_k=\Theta_k-\Theta_{0k}$ 和

$$f(\Delta_k):=tr(\Delta_k\Sigma_{0k})-\Big[\log\det(\Delta_k+\Theta_{0k})-\log\det(\Theta_{0k})\Big]$$

可得以下表达式

$$\begin{aligned}&\sum_{k=1}^{K}f(\Delta_k)+\lambda\sum_{k=1}^{K}||\widetilde{\Theta}_k^-||_1+\rho\sum_{k<k'}||\widetilde{\Theta}_k^--\widetilde{\Theta}_{k'}^-||_1\\ \leq&-\sum_{k=1}^{K}\Big(tr(\Delta_k(\widehat{\Sigma}_k-\Sigma_{0k}))\Big)-tr(\Delta_2(\widehat{\Sigma}_2-\Sigma_{02}))+\lambda\sum_{k=1}^{K}||\Theta_{0k}^-||_1+\rho\sum_{k<k'}||\Theta_{0k}^--\Theta_{0k'}^-||_1.\end{aligned}\tag{82}$$

对于 $L_k\geq 1$、$k=1$、2、$\cdots$、K，Θ_{0k} 的最小特征值和最大特征值满足

$$1/L_k\leq\lambda_{min}(\Theta_{0k})\leq\lambda_{max}(\Theta_{0k})\leq L_k.\tag{83}$$

对于多种情形，我们选取一个常数 L，使其满足 $1/L\leq 1/L_k$ 和 $L_k\leq L$。通过类似的分析，对于 $(0、\ 1/2L]$ 中的 M，不等式 *(48)* 和不等式 *(49)* 仍然成立。

对于 K 组数据，根据不等式 *(46)* 和 *(49)*，不等式 *(82)* 可化为

$$\begin{aligned}&c\sum_{k=1}^{K}||\widetilde{\Theta}_k-\Theta_{0k}||_F^2+\lambda\sum_{k=1}^{K}||\widetilde{\Theta}_k^-||_1+\rho\sum_{k<k'}||\widetilde{\Theta}_k^--\widetilde{\Theta}_{k'}^-||_1\\ \leq&\sum_{k=1}^{K}\Big\{\lambda_0||\Delta_k^-||_1+||\widehat{\Sigma}_k^+-\Sigma_{0k}^+||_F||\Delta_k^+||_F\Big\}+\lambda\sum_{k=1}^{K}||\Theta_{0k}^-||_1+\rho\sum_{k<k'}||\Theta_{0k}^--\Theta_{0k'}^-||_1.\end{aligned}\tag{84}$$

因此，

$$\begin{aligned}&c\sum_{k=1}^{K}||\widetilde{\Theta}_k-\Theta_{0k}||_F^2+\lambda\sum_{k=1}^{K}\Big\{||\widetilde{\Theta}_k^-||_1-||\Theta_{0k}^-||_1\Big\}\\ \leq&\lambda_0\sum_{k=1}^{K}||\widetilde{\Theta}_k^--\Theta_{0k}^-||_1+\rho\sum_{k<k'}\Big\{||\Theta_{0k}^--\Theta_{0k'}^-||_1-||\widetilde{\Theta}_k^--\widetilde{\Theta}_{k'}^-||_1\Big\}\\ &+\sum_{k=1}^{K}\Big\{||\widehat{\Sigma}_k^+-\Sigma_{0k}^+||_F||\widetilde{\Theta}_k^+-\Theta_{0k}^+||_F\Big\}.\end{aligned}\tag{85}$$

当 $k=1$、2、$\cdots$、K 时，不等式 *(52)* 和 *(53)* 仍然成立。同理，我们有以下不等式

$$\begin{aligned}||\Theta_{0k}^--\Theta_{0k'}^-||_1-||\widetilde{\Theta}_k^--\widetilde{\Theta}_{k'}^-||_1=&||\Theta_{0k}^--\widetilde{\Theta}_k^-+\widetilde{\Theta}_k^--\widetilde{\Theta}_{k'}^-+\widetilde{\Theta}_{k'}^--\Theta_{0k'}^-||_1-||\widetilde{\Theta}_k^--\widetilde{\Theta}_{k'}^-||_1\\ \leq&||\widetilde{\Theta}_k^--\Theta_{0k}^-||_1+||\widetilde{\Theta}_{k'}^--\Theta_{0k'}^-||_1.\end{aligned}\tag{86}$$

因此，由等式 *(52)*、*(53)* 和 *(86)*，不等式 *(85)* 得出

$$\begin{aligned}
&c\sum_{k=1}^{K}||\widetilde{\Theta}_k-\Theta_{0k}||_F^2+\lambda\sum_{k=1}^{K}\Big\{||\Delta^-_{kS_k^c}||_1-||\Delta^-_{kS_k}||_1\Big\}\\
\leq&\lambda_0\sum_{k=1}^{K}\Big\{||\Delta^-_{kS_k^c}||_1+||\Delta^-_{kS_k}||_1\Big\}+\rho\sum_{k<k'}\Big\{||\Delta^-_{kS_k^c}||_1+||\Delta^-_{kS_k}||_1+||\Delta^-_{k'S_{k'}^c}||_1+||\Delta^-_{k'S_{k'}}||_1\Big\}\\
&+\sum_{k=1}^{K}\Big\{||\widehat{\Sigma}_k^+-\Sigma_{0k}^+||_F||\widetilde{\Theta}_k^+-\Theta_{0k}^+||_F\Big\}\\
\leq&\left(\frac{K(K-1)}{2}\rho+\lambda_0\right)\sum_{k=1}^{K}\Big\{||\Delta^-_{kS_k^c}||_1+||\Delta^-_{kS_k}||_1\Big\}+\sum_{k=1}^{K}\Big\{||\widehat{\Sigma}_k^+-\Sigma_{0k}^+||_F||\widetilde{\Theta}_k^+-\Theta_{0k}^+||_F\Big\}.
\end{aligned}\tag{87}$$

由于 K 是固定常数，且 $2(\frac{K(K-1)}{2}\rho+\lambda_0)<\lambda$，我们可以得到

$$\begin{aligned}
&2c\sum_{k=1}^{K}||\widetilde{\Theta}_k-\Theta_{0k}||_F^2+\lambda\sum_{k=1}^{K}||\Delta^-_{kS_k^c}||_1\\
\leq&3\lambda\sum_{k=1}^{K}||\Delta^-_{kS_k}||_1+2\sum_{k=1}^{K}\Big\{||\widehat{\Sigma}_k^+-\Sigma_{0k}^+||_F||\widetilde{\Theta}_k^+-\Theta_{0k}^+||_F\Big\}.
\end{aligned}\tag{88}$$

根据不等式 *(62)*，我们推得

$$\begin{aligned}
&2c\sum_{k=1}^{K}||\widetilde{\Theta}_k-\Theta_{0k}||_F^2+\lambda\sum_{k=1}^{K}||\Delta^-_{k}||_1\\
\leq&4\lambda\sum_{k=1}^{K}||\Delta^-_{kS_k}||_1+2\sum_{k=1}^{K}\Big\{||\widehat{\Sigma}_k^+-\Sigma_{0k}^+||_F||\widetilde{\Theta}_k^+-\Theta_{0k}^+||_F\Big\}
\end{aligned}\tag{89}$$

成立。此外，可得不等式 $||\Delta^-_{kS_k}||_1\leq\sqrt{s_k}||\Delta^-_{kS_k}||_F$。因此

$$\begin{aligned}
&2c\sum_{k=1}^{K}||\widetilde{\Theta}_k-\Theta_{0k}||_F^2+\lambda\sum_{k=1}^{K}||\Delta^-_{k}||_1\\
\leq&4\lambda\sum_{k=1}^{K}\left(\sqrt{s_k}||\Delta^-_{kS_k}||_F\right)+2\sum_{k=1}^{K}\Big\{||\widehat{\Sigma}_k^+-\Sigma_{0k}^+||_F||\widetilde{\Theta}_k^+-\Theta_{0k}^+||_F\Big\}.
\end{aligned}\tag{90}$$

根据 $xy\leq(x^2+y^2)/2$ 和不等式 *(66)*，不等式 *(90)* 推得

$$\begin{aligned}
&2c\sum_{k=1}^{K}||\widetilde{\Theta}_k-\Theta_{0k}||_F^2+\lambda\sum_{k=1}^{K}||\Delta^-_{k}||_1\\
\leq&\frac{1}{2}\sum_{k=1}^{K}\left(c||\Delta^-_{kS_k}||_F^2+\frac{16\lambda^2 s_k}{c}\right)+\frac{1}{2}\sum_{k=1}^{K}\left(c||\widetilde{\Theta}_k^+-\Theta_{0k}^+||_F^2+\frac{4||\widehat{\Sigma}_k^+-\Sigma_{0k}^+||_F^2}{c}\right)\\
\leq&c\sum_{k=1}^{K}||\Delta_k||_F^2+\frac{8\lambda^2\sum_{k=1}^{K}s_k}{c}+\frac{2\sum_{k=1}^{K}||\widehat{\Sigma}_k^+-\Sigma_{0k}^+||_F^2}{c}.
\end{aligned}\tag{91}$$

因此，

$$c\sum_{k=1}^{K}||\Delta_k||_F^2+\lambda\sum_{k=1}^{K}||\Delta^-_{k}||_1\leq\frac{8\lambda^2\sum_{k=1}^{K}s_k}{c}+\frac{2\sum_{k=1}^{K}||\widehat{\Sigma}_k^+-\Sigma_{0k}^+||_F^2}{c}.\tag{92}$$

利用弗罗贝尼乌斯范数与上确界范数之间的关系，我们有

$$c\sum_{k=1}^{K}||\Delta_k||_F^2+\lambda\sum_{k=1}^{K}||\Delta_k^-||_1\leq\frac{8\lambda^2\sum_{k=1}^{K}s_k}{c}+\frac{2Kp\lambda_0^2}{c}. \tag{93}$$

根据不等式 *(93)*，我们得到

$$\sum_{k=1}^{K}||\Delta_k||_F^2\leq\frac{\lambda_0}{2cL}. \tag{94}$$

根据 $\lambda_0\leq\lambda/2$ 和条件 $\lambda\leq c/8L$，我们得到

$$\sum_{k=1}^{K}||\Delta_k||_F^2\leq\frac{1}{32L^2}. \tag{95}$$

取 $M=1/(2\sqrt{2}L)\bar{}<1/2L$，我们有

$$\sum_{k=1}^{K}||\Delta_k||_F^2\leq M^2/4. \tag{96}$$

$\hat{}$因此，$||\Delta_k||_F$ 有界于 $M/2$。此外，我们可以推导出 $||\Theta_k-\Theta_{0k}||_F\leq M$，这意味着我们可以用 Θ_k替换 Θ_k，从而不等式 *(93)* 对 Θ_k成立，即

$$c\sum_{k=1}^{K}||\widehat{\Theta}_k-\Theta_{0k}||_F^2+\lambda\sum_{k=1}^{K}||(\widehat{\Theta}_k-\Theta_{0k})^-||_1\leq\frac{8\lambda^2\sum_{k=1}^{K}s_k}{c}+\frac{2Kp\lambda_0^2}{c}, \tag{97}$$

这意味着

$$\begin{aligned}\sum_{k=1}^{K}|||\widehat{\Theta}_k-\Theta_{0k}|||_1\leq&\sum_{k=1}^{K}\left(||\widehat{\Theta}_k^+-\Theta_{0k}^+||_\infty+||\widehat{\Theta}_k^--\Theta_{0k}^-||_1\right)\\\leq&\sum_{k=1}^{K}\left(||\widehat{\Theta}_k-\Theta_{0k}||_F+||\widehat{\Theta}_k^--\Theta_{0k}^-||_1\right)\\\leq&K\left[\frac{\lambda\sqrt{8\sum_{k=1}^{K}s_k+\frac{Kp}{2}}}{c}+\frac{\lambda\left(8\sum_{k=1}^{K}s_k+\frac{Kp}{2}\right)}{c}\right]\\\leq&\frac{2K\lambda\left(8\sum_{k=1}^{K}s_k+\frac{Kp}{2}\right)}{c},\end{aligned} \tag{98}$$

由此完成证明。

A.4. 定理 *4* 的证明

证明 4. 由 *(92)* 可得

$$c\sum_{k=1}^{K}||\widehat{\Theta}_R^{[k]}-\Theta_{R0}^{[k]}||_F^2+\lambda\sum_{k=1}^{K}||(\widehat{\Theta}_R^{[k]}-\Theta_{R0}^{[k]})^-||_1\leq\frac{8\lambda^2\sum_{k=1}^{K}s_k}{c}+\frac{2\sum_{k=1}^{K}||(\widehat{\Theta}_R^{[k]}-\Theta_{R0}^{[k]})^+||_F^2}{c}, \tag{99}$$

并类似地推导出

$$c\sum_{k=1}^{K}||\widehat{\Theta}_R^{[k]}-\Theta_{R0}^{[k]}||_F^2+\lambda\sum_{k=1}^{K}||(\widehat{\Theta}_R^{[k]}-\Theta_{R0}^{[k]})^-||_1\leq\frac{8\lambda^2\sum_{k=1}^{K}s_k}{c}. \tag{100}$$

使用

$$\sum_{k=1}^{K}|||\widehat{\Theta}_{R}^{[k]}-\Theta_{R0}^{[k]}|||_{1}\leq\sum_{k=1}^{K}\left(||\widehat{\Theta}_{R}^{[k]}-\Theta_{R0}^{[k]}||_{F}+||(\widehat{\Theta}_{R}^{[k]}-\Theta_{R0}^{[k]})^{-}||_{1}\right) \tag{101}$$

我们有

$$\sum_{k=1}^{K}|||\widehat{\Theta}_{R}^{[k]}-\Theta_{R0}^{[k]}|||_{1}\leq\frac{16K\lambda\sum_{k=1}^{K}s_{k}}{c}. \tag{102}$$

最后，根据不等式（*79*），基于对 $||W_{0}^{[k]}||_{\infty}$ 和 $||\widehat{W}^{[k]}||_{\infty}$上界的分析，以及 $||$ $(\widehat{\Sigma}^{[k]}-\Sigma_{0}^{[k]}$ $)$ $||_{\infty}$的收敛速率，我们得出结论：

$$\sum_{k=1}^{K}|||\widehat{\Theta}_{w}^{[k]}-\Theta_{0}^{[k]}|||_{1}\frac{16K\lambda\sum_{k=1}^{K}s_{k}}{c}. \tag{103}$$

A.5. 定理*5*的证明

证明5。首先，我们证明余项以 $1/\sqrt{n}$ 的收敛速率依概率收敛。根据定理*1*，我们得到

$$||rem||_{\infty}\leq\sum_{k=1}^{2}||(\widehat{\Theta}^{[k]}-\Theta_{0}^{[k]})(\widehat{\Sigma}^{[k]}-\Sigma_{0}^{[k]})\Theta_{0}^{[k]}||_{\infty}+\sum_{k=1}^{2}||(\widehat{\Theta}^{[k]}-\Theta_{0}^{[k]})(\widehat{\Sigma}^{[k]}\widehat{\Theta}^{[k]}-\mathbf{I}_{p})||_{\infty} \tag{104}$$

定义

$$l(\Theta)=\sum_{k=1}^{2}\{tr(\widehat{\Sigma}^{[k]}\Theta^{[k]})-\log\det(\Theta^{[k]})\}+\lambda\sum_{k=1}^{2}||(\Theta^{[k]})^{-}||_{1}+\rho||(\Theta^{[1]}-\Theta^{[2]})^{-}||_{1}. \tag{105}$$

根据 *Karush-Kuhn-Tucker*（*KKT*）条件，我们得到

$$\widehat{\Sigma}^{[1]}-(\widehat{\Theta}^{[1]})^{-1}+(\lambda+\rho)\widehat{Z}^{[1]}=0, \tag{106}$$

and

$$\widehat{\Sigma}^{[2]}-(\widehat{\Theta}^{[2]})^{-1}+(\lambda-\rho)\widehat{Z}^{[2]}=0, \tag{107}$$

其中 $\widehat{Z}_{ij}^{[k]}=sign(\widehat{\Theta}_{ij}^{[k]})$ 若 $\widehat{\Theta}_{ij}^{[k]}\neq0$，且满足 $||\widehat{Z}^{[k]}||_{\infty}\leq1$。将方程（*106*）两边乘以 $\widehat{\Theta}^{[1]}$，得到

$$\mathbf{I}_{p}-\widehat{\Sigma}^{[1]}\widehat{\Theta}^{[1]}=(\lambda+\rho)\widehat{Z}^{[1]}\widehat{\Theta}^{[1]}. \tag{108}$$

同理，我们有

$$\mathbf{I}_{p}-\widehat{\Sigma}^{[2]}\widehat{\Theta}^{[2]}=(\lambda-\rho)\widehat{Z}^{[2]}\widehat{\Theta}^{[2]}. \tag{109}$$

因此，

$$\begin{aligned}||rem||_{\infty}\leq&\sum_{k=1}^{2}|||(\widehat{\Theta}^{[k]}-\Theta_{0}^{[k]})|||_{1}||(\widehat{\Sigma}^{[k]}-\Sigma_{0}^{[k]})||_{\infty}|||\Theta_{0}^{[k]}|||_{1}\\&+(\lambda+\rho)\sum_{k=1}^{2}|||(\widehat{\Theta}^{[k]}-\Theta_{0}^{[k]})|||_{1}||\widehat{Z}^{[k]}||_{\infty}|||\widehat{\Theta}^{[k]}|||_{1}.\end{aligned} \tag{110}$$

综上，我们得到

$$|||(\widehat{\Theta}^{[k]} - \Theta_0^{[k]})|||_1 \leq b(p+s)\lambda, \tag{111}$$

其中 b 是一个 常数，且与 L有关。根据施瓦茨不等式和韦尔不等式，可得

$$|||\Theta_0^{[k]}|||_1 \leq \sqrt{d+1}\Lambda_{\max}(\Theta_0^{[k]}). \tag{112}$$

$|||\widehat{\Theta}^{[k]}|||_1$ 的界由下式推出

$$|||\widehat{\Theta}^{[k]}|||_1 \leq |||\widehat{\Theta}^{[k]} - \Theta_0^{[k]}|||_1 + |||\Theta_0^{[k]}|||_1. \tag{113}$$

根据 λ的速率，我们得出

$$|||\widehat{\Theta}^{[k]}|||_1 \leq \sqrt{d+1}\Lambda_{\max}(\Theta_0^{[k]}). \tag{114}$$

此外，协方差为 $\Sigma_0^{[k]}$ 的次高斯随机向量意味着 $||\widehat{\Sigma}^{[k]} - \Sigma_0^{[k]}||_\infty = O_p(\sqrt{\log(p)/n})$，其中 O_p 表示依概率有界。可得

$$\begin{aligned} ||rem||_\infty \quad \leq \quad & \frac{4\lambda(8s_1+8s_2+p)}{c}\sqrt{\frac{\log p}{n}}\sqrt{d+1}\max\{\Lambda_{\max}(\Theta_0^{[1]}), \Lambda_{\max}(\Theta_0^{[2]})\} \\ & + (\lambda+\rho)\frac{4\lambda(8s_1+8s_2+p)}{c}\sqrt{d+1}\max\{\Lambda_{\max}(\Theta_0^{[1]}), \Lambda_{\max}(\Theta_0^{[2]})\}. \end{aligned} \tag{115}$$

对于 $\lambda \asymp \rho$, $||rem||_\infty$，其依概率有界于$\tilde{b}(p+s)\sqrt{d+1}\lambda^2$，其中 $\tilde{b}$ 是与 L有关的常数。根据条件$(p+s)\sqrt{d} = o(\sqrt{n}/\log\, p)$, 有 $||rem||_\infty = o_p(1/\sqrt{n})$。根据$(\widehat{\Theta}^{[k]})_{\overline{ii}}(\widehat{\Theta}^{[k]})_{jj} + (\widehat{\Theta}^{[k]})_{ij}^2$的有界四阶矩以及林德伯格中心极限定理，我们完成了定理*5*的证明。

A.6. 定理 *6* 的证明

证明 6. 定理 *6* 的结论可由论证 *(111)-(114)* 得出。对于加权版本， $||rem||_\infty$ 可由 $\tilde{b}s\sqrt{d+1}\lambda^2$界定，证毕。