arXiv:2601.09686v1 [stat.ME] 14 Jan 2026

LARGE：估计高斯图模型的一种局部自适应正则化方法

HaNguyen *[1] 和 Sumanta Basu †[1]

[1]康奈尔大学统计与数据科学系

2026年1月15日

摘要

图形 Lasso（GLASSO）是一种广泛用于学习高维无向高斯图模型（GGM）的算法。给定来自多元正态分布的独立同分布观测，GLASSO 通过最大化对数似然，并对非对角元素施加 ℓ_1 惩罚项来估计精度矩阵。然而，在这种无监督情境下，选择最优正则化参数 λ 仍然是一项重大挑战。一个众所周知的问题是，现有方法（例如样本外似然最大化）只选择单一的全局 λ，而未考虑变量尺度或偏方差的异质性。将数据标准化为单位方差虽是常见的权宜做法，但已被证明会对图恢复产生负面影响。解决图估计中的节点级自适应调参问题，对于计算神经科学等应用至关重要；在这些应用中，脑网络是根据异质性极强的区域特异性 fMRI 数据构建的。

在这项工作中，我们提出了图估计的局部自适应正则化（LARGE），这是一种自适应学习节点级调优参数的方法，旨在改进图估计与选择。在 GLASSO 的每一步块坐标下降中，我们扩展节点级 Lasso 回归，使其能够联合估计回归系数和误差方差，进而指导节点级惩罚项的自适应学习。我们还采用一种 *guiding* 程序，优先处理边际相关系数较高的节点对，以促进真实边的早期检测。我们的方法在稀疏性控制方面带来了理念上的转变：用户无需依赖对尺度敏感的惩罚项 λ，而是指定显著性水平 α，借助标准假设检验提升结果的可解释性。在模拟中，LARGE 在图恢复方面始终优于基准方法，在重复实验中表现出更高的稳定性，并且在最具挑战性的模拟设置中取得了最高的估计准确性。我们通过使用真实 fMRI 数据集估计脑功能连接，展示了该方法的实际效用。

1 引言

无向高斯图模型（GGMs）为表示多元数据的网络结构提供了一个有用的框架。设 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ 为一个无向图，其中顶点集 $\mathcal{V} = \{1, 2, \ldots, p\}$ 表示一个 p 维随机向量 $X = [X_1, X_2, \ldots, X_p]^\top$，边集 $\mathcal{E}$ 则刻画节点之间的条件依赖。在 $\mathcal{G}$ 中，两个节点 (j, k) 之间不存在边，意味着相应的分量 (X_j, X_k) 在给定所有其余变量的条件下相互独立。当 X 服从多元正态分布 $X \sim \mathcal{N}_p(0, \Sigma)$ 时，这种条件独立结构由

*电子邮件：tn325@cornell.edu

†电子邮件：sumbose@cornell.edu

精度矩阵 $\Theta = \Sigma^{-1}$ 的稀疏模式表示。图选择（结构学习）和精度矩阵的估计是重要的统计任务，在遗传学、金融和社交网络等多个领域都有应用 [Wang and Huang, 2014；Fan et al., 2011；Costantini et al.，2019]。

在低维情形下，精度矩阵的最大似然估计量（MLE）就是样本协方差矩阵的逆。但在高维情形下，当 $p > n$ 时，MLE 未定义；或者当 p 较大时，MLE 可能不稳定。Meinshausen 和 B¨uhlmann [2006]，提出了最早的一些用于学习高维高斯图模型的正则化方法之一，即采用逐节点回归（NWR）恢复精度矩阵的稀疏结构。该方法利用高斯图模型的条件独立结构，对每个节点分别针对所有其他节点拟合 Lasso 回归。通常使用交叉验证（CV）为每个回归选择一个单独的调优参数，从而得到 p 个不同的选定值。

与 NWR 不同，当研究者既关注结构学习，也关注完整精度矩阵的估计时，图形 Lasso（GLASSO）是一种广泛使用的方法。它通过最小化负对数似然，并对 Θ [的非对角元素添加 ℓ_1 惩罚项来估计精度矩阵（Yuan and Lin，2007；Banerjee et al.， 2008]）。GLASSO 在高维情形下能够得到稀疏且易于解释的精度矩阵，并将问题转化为一个可通过块坐标下降（BCD）高效求解的凸优化问题[（Friedman et al.， 2008]）。GLASSO 实现中的一个关键环节是选择调优参数λ，它控制稀疏程度，但在实践中，调参仍然颇具挑战。

一种常见的 λ 选择方法是样本外似然最大化或 K 折交叉验证 [Friedman et al.,2008]。该方法将数据划分为训练集和测试集，并通过网格搜索找到在留出测试集上似然得分最高的 λ 。已有其他网格搜索策略被提出，用于选择能够优化特定信息论准则的调优参数。例如，风险膨胀准则（RIC）通过置换消除变量之间的所有关联，并选择使估计的精度矩阵为对角矩阵的最小 λ [Lysen，2009]。直观而言，该方法降低了在真实空图中出现假阳性的可能性。Foygel and Drton [2010] 提出选择能优化扩展贝叶斯信息准则（EBIC）的 λ ，以平衡模型拟合与图密度。另一种方法是 Liu et al. [2010] 提出的基于子抽样的方法，称为正则化选择稳定性方法（StARS）。该方法将数据划分为 K 个子集，并选择能够使不同子样本间估计边集稳定性最高的 λ 。

现有方法的一个根本局限在于，无论精度矩阵的对角元素存在何种异质性，它们都使用单一λ进行相同的惩罚。在高斯图模型中， Θ 的对角元素表示给定所有其他变量时各变量的条件精度（偏方差的倒数）。由于这些偏方差在不同变量之间可能存在显著差异，对所有元素统一应用 λ 会忽略逐节点条件变异性的这种差异。图模型文献已充分记录了这一问题 [Sun 和 Zhang，2013；Ren 等，2015；Jankov´a 和 van de Geer，2018；Carter 等，2023]。我们在图3中展示了忽略偏方差异质性的代价：在模拟网络中，依赖单一 λ 的方法往往会对不同区域产生低估或高估。为处理异质性调优，GLASSO 算法（通过 R 软件包 glasso 实现）支持使用矩阵进行逐元素惩罚，其中第（j， k）个元素为 $\sqrt{\lambda_j \lambda_k}$，而 $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_p]^\top$ 是用户指定向量 [Friedman et al.， 2008]。然而，目前尚无通用指南说明如何恰当地选择这些参数。

作为一种权宜之计，通常建议在进行图估计之前对观测数据进行标准化，使其样本方差为单位方差。然而，Carter 等人 [2023] 通过模拟表明，标准化实际上可能会损害结构学习，在某些情况下，甚至会导致无论 λ 取何值，都无法恢复底层图模型。为解决这一问题，他们提出了偏相关图形套索（PC-GLASSO）方法，同时估计 Θ 的对角元素和偏相关，从而得到完整精度矩阵的估计量。该方法将估计目标从精度矩阵转为尺度不变偏相关矩阵，并使用通过 BIC 准则选择的全局调优参数 λ。对角元素采用惩罚项 $\text{pen}_{ii} = 2\log(\Theta_{ii})$，以确保 PC-GLASSO 具有尺度不变性。

在噪声水平和变量尺度存在异质性的现实应用中，解决 GLASSO 的异质调参问题至关重要。例如，在神经科学应用中，研究人员通常使用 GLASSO 根据功能性磁共振成像（fMRI）数据估计大脑连接 [Smith 等，2011；Zhu 和 Cribben， 2018]。然而，脑区往往具有各自特有的噪声特征，并受到受试者运动和生理变异的影响 [Birn 等，2014；Liu， 2016], ，这些因素都会导致偏方差存在异质性。估计功能连接时若依赖单一的全局调优参数，可能会产生系统性误差：在某些子网络中漏掉真实连接，同时在其他子网络中误判出伪边。这类失真会损害基于连接的生物标志物的有效性，而这类标志物被广泛用于研究认知表现和预测神经系统疾病 [Marek 等， 2022]。

本文提出图估计的局部自适应正则化（LARGE），这是一种自适应学习节点级调优参数的方法。该方法以 GLASSO 为基础，旨在提升高斯图模型的结构学习能力和估计准确性。首先，我们利用 GLASSO 与 NWR 之间广为人知的概念联系：在每个块坐标下降步骤中，我们扩展内部 Lasso 子问题，同时跟踪每个回归的误差方差。随后，使用 p噪声方差估计量自适应地引导 p节点级调优参数的选择。其次，我们发现，若未将方差标准化为单位方差，GLASSO 在 Lasso路径中往往偏向边际方差较大的边，因为阈值化基于原始协方差进行。为此，我们引入一种引导程序，通过优先考虑绝对边际相关性较高的节点对，促进真实边的早期检测。第三，我们的算法采用用户定义的显著性阈值 α执行序贯 F 检验以进行边选择，取代对尺度敏感的惩罚项 λ。这种重新表述使得通过假设检验直接控制图稀疏性成为可能，并提升了方法对实践者的可解释性。最后，我们的 R软件包 large 可通过 https://github.com/hanguyen97/large 获取，其中计算密集型函数以 C 语言编写++。

我们在低维和高维情形下对模拟数据进行了广泛的数值实验，以评估 LARGE 的表现。首先，我们的分析证实，与现有方法相比，使用尺度自适应调优参数能够提升图选择性能。这证明了一个概念：与在所有回归中使用单一全局 λ 相比，局部自适应惩罚项能够增强图恢复。我们还发现，LARGE 在重复实验中的结果变异性明显更低，表明其稳定性更好。其次，当 $n > p$ 时，LARGE 与其他基准方法的表现相当；但在最具挑战性的 $n = p$ 情形下，我们的方法始终优于它们。最后，我们将 LARGE 应用于人类连接组计划（HCP）的一组真实功能性磁共振成像数据，以估计功能连接，从而展示其实际应用价值。

本文其余部分安排如下。第2节介绍 GLASSO 和 Autotune Lasso 的背景知识；Autotune Lasso 是用于学习节点级惩罚项的基础方法。第3节介绍我们提出的方法 LARGE，并详细说明该算法的关键设计选择，包括调优初始化、自适应正则化，以及通过显著性检验进行稀疏性控制。第4节和第5节分别在模拟数据和真实功能性磁共振成像数据上评估我们的方法的性能。

符号说明。 在全文中，对于向量 $\beta = [\beta_1, \beta_2, \ldots, \beta_p]^\top \in \mathbb{R}^p$，我们用 $\beta_{-j} \in \mathbb{R}^{p-1}$ 表示移除第 j 个元素后得到的向量。我们用记号 $\|\beta\|_q$ 表示 β 的 q 范数，即，对于任意 $q > 0$，$(\sum_i |\beta_i|^q)^{1/q}$。一个 p 维随机向量记为 $X = [X_1, X_2, \ldots, X_p]^\top \in \mathbb{R}^p$。设 $X_1, X_2, \ldots, X_n$ 是来自 X 的独立同分布（i.i.d.）随机样本，其中每个样本 $X_i = [X_{i1}, X_{i2}, \ldots, X_{ip}]^\top$， $1 \le i \le n$。我们将得到的数据矩阵 $\mathcal{X} \in \mathbb{R}^{n \times p}$ 记为

$$\mathcal{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}.$$

对于矩阵 A，我们用 $A^\top$ 表示其转置，用 $A_{\cdot j}$ 表示其第 j列，用$A_{\cdot\ -j}$ 表示从 A中删除第 j列后得到的矩阵。矩阵元素用下标表示，即 A_{ij}。对于两个矩阵 A 和 B， A 与 B之间的 Frobenius内积记为 $\langle A,\ B\rangle$，定义为 tr($A^\top B$)。

若 A 是方阵，tr(A)表示其迹，即其特征值之和。矩阵 A 的弗罗贝尼乌斯范数为 $\|A\|_F = \sqrt{\langle A,\ A\rangle}$。我们用 $\mathbb{S}^p_+$ 表示半正定 $p \times p$ 矩阵锥，$\mathbb{S}^p_+ := \{A \in \mathbb{R}^{p\times p} \mid A = A^\top,\ A \succeq 0\}$，并用 $\mathbb{S}^p_{++}$ 表示正定 $p \times p$ 矩阵集合，$\mathbb{S}^p_{++} := \{A \in \mathbb{R}^{p\times p} \mid A = A^\top,\ A \succ 0\}$。

2 背景

我们首先简要介绍高斯图模型框架。设 $\mathcal{G} = (\mathcal{V},\ \mathcal{E})$ 是一个顶点集为 $\mathcal{V} = \{1, 2, \ldots, p\}$、边集为 $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ 的无向图。考虑一个 p 维随机向量 $X = [X_1, X_2, \ldots, X_p]^\top$，其中每个分量都由图中的一个顶点索引。若对于任意一对顶点 j 和 k，在给定所有其他变量的条件下，相应变量 X_j 和 X_k 条件独立，则称 X 的分布遵循由 $\mathcal{G}$ 定义的底层图模型。假设 $X \sim \mathcal{N}_p(0,\ \Sigma)$，可以证明，逆协方差矩阵（亦称精度矩阵） $\Theta = \Sigma^{-1}$ 通过其非对角元素的稀疏模式，编码了 X 各元素之间的条件独立关系。具体而言，

$$X_j \perp\!\!\!\perp X_k \mid X_{\mathcal{V}\setminus\{(j,k)\}} \ \Leftrightarrow\ \Theta_{jk} = 0 \ \Leftrightarrow\ (j,k) \notin \mathcal{E}. \tag{2.1}$$

非零元素的大小反映了条件关联的强度。设有 n 个独立同分布样本X_1，.....，$X_n \in \mathbb{R}^p$，我们的目标是估计精度矩阵，并恢复底层图 $\mathcal{G}$ 的边集 $\mathcal{E}$。这也称为（逆）协方差选择问题 [Dempster，1972]。

2.1 使用 GLASSO 进行精度矩阵估计

为解决此问题，GLASSO 估计量在正定精度矩阵空间上，最小化带有 ℓ_1 惩罚项的负对数似然:

$$\hat{\Theta}_{\text{GLASSO}} := \arg\min_{\Theta \in \mathbb{S}^p_{++}} \left\{ -\log\det(\Theta) + \operatorname{tr}(S\Theta) + \lambda \|\Theta\|_{1,\text{off}} \right\}, \tag{2.2}$$

其中，$S = \frac{1}{n}\sum_{i=1}^n X_i X_i^\top$ 是样本协方差矩阵，$\|\Theta\|_{1,\ \text{off}} = \sum_{i\neq j} |\Theta_{ij}|$，且 $\lambda > 0$ 是调优参数。该凸优化问题的 Karush-Kuhn-Tucker（KKT）条件为:

$$-\Theta^{-1} + S + \lambda \cdot \Gamma = 0 \text{ where } \Gamma_{ij} = \begin{cases} \operatorname{sign}(\Theta_{ij}) & \text{if } \Theta_{ij} \neq 0, \\ \in [-1, 1] & \text{if } \Theta_{ij} = 0. \end{cases} \tag{2.3}$$

Friedman et al. [2008] 提出了一种块坐标下降算法，通过迭代更新 Θ^{-1} 的一行和一列，同时保持其余部分不变，从而高效地求解该系统。令 W 表示 Θ^{-1} 的当前估计。在每次迭代中，将目标行和列置换到最后一个位置，并将 W 和 S 划分如下:

$$W = \begin{bmatrix} W_{11} & w_{12} \\ w_{12}^\top & w_{22} \end{bmatrix}, \quad S = \begin{bmatrix} S_{11} & s_{12} \\ s_{12}^\top & s_{22} \end{bmatrix}.$$

其中，W_{11}，$S_{11} \in \mathbb{R}^{(p-1)\times(p-1)}$、$w_{12}$，$s_{12} \in \mathbb{R}^{p-1}$，以及 w_{22}，$s_{22} \in \mathbb{R}$。为使分块表示更清晰，这里用小写字母表示列向量和矩阵元素。利用分块矩阵求逆，（2.3）式的右上块可改写为

$$W_{11}\beta - s_{12} + \lambda \cdot \Gamma_{12} = 0, \quad \text{where } \beta = -\frac{\Theta_{12}}{\Theta_{22}} \in \mathbb{R}^{p-1}. \tag{2.4}$$

˄令 $V = W_{11}$。这将得到一种修正的 Lasso 形式，可通过坐标下降（CD）高效地更新 β：

$$\hat{\beta}_k \leftarrow \mathcal{S}_\lambda\left(s_{12k} - \sum_{l\neq k} V_{kl}\hat{\beta}_l \right) / V_{kk}; \quad k = 1, 2, \ldots, p-1, \tag{2.5}$$

其中， $\mathcal{S}_\lambda(z) = \text{sign}(z)(|z| - \lambda)_+$ 表示软阈值函数。GLASSO 算法由一个外循环和一个内循环组成：外循环每次更新 W 的一行和一列，内循环则通过对 β 进行坐标逐点更新来求解修正的 ˆ Lasso 问题 (2.4)，直至收敛。

***GLASSO*与*NWR*的联系。**众所周知，GLASSO 与 NWR[Hastie等，2015b], 密切相关。NWR 使用 Lasso 将每个节点对其余节点进行回归 j ，以估计各节点的条件依赖 $\mathcal{V} \setminus \{j\}$ 。令 $y = \mathcal{X}_{\cdot j}$ 表示数据矩阵 $\mathcal{X}$的第 j列，令$\mathcal{Z} = \mathcal{X}_{\cdot,-j}$ 表示移除第 j列后得到的矩阵。NWR 通过求解以下问题来估计节点 j 的稀疏邻域:

$$\hat{\beta}^{\text{NWR}} := \arg\min_{\beta \in \mathbb{R}^{p-1}} \left\{ \frac{1}{2n}(y - \mathcal{Z}\beta)^\top (y - \mathcal{Z}\beta) + \lambda \|\beta\|_1 \right\}, \tag{2.6}$$

其中 $\|\beta\|_1 = \sum_{k=1}^{p-1} |\beta_k|$，且 $\lambda > 0$ 是调优参数。相应的次梯度条件 is:

$$\frac{1}{n}\mathcal{Z}^\top \mathcal{Z}\beta - \frac{1}{n}\mathcal{Z}^\top y + \lambda \,\text{sign}(\beta) = 0. \tag{2.7}$$

ˆ β^{NWR} 的逐坐标更新形式为:

$$\hat{\beta}_k^{\text{NWR}} \leftarrow \mathcal{S}_\lambda \left(\frac{1}{n}\mathcal{Z}_{\cdot k}^\top y - \frac{1}{n}\sum_{l \neq k} \mathcal{Z}_{\cdot k}^\top \mathcal{Z}_{\cdot l} \hat{\beta}_l^{\text{NWR}} \right) \Big/ \left(\frac{\|\mathcal{Z}_{\cdot k}\|_2^2}{n} \right). \tag{2.8}$$

与 (3.3) 比较可知，$s_{12,k}$ 是 $\frac{1}{n} Z_{\cdot k}^\top y$的对应项。同样， $\sum_{\ell \neq k} V_{k\ell}$ ˆβ_ℓ 和 $\frac{1}{n}\sum_{\ell \neq k} \mathcal{Z}_{\cdot k}^\top \mathcal{Z}_{\cdot \ell}$ ˆβ_ℓ^{NWR}均用于估计 W_{11}。正如 Friedman et al. [2008], 所指出的，GLASSO 会在每个块坐标下降步骤中迭代重新计算 NWR 回归，以更新精度矩阵。这一联系也凸显了一个关键局限：GLASSO 实际上在内部执行 p 个独立回归，却对所有回归统一使用一个全局调优参数 λ 。这种统一惩罚忽略了偏方差的异质性，而这种异质性反映在 Θ的对角元素中。更具适应性的方法是在每次区块更新时为各节点分配节点级调优参数λ_j 。然而，通过网格搜索寻找多个 λ_j 会耗费大量计算资源，而且由于缺乏明确的联合选择策略，也颇具挑战性。

为什么数据标准化并不足够。 对标准化数据运行 GLASSO，相当于求解:

$$\hat{\Theta}_{\text{norm}} = \arg\min_{\Theta \in \mathbb{R}_{++}^p} \{ -\log\det(\Theta) + \text{tr}(R\Theta) + \lambda \|\Theta\|_{1,\text{off}} \}, \tag{2.9}$$

其中，ˆ Θ_{norm} 估计逆相关矩阵， $D =$diag(S)，而 $R = D^{-1/2} S D^{-1/2}$ 是样本ˆ相关矩阵。 Θ_{norm} 本身具有独立的研究意义，并且与加权GLASSO估计量密切相关:

$$\hat{\Theta}_{\text{w}} = \arg\min_{\Theta \in \mathbb{R}_{++}^p} \{ -\log\det(\Theta) + \text{tr}(S\Theta) + \lambda \sum_{i \neq j} D_{ii}^{1/2} D_{jj}^{1/2} |\Theta_{ij}| \}, \tag{2.10}$$

其中，ˆ$\Theta_{\text{norm}} = D^{1/2}$ˆ$\Theta_{\text{w}} D^{1/2}$ [Jankov´a 和 van de Geer， 2018]。ˆ Θ_{norm}的对角元素记为 $D_{\text{norm}} =$ ˆ ˆ diag(Θ_{norm}), 无法保证经过标准化或保持一致；事实上，$(D_{\text{norm}})_{ii} = (\Theta_{\text{w}})_{ii} D_{ii}$。因此，在所有节点上使用单一的全局调优参数 λ 并不理想。

这一挑战进一步凸显了估计 Θ时采用 p 不同调优参数 λ_j 的必要性。为解决这一问题，我们针对 GLASSO 算法中的每个 Lasso 子问题，采用 Sadhukhan 等人 [2025] 提出的 Autotune Lasso 程序。Autotune Lasso 会跟踪噪声方差的估计值，并据此在回归问题中自适应地确定调优参数。这种方法在稀疏回归模型中展现出良好的效果，我们将在下一节中介绍。

2.2 通过 Autotune Lasso 自动调参

考虑线性模型 $y = \mathcal{Z}\beta + \varepsilon$，其中 $\varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ 来自 NWR问题 (2.6)。令 $\mathcal{Z}$ 的每一列都经过标准化，使其方差为单位方差。Autotune Lasso 通过优化带有 ℓ_1 惩罚项的完整似然，同时估计回归系数 β和误差方差 σ^2：

$$(\hat{\beta}, \hat{\sigma}^{-2}) := \arg \min_{(\beta, \sigma^{-2}) \in \mathbb{R}^{p-1} \times \mathbb{R}_+} \left\{ \frac{1}{2} \log(\sigma^2) + \frac{1}{2n\sigma^2} \|y - \mathcal{Z}\beta\|_2^2 + \lambda_0 \|\beta\|_1 \right\}, \tag{2.11}$$

其中 $\lambda_0 = \frac{1}{\text{Var}(y)} \| \frac{1}{2n} \mathcal{Z}^\top y \|_\infty$ 为固定值。对 β和 σ^{-2}求偏导数，可得到以下闭式更新：

$$\hat{\sigma}^2 \leftarrow \frac{\|y - \mathcal{Z}\hat{\beta}\|_2^2}{n}, \tag{2.12}$$

$$\begin{aligned} \hat{\beta}_k &\leftarrow \mathcal{S}_{\lambda_0 \hat{\sigma}^2} \left(\frac{1}{n} \mathcal{Z}_{\cdot k}^\top y - \frac{1}{n} \sum_{l \neq k} \mathcal{Z}_{\cdot k}^\top \mathcal{Z}_{\cdot l} \hat{\beta}_l \right) \\ &= \mathcal{S}_{\lambda_0 \hat{\sigma}^2} \left(\frac{1}{n} \langle \mathcal{Z}_{\cdot k}, r_k \rangle \right) \end{aligned} \tag{2.13}$$

$\hat{}$ 其中 $r_k = y - \sum_{l \neq k} \mathcal{Z}_{\cdot l} \beta_l$ 为部分残差。由于目标函数 (2.11) 关于 β 和σ^{-2}是双凸的，Sadhukhan 等人 [2025] 采用交替估计方法，每次更新一组参数，同时固定另一组参数。对于任意给定的 σ^2，求解 β 本质上就是求解调优参数为 $\lambda := \lambda_0 \hat{\sigma}^2$的 Lasso。因而，迭代更新 $\hat{\sigma}^2$ 会自动地、以数据驱动的方式，针对一系列 λ 值求解一系列 Lasso 问题。

为提高 σ^2的有限样本估计，Sadhukhan 等人 [2025] 利用精心构建的线性模型的均方误差 (MSE) 作为估计量。在每次迭代中，Autotune Lasso 根据偏残差r_k的标准差 (SD)，对回归系数 β_k 的重要性进行排序。Autotune Lasso 采用序贯 F 检验指导的前向选择，在一组预测变量上拟合线性模型，并利用均方误差估计 σ^2：

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - |\hat{S}|}, \tag{2.14}$$

其中，$|\hat{S}|$ 表示选定变量的数量。为完整起见，我们在算法 1 中给出了 Autotune Lasso 过程的简化版本。

3 方法

本节将介绍我们的方法——图估计的局部自适应正则化。我们扩展了 GLASSO估计量，使其能够使用按列调节参数。

$$\hat{\Theta}_{\text{LARGE}} := \arg \min_{\Theta \in \mathbb{S}_{++}^p} \left\{ -\log\det(\Theta) + \text{tr}(S\Theta) + \sum_{j=1}^p \lambda_j \|\Theta_{\cdot j}\|_{1,\text{off}} \right\}, \tag{3.1}$$

其中$\|\Theta_{\cdot j}\|_{1,\text{ off}} = \sum_{k \neq j} |\Theta_{kj}|$。下面，我们将逐步介绍如何使用 LARGE 算法求解目标函数（3.1）。我们在算法 2 中概述完整流程。在附录中，我们给出了 LARGE 的另一种形式，该形式同时对 Θ 的对角元素和非对角元素进行惩罚。

算法 1 Autotune Lasso [**Sadhukhan 等人，2025**]

输入： y、归一化预测变量 $\mathcal{Z}$ （即$\|\mathcal{Z}_{\cdot j}\|_2^2 = n$）、序贯 F 检验显著性水平 α1: 初始化 $\hat{\beta} \leftarrow 0$、$r \leftarrow y$、$\hat{\sigma}^2 \leftarrow$Var($y$)、 $\lambda_0 \leftarrow \frac{1}{\mathrm{Var}(y)} \left\| \frac{\mathcal{Z}^\top y}{2n} \right\|_\infty$、预测变量排序 $\leftarrow \{1, \ldots, p-1\}$、支持集 $\leftarrow \emptyset$、sigma 更新标志 $\leftarrow$TRUE2: **当** error $\geq$误差容差 **时**3: 支持集$^{(\mathrm{old})} \leftarrow$支持集，$\hat{\beta}^{(\mathrm{old})} \leftarrow \hat{\beta}$4: **对于** k （属于预测变量排序）**执行**

5: $\hat{\beta}_k \leftarrow$ 软阈值$_{\lambda=\lambda_0}\hat{\sigma}^2 \left(\frac{1}{n}\mathcal{Z}_{\cdot k}^\top r + \hat{\beta}_k\right)$6: $r \leftarrow r + \mathcal{Z}_{\cdot k}\left(\hat{\beta}_k^{(\mathrm{old})} - \hat{\beta}_k\right)$7: 结束循环 8: 若 sigma 更新标志 $\triangleright$ 估计 β ==TRUE 则 $\triangleright$ 估计 σ^2预测变量排序 $\leftarrow$ 按 SD(r_1) . . . , SD(r_{p-1}) 降序排列 10: 使用序贯 F 检验拟合普通最小二乘模型，以进行变量选择，截断值为 $= F_{\alpha;1,n-i}$11: 支持集 $\leftarrow \{k:\ \mathcal{X}_{\cdot k}$ 选入普通最小二乘模型$\}$12: $\hat{\sigma}^2 \leftarrow \frac{\sum_{i=1}^n (y_i - y_i)^2}{n - |\mathtt{Support.Set}|}$ 13: 若支持集 $\subseteq$支持集$^{(\mathrm{old})}$ 则 14: sigma 更新标志 $\leftarrow$FALSE $\triangleright$ 提前停止$\hat{\sigma}^2$ 更新 15: 结束若 16: 结束若 17: 误差 $\leftarrow \|\hat{\beta} - \hat{\beta}^{(\mathrm{old})}\|_1 / \|\hat{\beta}^{(\mathrm{old})}\|_1$18: 结束循环 输出：$\hat{\beta}, \hat{\sigma}^2, \lambda$

初始化节点级调优参数。 为初始化优化过程，我们为每个分量 $\mathcal{X}_{\cdot j}$指定节点级调优参数 λ_j。按照 Sadhukhan 等人 [2025], 的做法，我们令 $\lambda_{0j} = \frac{1}{\mathrm{Var}(\mathcal{X}_{\cdot j})} \max_{k \neq j} \left|\frac{1}{2n}\langle \mathcal{X}_{\cdot j}, \mathcal{X}_{\cdot k}\rangle\right|$，并将噪声方差估计初始化为 $\hat{\sigma}_j^2 = \mathrm{Var}(\mathcal{X}_{\cdot j})$。代入 λ_j 的表达式后，节点 j 的 Lasso 子问题的初始调优值为 $\lambda_j = \max_{k \neq j} \left|\frac{1}{2n}\langle \mathcal{X}_{\cdot j}, \mathcal{X}_{\cdot k}\rangle\right|$。

将 ***Lasso*** 子问题替换为 ***Autotune Lasso***。 在算法的核心部分，我们遵循 GLASSO 框架，并将逐节点 Lasso 子问题替换为 Autotune Lasso 回归，以自适应地为每个分量 $\mathcal{X}_{\cdot j}$ 学习调优参数 λ_j 。这是与 GLASSO 的主要区别：我们不使用全局惩罚项，而是根据局部噪声水平更新 λ_j ，并用它对 β 进行阈值处理。在每个块坐标下降步骤中，我们将 KKT 条件 (2.3) 的右上区块重新表述为

$$W_{11}\beta - s_{12} + \lambda_j \cdot \Gamma_{12} = 0, \quad \text{where } \beta = -\frac{\Theta_{12}}{\Theta_{22}} \text{ and } \lambda_j = \hat{\sigma}_j^2 \lambda_{0j}. \tag{3.2}$$

$\hat{}$ 这引出了一个交替最小化方案，其中 β 通过坐标下降进行更新：

$$\hat{\beta}_k \leftarrow \mathcal{S}_{\lambda_j = \hat{\sigma}_j^2 \lambda_{0j}} \left(s_{12k} - \sum_{l \neq k} V_{kl} \hat{\beta}_l \right) / V_{kk}, \tag{3.3}$$

并使用精心构造的线性模型的均方误差来更新逐节点噪声方差$\hat{\sigma}_j^2$：

$$\hat{\sigma}_j^2 \leftarrow \frac{\sum_{i=1}^n \left(X_{ij} - \hat{X}_{ij}\right)^2}{n - |\hat{S}_j|}, \tag{3.4}$$

其中， $|\hat{S}_j|$ 表示选定边的数量。按照第 2.2 小节中的描述，通过前向选择和序贯F检验构建线性模型。随后，我们针对节点 j 局部更新惩罚水平

Al算法 2 图估计的局部自适应正则化（LARGE）

输入： 数据矩阵 $\mathcal{X} \in \mathbb{R}^{n\times p}$，序贯 F 检验显著性水平 α

对于 $j = 1,\ldots,p$，初始化 $\hat{\sigma}_j^2 \leftarrow \mathrm{Var}(\mathcal{X}_{\cdot j})$，$\lambda_{0j} \leftarrow \frac{1}{\mathrm{var}(\mathcal{X}_{\cdot j})}$ 最大值$_{k\neq j}\left|\frac{1}{2n}\langle \mathcal{X}_{\cdot j}, \mathcal{X}_{\cdot k}\rangle\right|$

令 $S \leftarrow$ 协方差$(\mathcal{X})$；初始化 $W \leftarrow S$

重复 对于 $j = 1$、2、p:

(a) 划分 $W = \begin{bmatrix} W_{11} & w_{12} \\ w_{12}^\top & w_{22} \end{bmatrix}$，$S = \begin{bmatrix} S_{11} & s_{12} \\ s_{12}^\top & s_{22} \end{bmatrix}$.

(b) 使用 Autotune Lasso 求解 $W_{11}\beta - s_{12} + \lambda_j \cdot \mathrm{sign}(\beta) = 0$：

令 $V \leftarrow W_{11}$；初始化 $\hat{\beta} \leftarrow 0$

初始化预测变量排序 $_j \leftarrow \{1,\ldots,p-1\}$，支持集 $_j \leftarrow \emptyset$，sigma 更新标志 $_j \leftarrow$ TRUE

while 误差 $\geq$ 误差容差 **do**

支持集 $_j^{(\text{旧})} \leftarrow$ 支持集$_j$，$_\beta^{(\text{old})} \leftarrow$

对于 k 在预测变量排序$_j$ **中执行** ▷ 估计 β

$\beta_k \leftarrow$ 软阈值$_{\lambda_j = \hat{\sigma}_j^2 \cdot \lambda_{0j}}\left(s_{12k} - \sum_{l\neq k} V_{kl}\cdot \beta_l\right)/V_{kk}$

结束循环

如果 sigma 更新标志$_j$ ==为 TRUE **则** ▷ 估计 σ_j^2

预测变量排序$_j \leftarrow$按以下内容降序排列：k

$$\begin{cases} |\text{相关系数}(\mathcal{X}_{\cdot j}, & \text{如果是第 8 行的第一次迭代} \\ \quad \text{标准差} & \\ \mathcal{X}_{\cdot k})\| \mathcal{X}_{\cdot j} - \sum_{l\notin\{j,k\}} \mathcal{X}_{\cdot l}\beta_l, & \text{否则} \end{cases}$$

使用序贯F检验进行变量选择，并以 $= F_{\alpha;1,n-i}$ 为截断值来拟合线性模型

支持集 $_j \leftarrow \{k : \mathcal{X}_{\cdot k}$ 已选入线性模型$\}$

$\hat{\sigma}_j^2 \leftarrow \frac{\sum_{i=1}^n (X_{ij} - \hat{X}_{ij})^2}{n - |\texttt{Support.Set}_j|}$

如果 支持集$_j \subseteq$支持集$_j^{(\text{old})}$ **则**

sigma 更新标志$_j \leftarrow$FALSE

结束如果

结束如果

结束循环

(c) $w_{12} \leftarrow W_{11}\hat{\beta}$ 并进行对称化

直到 所有 j均收敛

对于 $i = 1,\ldots,p$, $\hat{\Theta}_{jj} \leftarrow \hat{\sigma}_j^{-2}$ 以及 $\hat{\Theta}_{j,-j} \leftarrow -\hat{\beta}\hat{\Theta}_{jj}$ 并进行对称化

输出： $\hat{\Theta}$, $(\hat{\sigma}_j^2, \lambda_j)$ 表示 $j = 1,\ldots,p$

即 $\lambda_j = \lambda_{0j}\hat{\sigma}_j^2$，根据其对应 Lasso 子问题的噪声水平调整正则化强度。

LARGE 算法保留了 GLASSO 的外循环，每次更新 W 的一行和一列，而内循环则运行 Autotune Lasso，直至收敛。对于第一个之后的块坐标下降步骤（算法 2 第 3 行），我们使用 λ_j 在上一次外循环迭代中的值作为初始值，而不是在 $\lambda_j = \max_{k\neq j}\left|\frac{1}{2n}\langle \mathcal{X}_{\cdot j}, \mathcal{X}_{\cdot k}\rangle\right|$ 处重新初始化。这保留了迭代间的信息，并有助于块坐标下降沿着正则化路径摆脱停滞。

引导初始 ***Lasso***路径。根据公式（3.3），我们可以看到，每个系数 β_k 都通过对原始协方差进行阈值化来更新。当各个变量的方差存在较大差异时，这些协方差会受到具有较大边际方差的特征的主导。这有助于解释为什么通常建议在应用 Lasso 或 GLASSO 之前对变量进行标准化 [Hastie 等人，2015a]。在我们的设定中，与目标节点 j 相连的高方差边往往会产生更大的偏残差，从而导致伪边沿着 Lasso路径过早地被纳入。

为解决这一问题，我们在内循环的第一次迭代中（算法 2 的第 8 行），在使用序贯 F 检验对 $\hat{}$ 线性模型执行变量选择之前，通过根据绝对边际相关性对边重要性进行排名，来引导初始路径。具体而言，对于每个目标节点 j，我们根据其绝对边际相关性对 β_k的重要性进行排名：

$$|\mathrm{Cor}(\mathcal{X}_{\cdot j}, \mathcal{X}_{\cdot k})| = \frac{|\langle \mathcal{X}_{\cdot j}, \mathcal{X}_{\cdot k} \rangle|}{\|\mathcal{X}_{\cdot j}\|_2 \, \|\mathcal{X}_{\cdot k}\|_2}. \tag{3.5}$$

这一步的灵感来自 Fan 和 Lv 的确定性独立筛选（SIS）规则 [2008], ：该规则按边际相关性对变量排序，以便在模型拟合前降低维度。我们的目标是，在一开始优先考虑与目标节点 j 具有较高边际相关性的边，从而获得更准确的初始线性模型和 $\hat{}\sigma_j^2$ 的估计。

我们使用合成数据进行了一项简单模拟，以了解引导程序如何影响噪声方差估计 $\hat{}\sigma_j^2$ 的准确性。我们从 $X \sim \mathcal{N}$ (0, Θ^{-1}) 生成独立同分布数据，其中 (n, p) =(200，50)。为模拟异质噪声水平，我们构造了一个三对角矩阵 $\Theta \in \mathbb{R}^{50\times 50}$，其中包含两个不同的子区块： $\Theta_{ii} = 10$, $\Theta_{i,\ i+1} = \Theta_{i+1,\ i} = 3$，适用于 $i = 1$, ..., 25；以及 $\Theta_{ii} = 1$, $\Theta_{i,\ i+1} = \Theta_{i+1,\ i} = 0.3$，适用于 $i = 26$, ..., 50。我们运行 LARGE，采用 $\alpha = 2\%$ 的 F 检验显著性水平，并重复实验 20 次。为进行评估，我们将随机选取节点的噪声方差估计 $\hat{}\sigma_j^2$ 与预言机线性模型中相应的估计值进行比较。

如图 1 所示，与预言机线性基准方法相比，未采用引导过程时，$\hat{}\sigma_j^2$ 的估计值始终偏高。相反，采用引导后，估计值与线性模型参照更加接近，表明其正逐渐收敛至真实误差方差。

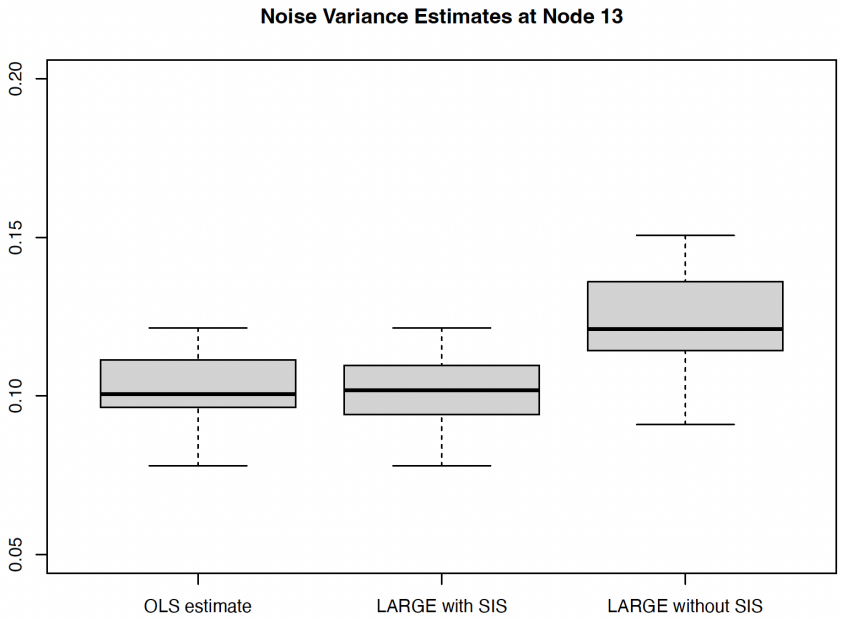


图 1： 在三对角精度矩阵下，随机选取的节点的噪声方差估计值$\hat{}\sigma_j^2$ 箱线图。数据生成自 $X \overset{iid}{\sim} \mathcal{N}(0, \Theta^{-1})$，其中 (n, p) =(200， 50)，汇总了 *20* 次重复实验的结果。引导过程得到的估计值与预言机线性模型的结果更加接近。

通过显著性检验控制稀疏性。 对于每个节点，我们使用序贯 F 检验这一经典的逐步回归工具来执行变量选择，并构建用于估计 σ_j^2 的线性模型。该过程会考察一系列嵌套线性模型，以及随着更多变量（边）加入而产生的相应残差平方和 (RSS)，并在残差平方和的下降不再具有统计

显著性（达到所选显著性水平 α [）时停止 2020]。这使得 α 成为唯一的调优参数，从而能够直接且易于解释地控制图稀疏性，有别于依赖尺度敏感的惩罚项参数 λ 的方法。通过利用经典的 F统计量，该框架避免了变量尺度问题，并允许实践者根据应用场景选择 α 。

4 数值实验

在本节中，我们通过全面的模拟研究评估 LARGE 的图选择和估计准确性。根据不同模拟设置下的经验表现，F 检验显著性水平设为 $\alpha = 2\%$。我们发现，虽然设为 $\alpha = 5\%$ 时真阳性率相近，但假阳性率略高。

基准方法：我们将本方法与 GLASSO 现有的调优参数选择方法进行比较，包括 K折交叉验证、RIC、扩展贝叶斯信息准则和 StARS。我们使用 CVglasso软件包 [Galloway，2018] 运行五折交叉验证，并使用 huge软件包进行 RIC、扩展贝叶斯信息准则和 StARS [Jiang et al.，2021]。为确保公平比较，我们使用每种方法选择调优参数，然后应用 glasso软件包中的原生函数进行估计 Θ [Friedman et al.，2019]。

此外，我们还与另外两种方法进行比较：(1) PC-GLASSO [Carter et al.，2023]，，该方法同样通过将图估计分解为对偏相关和 Θ 的对角元素分别进行惩罚估计，来处理异质调参问题；(2) 调参不敏感图估计与回归（TIGER） [Liu and Wang，2017]，，这是一种具有调参不敏感性质的常用方法。

生成模型： 基于 Ren 等人 [2015], 的设定，我们构造了由三个大小相等、对角值比例为 1:1:1 的区块组成的 $p \times p$ 精度矩阵 Θ。三个区块的对角元素分别设为（α_1、 α_2、 α_3） = （10、1、0.5）。该设定反映了节点级偏方差可能存在显著差异这一事实，因此单一全局调优参数很难取得良好效果。对于 Θ，我们考虑以下数据生成过程（DGP）：

1. **DGP1：** 带宽为1的图模型。在每个区块内，元素对于 $k \in \{1, 2, 3\}$遵循 $\Theta_{i,\ i+1} = \Theta_{i+1,\ i} = 0.3\alpha_k$ 。

2. **DGP2：** 带宽为2的图模型，其中 $\Theta_{i,\ i+1} = \Theta_{i+1,\ i} = 0.3\alpha_k$ 以及 $\Theta_{i,\ i+2} = \Theta_{i+2,\ i} = 0.2\alpha_k$。 3. **DGP3：** 块图模型，包含三个区块，其中每个区块都是稀疏性为 1% 的随机图。非零元素设为 $\Theta_{ij} = 0.4\alpha_k$。 4. **DGP4：** 枢纽图模型，包含 $p/15$ 个枢纽节点。非对角元素按 α_k进行缩放。5. **DGP5：** 稀疏性为 1% 的随机图模型。非零元素设为 0.4，然后根据对角元素 α_k 进行缩放。

对于每种设定，我们从多元高斯分布 $\mathcal{N}_p(0, \Theta^{-1})$ 中抽取随机样本，其中 $n \in \{300, 500\}$ 和 $p \in \{100, 300\}$。所有实验均重复 50 次。

性能指标：为评估恢复正确稀疏结构的能力，我们计算受试者工作特征曲线下的面积（曲线下面积）。为评估估计准确性，我们计算估计的精度矩阵与真实精度矩阵之间的相对均方误差（RMSE）。具体而言，我们关注非对角元素：

$$\mathrm{RMSE}_{\mathrm{off}} = \frac{\|\hat{\Theta}_{\mathrm{off}} - \Theta_{\mathrm{off}}\|_F^2}{\|\Theta_{\mathrm{off}}\|_F^2}, \tag{4.1}$$

其中， Θ_{off} 表示将对角元素设为零后的矩阵 Θ 。我们仅针对非对角元素计算RMSE，因为它们编码了变量之间的条件依赖，并直接决定图估计性能。我们报告每项性能指标在50次重复实验中的均值和标准差（SD）。

4.1 图选择

表1报告了曲线下面积数值，图2展示了示例 ROC 曲线。LARGE 在几乎所有设置下都取得最高的曲线下面积，表明其在异质节点噪声水平下具有更优的图恢复能力。值得注意的是，LARGE 在重复实验中的标准差也最低，说明其性能更加稳定。正如预期，交叉验证方法所得曲线下面积的标准差明显更高，这可能是因为其对重复实验中随机训练-测试集划分较为敏感。有趣的是，随着图数据生成过程变得更具挑战性，例如在 $p = 300$ 情形下，样本量从$n = 500$ 降至 $n = 300$ 时，PC-GLASSO 和 TIGER 的性能会急剧下降。

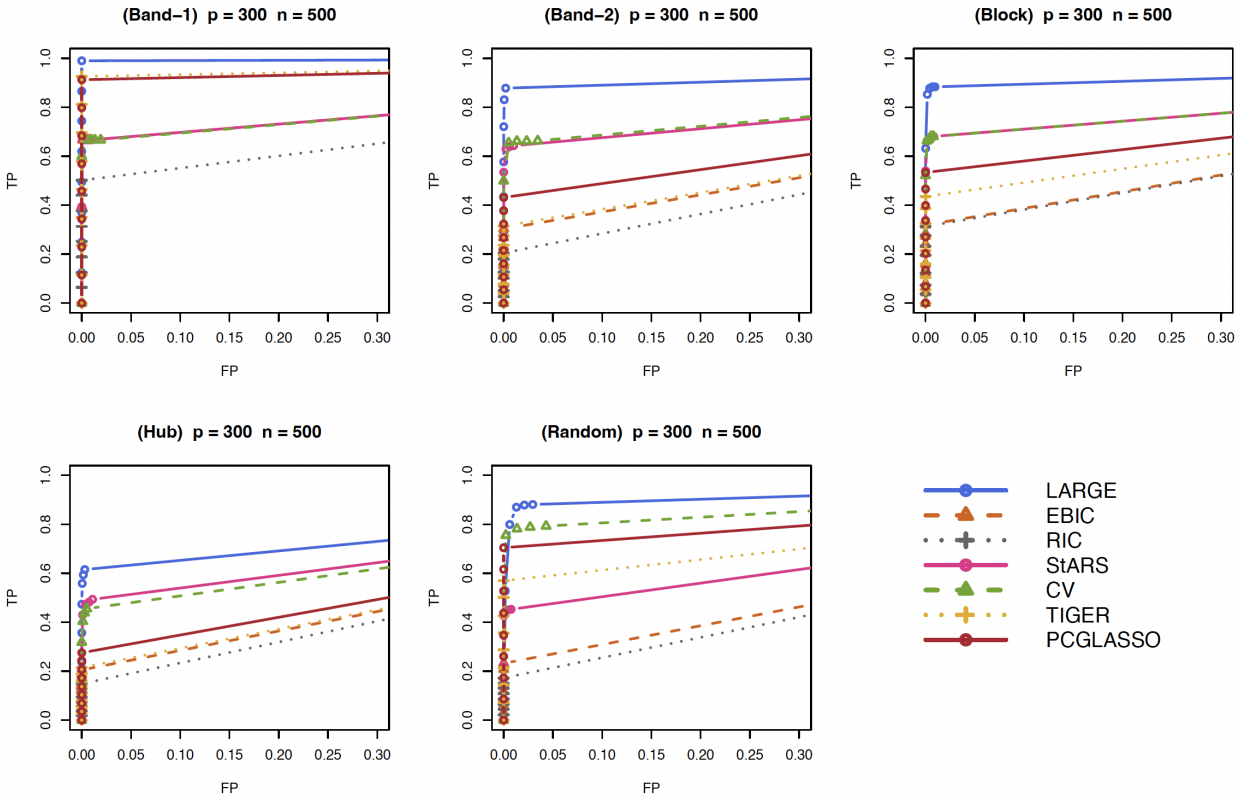
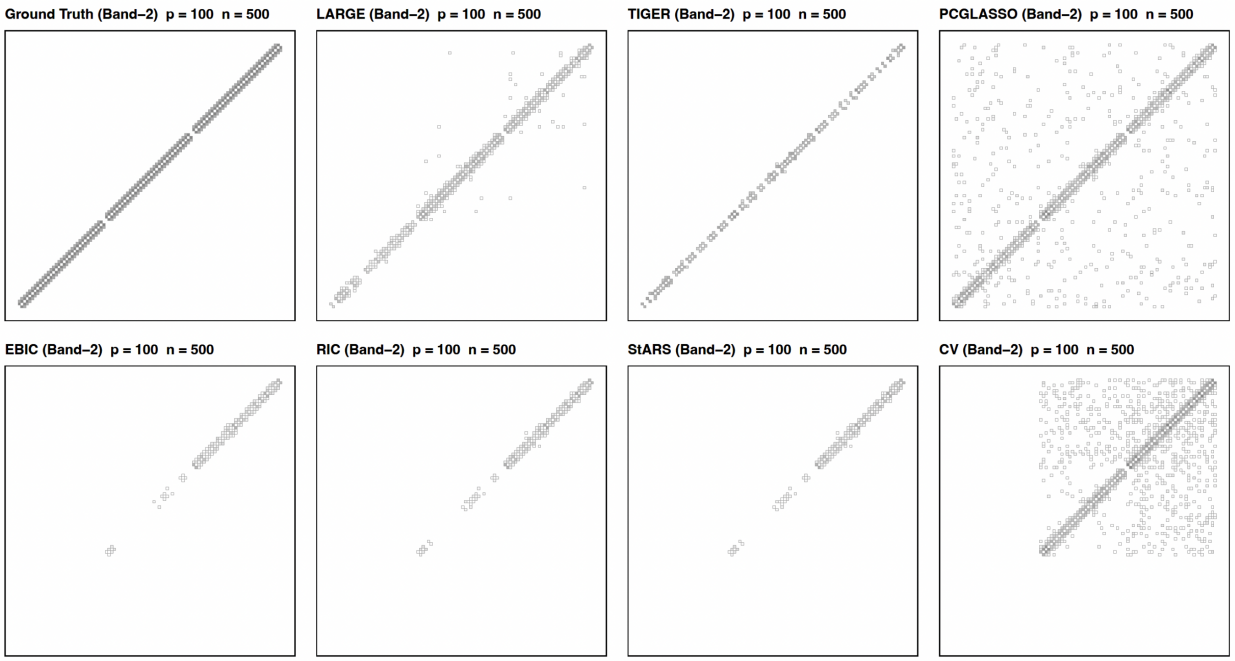


图2：所有模拟设置下单次重复实验的 *ROC* 曲线，其中 $(n,\ p) = (500,\ 300)$。在这一高维情形下，LARGE 在各种图结构中始终取得最高的曲线下面积。

为说明将调优参数适配于局部噪声水平的重要性，我们在图3中展示了带宽为2的图模型下各种方法估计的精度矩阵。在这个简单示例中，LARGE 正确识别了所有真实边，且假阳性极少。相比之下，EBIC、RIC 和 StARS（图3第二行）选择了过于保守的 λ 值，导致估计不足并遗漏关键边，尤其是网络第一区块内的边。与此同时，五折交叉验证系统性地高估了图中的某些区域，同时低估了其他区域。这一可视化凸显了自适应调优在平衡灵敏度和特异度方面的优势，并优于固定全局调优方法。

族	LARGE	EBIC	RIC	曲线下面积 StARS	五折交叉验证	TIGER	PC-GLASSO
第1带							
$p = 100$							
$n = 300$	0.99 (0.01)	0.75 (0.06)	0.74 (0.02)	0.84 (0)	0.84 (0.02)	0.97 (0.01)	**1 (0)**
$n = 500$	**1 (0)**	0.84 (0)	0.82 (0.01)	0.84 (0)	0.90 (0.08)	**1 (0)**	**1 (0)**
$p = 300$							
$n = 300$	**0.98 (0.01)**	0.70 (0.01)	0.68 (0.01)	0.83（0）	0.83（0）	0.93（0.01）	0.95（0.01）
$n = 500$	**1（0）**	0.83 (0)	0.77 (0.02)	0.83 (0)	0.83 (0)	**1 (0)**	**1 (0)**
Band-2							
$p = 100$							
$n = 300$	**0.90 (0.01)**	0.65 (0.01)	0.61 (0.01)	0.79 (0.02)	0.83 (0.01)	0.61 (0.01)	0.75 (0.01)
$n = 500$	0.96 (0.01)	0.67 (0.01)	0.67 (0.01)	0.81 (0.01)	0.84 (0)	0.76 (0.01)	**0.99 (0)**
$p = 300$							
$n = 300$	**0.86 (0.01)**	0.63 (0.01)	0.57 (0.01)	0.80 (0.01)	0.79 (0.02)	0.55 (0)	0.50 (0.01)
$n = 500$	**0.94 (0.01)**	0.66 (0.01)	0.62 (0.01)	0.82 (0)	0.83 (0)	0.69 (0.01)	0.75 (0.01)
区块							
$p = 100$							
$n = 300$	**0.87 (0.05)**	0.66 (0.07)	0.62 (0.06)	0.78 (0.06)	0.71 (0.11)	0.61 (0.05)	0.65 (0.15)
$n = 500$	**0.92 (0.04)**	0.67 (0.05)	0.66（0.05）	0.79（0.06）	0.76（0.1）	0.81（0.06）	0.73（0.12）
$p = 300$							
$n = 300$	**0.86 (0.02)**	0.64 (0.02)	0.59 (0.02)	0.78 (0.03)	0.73 (0.05)	0.55 (0.01)	0.50 (0)
$n = 500$	**0.92 (0.01)**	0.66 (0.02)	0.65 (0.02)	0.82 (0.02)	0.81 (0.04)	0.73 (0.03)	0.80 (0.02)
Hub							
$p = 100$							
$n = 300$	**0.80 (0.02)**	0.60 (0.01)	0.58 (0.01)	0.71 (0.01)	0.76 (0.03)	0.56 (0.01)	0.63 (0.05)
$n = 500$	0.84 (0.01)	0.61 (0.01)	0.62 (0.01)	0.74 (0.01)	0.79 (0.02)	0.66 (0.01)	**0.87 (0.01)**
$p = 300$							
$n = 300$	**0.76 (0.01)**	0.59 (0.01)	0.54 (0.01)	0.72 (0.01)	0.71 (0.03)	0.52 (0)	0.50 (0)
$n = 500$	**0.81（0.01）**	0.60（0.01）	0.59（0.01）	0.75（0.01）	0.76（0.02）	0.60 (0.01)	0.64 (0.01)
随机							
$p = 100$							
$n = 300$	**0.94 (0.02)**	0.60 (0.03)	0.58 (0.02)	0.70 (0.03)	0.70 (0.08)	0.61 (0.03)	0.74 (0.12)
$n = 500$	**0.96 (0.02)**	0.61 (0.03)	0.64 (0.03)	0.70 (0.03)	0.77 (0.11)	0.82 (0.03)	0.80 (0.09)
$p = 300$							
$n = 300$	**0.93 (0.01)**	0.61 (0.02)	0.56（0.01）	0.72（0.01）	0.72（0.03）	0.61（0.01）	0.53（0.07）
$n = 500$	**0.95 (0.01)**	0.63 (0.01)	0.62 (0.01)	0.73 (0.01)	0.86 (0.06)	0.83 (0.01)	0.90（0.01）

表1：*LARGE* 和基准方法在所有模拟设置下的平均曲线下面积。粗体数值表示各设置下的最佳表现。括号内为标准差。结果为 50 次重复实验的均值。



ˆ 图3：估计的精度矩阵 Θ ，针对带宽为2的图模型，以绝对相关性尺度绘制。数据由 $X \overset{iid}{\sim} \mathcal{N}(0, \Theta^{-1})$ 生成，其中$(n, p) = (500, 100)$。在第二行中，依赖全局调优参数的方法（扩展贝叶斯信息准则、RIC、StARS 和交叉验证）在存在异质局部噪声水平时，往往会对网络的不同区域估计不足或过度。

4.2 估计

图估计性能的结果见表2。LARGE 的非对角均方根误差与表现最佳的基准方法 PC-GLASSO 相当。与图选择结果一致，在更具挑战性的数据生成过程中，例如 $(n, p) = (300, 300)$，LARGE 往往优于 PC-GLASSO。我们还观察到，能够适应局部噪声水平的方法（如 LARGE 和 PC-GLASSO），或对调优参数不敏感的方法（如 TIGER），通常比依赖单一全局λ的方法具有更高的估计准确性。

4.3 收敛

我们使用相对弗罗贝尼乌斯范数监测 LARGE 的收敛情况：

$$\frac{\|W_{\text{new}} - W_{\text{old}}\|_F}{\|W_{\text{old}}\|_F} < e. \tag{4.2}$$

在实践中， LARGE 通常在 20 次迭代内收敛。我们在带宽为2的情形中观察到一次未收敛，在枢纽情形中观察到三次未收敛。为确保在这些更具挑战性的数据生成过程中收敛，我们将 $e = 0.005$ 用作 $p = 100$的收敛阈值，并将 $e = 0.05$放宽至 $p = 300$。尽管这一调整使算法成功收敛，但也凸显了一项实际局限：用户可能需要根据问题的维度和结构复杂性放宽收敛阈值。

族	LARGE	EBIC	RIC	均方根误差$_{off}$ StARS	五折交叉验证	TIGER	PC-GLASSO
第1带							
$p = 100$							
$n = 300$	0.46 (0.05)	1 (0)	1 (0)	0.99 (0)	0.99 (0.03)	0.84 (0.01)	**0.31 (0.05)**
$n = 500$	0.32 (0.03)	0.99 (0)	0.99 (0)	0.99 (0)	0.71 (0.34)	0.82 (0.01)	**0.23 (0.11)**
$p = 300$							
$n = 300$	**0.56 (0.03)**	1 (0)	1 (0)	0.99 (0)	0.99 (0)	0.86 (0.01)	0.77 (0.01)
$n = 500$	0.38 (0.02)	0.99 (0)	1 (0)	0.99 (0)	0.99 (0)	0.83 (0)	**0.29 (0.02)**
Band-2							
$p = 100$							
$n = 300$	**0.72 (0.04)**	1 (0)	1 (0)	1 (0)	0.99 (0)	0.96 (0.01)	0.82 (0.02)
$n = 500$	0.59 (0.04)	1 (0)	1 (0)	0.99 (0)	0.99 (0)	0.91 (0.01)	**0.3 (0.03)**
$p = 300$							
$n = 300$	**0.88 (0.01)**	1 (0)	1 (0)	1 (0)	1 (0)	0.98 (0)	1 (0)
$n = 500$	**0.70 (0.02)**	1 (0)	1 (0)	0.99 (0)	0.99 (0)	0.94 (0)	0.86 (0.01)
区块							
$p = 100$							
$n = 300$	0.90 (0.08)	1 (0)	1 (0)	1 (0)	1 (0)	0.97 (0.03)	**0.89 (0.12)**
$n = 500$	**0.79 (0.12)**	1 (0)	1 (0)	0.99 (0)	1 (0)	0.91 (0.03)	0.86 (0.1)
$p = 300$							
$n = 300$	**0.91 (0.03)**	1 (0)	1 (0)	1 (0)	1 (0)	0.98 (0.01)	1 (0.01)
$n = 2000$	**0.81 (0.04)**	1 (0)	1 (0)	0.99 (0)	0.99 (0)	0.93 (0.01)	**0.81 (0.04)**
Hub							
$p = 100$							
$n = 300$	**0.84 (0.06)**	1 (0)	1 (0)	1 (0)	0.99 (0)	0.98 (0.01)	0.92 (0.03)
$n = 500$	0.77 (0.03)	1 (0)	1 (0)	1 (0)	0.99 (0.02)	0.95 (0.01)	**0.5 (0.03)**
$p = 300$							
$n = 300$	**0.95 (0.01)**	1 (0)	1 (0)	1 (0)	1 (0)	0.99 (0)	1 (0)
$n = 500$	**0.88 (0.01)**	1 (0)	1 (0)	1 (0)	0.99 (0)	0.97 (0)	0.93 (0.01)
随机							
$p = 100$							
$n = 300$	0.87 (0.09)	1 (0)	1 (0)	0.99 (0)	0.99 (0.02)	0.94 (0.03)	**0.82 (0.13)**
$n = 500$	0.76 (0.08)	1 (0)	1 (0)	0.99 (0)	0.96 (0.05)	0.85 (0.03)	**0.74 (0.21)**
$p = 300$							
$n = 300$	**0.85 (0.02)**	1 (0)	1 (0)	0.99 (0)	0.99 (0)	0.94 (0.01)	0.99 (0.02)
$n = 500$	0.77 (0.03)	1 (0)	1 (0)	0.99 (0)	0.93 (0.04)	0.83 (0.01)	**0.64 (0.04)**

表2：*LARGE* 和基准方法在所有模拟设置下的非对角均方根误差均值。粗体数值表示各设置下的最佳表现。括号内为标准差。结果取自50次重复实验的均值。

5 数据分析

功能连接（FC）指通过功能性磁共振成像（fMRI）测得的区域血氧水平依赖（BOLD）信号之间的时间依赖模式。为估计功能连接，首先将大脑划分为具有明确解剖定义的区域。对于每个区域，功能性磁共振成像扫描会记录随时间变化的 BOLD 信号，该信号反映每个时间点的神经活动水平。功能连接与认知功能相关，并常用于预测健康个体的认知表现 [Seeley et al., 2007; van den Heuvel et al., 2009]。

传统上，功能连接通过皮尔逊相关来计算，以量化区域两两之间的线性关联 [Li et al., 2019; He et al., 2020; Dhamala et al., 2021]。不过，也有研究探索其他度量方法，例如经验协方差和蒂霍诺夫精度矩阵 [Pervaiz et al., 2020]。在本分析中，我们使用 LARGE 将功能连接网络估计并可视化为高斯图模型。此处使用的数据集是来自人类连接组计划（HCP）青年成人 S1200 版本的公开高分辨率预处理功能性磁共振成像数据 [Essen et al., 2013]。我们分析了 1003 名健康成年人在四次完整静息态功能磁共振成像扫描中、同一时间范围内观测到的时间序列数据，每次扫描包含 $n = 1200$ 个时间点。

预处理：磁共振成像是在圣路易斯华盛顿大学使用西门子 Skyra 3T 扫描仪采集的。每位受试者均接受了四次梯度回波平面回波成像静息态功能性磁共振成像（静息态功能磁共振成像）扫描，TR 为 $= 720$ ms，TE 为 $= 33.1$ ms，体素尺寸为 2.0 mm 各向同性，视野为 $= 208 \times 180$ mm^2, 翻转角为 $= 52°$，矩阵为 $= 104 \times 90$，共 72 层。在两个扫描时段中，每个时段均采集了两次 15 分钟的静息态功能磁共振成像扫描。每次扫描包含 1200 个时间点，因此每位受试者的四次扫描总计包含 4800 个时间点。HCP 联盟 [Smith 等人（ 2011]）对每位受试者每次静息态功能磁共振成像扫描的数据进行了预处理。数据经过最小限度的预处理 [Glasser 等人（ 2013] ），并使用 ICA+FIX [Griffanti 等人，2014；Salimi-Khorshidi 等人（ 2014]）去除了伪影。我们采用 Jamison 等人（[2024], ）所述的标准后处理流程，识别运动和全局信号异常时间点，回归去除组织特异性及运动相关干扰时间序列，并进行时间滤波。异常时间点在干扰回归和时间滤波中均予以排除。使用 FreeSurfer [Fischl 等人（ 1999], ）将灰质划分为 86 个皮层、皮层下及小脑区域，并将各区域内所有体素的时间序列均值计算为该区域的时间序列。

结果：对于随机选取的一名受试者，我们首先通过求样本协方差矩阵的逆，绘制其第一次功能性磁共振成像扫描（$n = 1200$ 个时间点）得到的样本偏方差直方图。如图4所示，偏方差存在显著差异，这表明在估计功能连接时，采用局部自适应节点惩罚可能会有所帮助。

为了直观呈现并解读功能连接网络，我们使用 LARGE 和基准方法，根据每次扫描中 86 个脑区的功能磁共振成像时间序列估计精度矩阵。由于得到的偏方差相对较小，我们对精度矩阵的对角元素进行惩罚，以提高数值稳定性。随后，我们对四次扫描得到的精度矩阵取平均，并在图5中展示最终网络。我们不将 RIC 纳入此次基准方法比较，因为它得到的估计过于稀疏，难以解读。

所有方法似乎都能有效捕捉潜在的稀疏结构。它们估计的精度矩阵保留了已知的生理模式，其中最显著的是双侧同源区域之间的强连接（即左右半球中相对应的区域），这一点已在文献中得到充分证实[Zuo et al., 2010]。此外，我们还观察到一种将左右大脑皮层半球分隔开的模块化块结构，以及一个明显的皮层下簇 [Power et al., 2011]。在这些方法中，LARGE 产生的解最为稀疏（密度为 $= 31\%$），这通常会抑制较弱的连接，并可能增强可解释性。相比之下，CV、扩展贝叶斯信息准则和 StARS 产生更稠密的网络（密度为 $> 80\%$），可能会引入额外噪声。

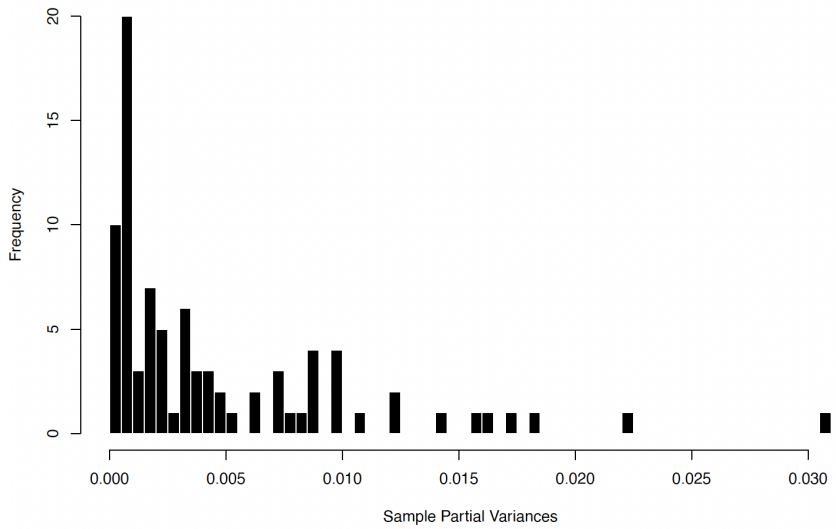


图4： 一次功能性磁共振成像扫描中，*86* 个脑区在 $n = 1200$ 个时间点的样本偏方差分布。

6 结论

本文提出了一种用于估计稀疏精度矩阵的自适应框架，该框架针对逆协方差矩阵的每一列，学习逐节点且尺度适当的 ℓ_1-惩罚项。我们的方法在分块坐标下降算法中加入了一个内循环，用于动态跟踪节点噪声方差估计，进而据此推导局部自适应调优参数。为进一步增强边恢复能力，我们引入了一种变量选择策略，引导 Lasso路径优先考察边际相关性较高的节点对，从而能够在图中更早地检测出真实边。总体而言，我们的方法实现了一个概念转变：从全局调优且对尺度敏感的正则化 λ，转向通过用户指定的显著性水平 α，以更具可解释性、基于假设检验的方式控制稀疏性。

我们的方法展现出良好的实证性能，也由此引出了若干令人振奋的研究方向。在算法方面，如何评估基于 ℓ_1 的方法是否适用，仍是一个有待解决的问题，例如，可以通过拟合优度检验来评估真实图是否近似稀疏。在计算方面，若能更高效地实现关键步骤，算法将会受益，例如，按偏残差的标准差对变量排序，或执行用于变量选择的序贯 F 检验。在理论方面，对内循环中具有双凸目标函数的块坐标下降算法进行严格的收敛性分析，将有助于进一步明确所提方法的适用范围及其理论保证。这些重要问题留待未来研究。

致谢

SB感谢美国国立卫生研究院项目R01GM135926以及美国国家科学基金会项目DMS-1812128、DMS-2210675和CAREER DMS-2239102提供的部分资助。HN和SB感谢Amy Kuceyeski博士和Keith Jamison博士就神经科学应用的分析与解读提供的深刻见解。

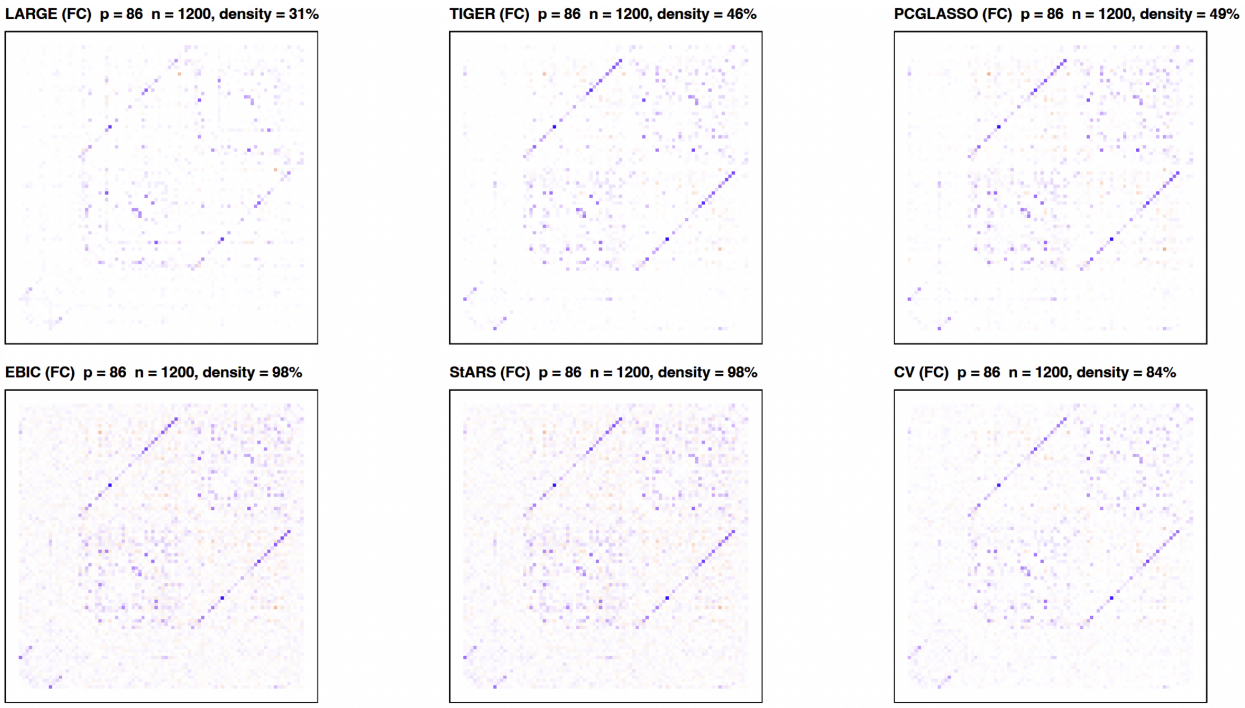


图5：单个个体的功能连接网络由 *LARGE* 和基准方法估计，并以相关系数尺度进行可视化。所有方法均恢复了已知的大脑连接生物学模式。

参考文献

O. Banerjee、L. El Ghaoui和A. d'Aspremont。通过稀疏最大似然估计对多元高斯数据或二元数据进行模型选择。机器学习研究杂志，9：485–516，2008年6月。ISSN 1532-4435。R. M. Birn、M. D. Cornejo、E. K. Molloy、R. Patriat、T. B. Meier、G. R. Kirk、V. A. Nair、M. E. Meyerand和V. Prabhakaran。生理噪声校正对静息态功能连接重测信度的影响。Brain Connectivity，4(7)：511–522，2014年9月。数字对象唯一标识符：10.1089/brain.2014.0284。网址：https://doi.org/10.1089/brain.2014.0284。J. S. Carter、D. Rossell和J. Q. Smith。偏相关图形套索。《斯堪的纳维亚统计学杂志》，2023。G. Costantini、J. Richetin、E. Preti、E. Casini、S. Epskamp和M. Perugini。人格网络的稳定性与变异性：网络心理测量学最新进展教程。《人格与个体差异》，136：68–78，2019。ISSN 0191-8869。数字对象唯一标识符：https://doi.org/10.1016/j.paid.2017.06.011。网址：https://www.sciencedirect.com/science/article/pii/S0191886917304002。动态人格心理学。A. P. Dempster。协方差选择。《生物计量学》，28(1)：157–175，1972年。数字对象唯一标识符：10.2307/2528966。网址：https://doi.org/10.2307/2528966。E. Dhamala、K. W. Jamison、A. Jaywant、S. Dennis和A. Kuceyeski。不同的功能连接和结构连接可预测健康成年人的晶体智力与流体认知。《人脑图谱》，42(10)：3102–3118，2021年7月。数字对象唯一标识符：10.1002/hbm.25420。2021年4月8日在线发表。D. C. V. Essen、S. M. Smith、D. M. Barch、T. E. Behrens、E. Yacoub和K. Ugurbil。华盛顿大学—明尼苏达大学人类连接组计划：概述。NeuroImage，80：62–79，2013。ISSN 1053-8119。

数字对象唯一标识符：10.1016/j.neuroimage.2013.05.041。网址 https://www.sciencedirect.com/science/article/pii/ S1053811913005351。J. Fan 和 J. Lv。超高维特征空间的确定性独立筛选。英国皇家统计学会会刊：B辑（统计方法学），70(5):849–911，2008年。数字对象唯一标识符：10.1111/j.1467-9868. 2008.00674.x。J. Fan、J. Lv 和 L. Qi。经济学中的稀疏高维模型。经济学年度评论，3: 291–317，2011。数字对象唯一标识符：10.1146/annurev-economics-061109-080451。B. Fischl、M. I. Sereno 和 A. M. Dale。基于皮层表面的分析：II. 膨胀、展平和基于表面的坐标系统。NeuroImage，9(2):195–207，1999。ISSN 1053-8119。数字对象唯一标识符：10.1006/nimg.1998.0396。R. Foygel 和 M. Drton。高斯图模型的扩展贝叶斯信息准则。神经信息处理系统进展，23:2020–2028，2010。J. Friedman、T. Hastie 和 R. Tibshirani。利用图形 Lasso 进行稀疏逆协方差估计。生物统计学，9(3):432–441，2008 年 7 月。数字对象唯一标识符：10.1093/biostatistics/kxm045。网址 https://doi.org/10. 1093/biostatistics/kxm045。J. Friedman、T. Hastie 和 R. Tibshirani。glasso：图形 Lasso——高斯图模型估计，2019。网址 https://CRAN.R-project.org/package=glasso。R软件包版本 1.11。M. Galloway。套索惩罚精度矩阵估计，2018。网址 https://github.com/MGallow/ CVglasso。R软件包版本 1.0。M. F. Glasser、S. N. Sotiropoulos、J. A. Wilson、T. S. Coalson、B. Fischl、J. L. Andersson、J. Xu、S. Jbabdi、M. Webster、J. R. Polimeni 等。人类连接组计划的最简预处理流程。NeuroImage，80:105–124，2013。L. Griffanti、G. Salimi-Khorshidi、C. F. Beckmann、E. J. Auerbach、G. Douaud、C. E. Sexton、E. Zsoldos、K. P. Ebmeier、N. Filippini、C. E. Mackay 等。基于ICA的伪影去除及加速功能磁共振成像采集，以改进静息态网络成像。NeuroImage，95:232–247，2014。T. Hastie、R. Tibshirani 和 M. Wainwright。《稀疏统计学习：套索及其推广》，第2章，第25–54页。Chapman and Hall/CRC，纽约，2015a。T. Hastie、R. Tibshirani 和 M. Wainwright。《稀疏统计学习：套索及其推广》。CRC出版社，博卡拉顿，佛罗里达州，2015b。T. He、R. Kong、A. J. Holmes、M. Nguyen、M. R. Sabuncu、S. B. Eickhoff 和 B. T. T. Yeo。深度神经网络和核回归在功能连接预测行为和人口统计学特征方面的准确度相当。NeuroImage，206:116276，2020。数字对象唯一标识符：10.1016/j.neuroimage.2019.116276。K. Jamison、Z. Gu、Q. Wang、C. Tozlu、M. Sabuncu 和 A. Kuceyeski。推出 krakencoder：一个统一的脑连接组转换与融合工具。bioRxiv，页码 2024.04.12.589274，2024年10月8日。数字对象唯一标识符：10.1101/2024.04.12.589274。预印本。J. Jankov´a 和 S. van de Geer。高维图模型中的推断。载于《图模型手册》，第325–350页。CRC出版社，2018。H. Jiang、X. Fei、H. Liu、K. Roeder、J. Lafferty、L. Wasserman、X. Li 和 T. Zhao。huge：高维无向图估计，2021。网址 https://CRAN.R-project.org/package=huge。R软件包版本 1.3.5。J. Li、R. Kong、R. Li´egeois、C. Orban、Y. R. Tan、N. B. Sun 和 B. T. T. Yeo。全局信号回归增强了静息态功能连接与行为之间的关联。NeuroImage，196: 126–141，2019。数字对象唯一标识符：10.1016/j.neuroimage.2019.04.016。

H. Liu 和 L. Wang。TIGER：一种对调参不敏感的高斯图模型最优估计方法。Electronic Journal of Statistics，11(1):241–294，2017。数字对象唯一标识符：10.1214/16-EJS1195。H. Liu、K. Roeder 和 L. Wasserman。针对高维图模型的正则化选择稳定性方法（StARS）。神经信息处理系统进展，24(2):1432–1440，2010。T. T. Liu。功能性磁共振成像信号中的噪声贡献：综述。NeuroImage，143:141–151，2016年12月。数字对象唯一标识符：10.1016/j.neuroimage.2016.09.008。网址 https://doi.org/10.1016/j.neuroimage.2016.09.008。在线发表于2016年9月6日。S. Lysen。置换包含准则：一种变量选择技术。博士论文，宾夕法尼亚大学，2009。S. Marek、B. Tervo-Clemmens、F. J. Calabro 等。可重复的全脑关联研究需要数千名参与者。Nature，603(7902):654–660，2022。数字对象唯一标识符：10.1038/s41586-022-04492-9。N. Meinshausen 和 P. B¨uhlmann。利用套索进行高维图和变量选择。Annals of Statistics，34(3):1436–1462，2006。数字对象唯一标识符：10.1214/009053606000000281。D. C. Montgomery。《统计质量控制导论》。John Wiley & Sons，第8版，2020。参见第4章。U. Pervaiz、D. Vidaurre、M. W. Woolrich 和 S. M. Smith。优化功能磁共振成像的网络建模方法。NeuroImage，211:116604，2020年5月。数字对象唯一标识符：10.1016/j.neuroimage.2020.116604。在线发表于2020年2月13日。J. D. Power、A. L. Cohen、S. M. Nelson、G. S. Wig、K. A. Barnes、J. A. Church、A. C. Vogel、T. O. Laumann、F. M. Miezin、B. L. Schlaggar 和 S. E. Petersen。人脑的功能网络组织。Neuron，72(4):665–678，2011。数字对象唯一标识符：10.1016/j.neuron.2011.09.006。网址 https://doi.org/10.1016/j. neuron.2011.09.006。Z. Ren、T. Sun、C.-H. Zhang 和 H. H. Zhou。大型高斯图模型估计中的渐近正态性与最优性。Annals of Statistics，43(3):991–1026，2015年6月。数字对象唯一标识符：10.1214/14-AOS1286。T. Sadhukhan、I. Wilms、S. Smeekes 和 S. Basu。Autotune：快速、准确且自动的套索调节参数选择。arXiv预印本 arXiv:2512.11139，2025。数字对象唯一标识符：10.48550/arXiv.2512.11139。G. Salimi-Khorshidi、G. Douaud、C. F. Beckmann、M. F. Glasser、L. Griffanti 和 S. M. Smith。功能性磁共振成像数据的自动去噪：结合独立成分分析与分类器的层次融合。NeuroImage，90:449–468，2014。W. W. Seeley、V. Menon、A. F. Schatzberg、J. Keller、G. H. Glover、H. Kenna、A. L. Reiss 和 M. D. Greicius。用于显著性处理和执行控制的可分离内在连接网络。Journal of Neuroscience，27(9):2349–2356，2007。数字对象唯一标识符：10.1523/JNEUROSCI.5587-06.2007。S. M. Smith、K. L. Miller、G. Salimi-Khorshidi、M. Webster、C. F. Beckmann、T. E. Nichols、J. D. Ramsey 和 M. W. Woolrich。功能磁共振成像的网络建模方法。NeuroImage，54(2):875–891，2011。ISSN 1053-8119。数字对象唯一标识符：10.1016/j.neuroimage.2010.08.063。T. Sun 和 C.-H. Zhang。利用缩放 Lasso 进行稀疏矩阵求逆。Journal of Machine Learning Research，14(1):3385–3418，2013年1月。ISSN 1532-4435。M. P. van den Heuvel、C. J. Stam、R. S. Kahn 和 H. E. Hulshoff Pol。功能脑网络的效率与智力表现。Journal of Neuroscience，29(23):7619–7624，2009年6月。数字对象唯一标识符：10.1523/JNEUROSCI.1443-09.2009。Y. R. Wang 和 H. Huang。利用表达数据重建基因网络的统计方法综述。Journal of Theoretical Biology，362:53–61，2014。ISSN 0022-5193。数字对象唯一标识符：https://doi.org/10.1016/j.jtbi.2014.03.040。网址 https://www.sciencedirect.com/science/article/pii/S0022519314001969。复杂疾病的基于网络的生物标志物。

M. Yuan 和 Y. Lin。高斯图模型中的模型选择与估计。Biometrika，94(1)：19–35，2007。数字对象唯一标识符：10.1093/biomet/asm018。Y. Zhu 和 I. Cribben。功能连接网络的稀疏图模型：最佳方法与自相关问题。Brain Connectivity，8(3)：139–165，2018 年 4 月。ISSN 2158-0014。数字对象唯一标识符：10.1089/ brain.2017.0511。网址：https://www.liebertpub.com/doi/10.1089/brain.2017.0511。出版商：Mary Ann Liebert, Inc., publishers。X.-N. Zuo、C. Kelly、A. Di Martino、M. Mennes、D. S. Margulies、S. Bangaru、R. Grzadzinski、A. C. Evans、Y.-F. Zang、F. X. Castellanos 等。共同成长与渐行渐远：功能同伦的生命周期发展轨迹中的区域与性别差异。Journal of Neuroscience，30(45)：15034–15043，2010。

A 附录

A.1 带对角惩罚的 LARGE

本节中，我们介绍 LARGE 的另一种表述和算法，其中 Θ 的对角线元素和非对角线元素均受到惩罚。

$$\hat{\Theta} := \arg\min_{\Theta \in \mathbb{S}^p_{++}} \left\{ -\log\det(\Theta) + \mathrm{tr}(S\Theta) + \sum_{j=1}^{p} \lambda_j \|\Theta_{\cdot j}\|_1 \right\}, \tag{A.1}$$

其中 $\|\Theta_{\cdot j}\|_1 = \sum_{k=1}^{p} |\Theta_{kj}|$。

算法 3 带对角惩罚的 LARGE

输入： 数据矩阵 $\mathcal{X} \in \mathbb{R}^{n\times p}$，序贯 F 检验显著性水平 α

1: 对于 $j = 1,\ldots,p$，初始化 $\hat{\sigma}_j^2 \leftarrow$ 方差($\mathcal{X}_{\cdot j}$)，$\lambda_{0j} \leftarrow \frac{1}{\text{方差}(\mathcal{X}_{\cdot j})}$ 最大值$_{k\neq j} \left|\frac{1}{2n}\langle \mathcal{X}_{\cdot j},\ \mathcal{X}_{\cdot k}\rangle\right|$

2: 令 $S \leftarrow$ 协方差($\mathcal{X}$)；初始化 $W \leftarrow S$

3: **重复** 对于 $j = 1, 2, \ldots, p$:

4: (a) 划分 $W = \begin{bmatrix} W_{11} & w_{12} \\ w_{12}^\top & w_{22} \end{bmatrix}$，$S = \begin{bmatrix} S_{11} & s_{12} \\ s_{12}^\top & s_{22} \end{bmatrix}$.

5: (b) 使用 Autotune Lasso 求解 $W_{11}\beta - s_{12} + \lambda_j \cdot \mathrm{sign}(\beta) = 0$：

6: 令 $V \leftarrow W_{11}$；初始化 $\hat{\beta} \leftarrow 0$

7: 初始化预测变量排序 $_j \leftarrow \{1,\ldots,p-1\}$，支持集 $_j \leftarrow \emptyset$，sigma 更新标志 $_j \leftarrow$ TRUE

8: **当** 误差 $\geq$ 误差容差 **时**

9: 支持集 $_j^{(\text{旧})} \leftarrow$ 支持集$_j$，$_\beta^{\hat{}\beta^{(\text{old})} \leftarrow \hat{}}$

10: **对于** k in Predictor.Ranking$_j$ **执行**

11: $\hat{}\beta_k \leftarrow$ 软阈值$_{\lambda_j = \hat{\sigma}_j^2 \cdot \lambda_{0j}} \left(s_{12k} - \sum_{l\neq k} V_{kl} \cdot \hat{}\beta_l\right) / (V_{kk} + \lambda_k)$

12: **结束循环**

13: **如果** sigma.update.flag$_j$ ==TRUE **则**

14: Predictor.Ranking$_j$ $\leftarrow$按以下内容降序 k 排列：

$$\begin{cases} |\mathrm{Cor}(\mathcal{X}_{\cdot j}, \mathcal{X}_{\cdot k})|, & \text{if first iteration of line 8} \\ \text{SD of } r_k = \mathcal{X}_{\cdot j} - \sum_{l\notin\{j,k\}} \mathcal{X}_{\cdot l}\hat{\beta}_l, & \text{otherwise} \end{cases}$$

15: 使用序贯F检验拟合线性模型进行变量选择，截断值为 $= F_{\alpha;1,\ n-i}$

16: 支持集 $_j \leftarrow \{k : \mathcal{X}_{\cdot k}$ 已选入线性模型$\}$

17: $\hat{\sigma}_j^2 \leftarrow \frac{\sum_{i=1}^n (X_{ij} - \hat{}X_{ij})^2}{n - |\texttt{Support.Set}_j|}$

18: **如果**支持集$_j \subseteq$支持集$_j^{(\text{old})}$ **则**

19: sigma 更新标志$_j$ $\leftarrow$FALSE

20: **结束如果**

21: **结束如果**

22: **结束循环**

23: (c) $w_{12} \leftarrow (W_{11} + \mathrm{diag}(\lambda_{-j})\mathrm{I})\ \hat{\beta}$ 并进行对称化

24: **直至**所有 j 均收敛

25: 对于 $i = 1,\ldots,p$，$\hat{\Theta}_{jj} \leftarrow (\hat{}\sigma_j + \lambda_j)^{-2}$ 和 $\hat{\Theta}_{j,-j} \leftarrow -\hat{\beta}\hat{\Theta}_{jj}$ 经对称化

输出： $\hat{\Theta}, (\hat{}\sigma_j^2, \lambda_j)$，其中 $j = 1,\ \ldots,\ p$