arXiv:2603.00955v1 [stat.ME] 1 Mar 2026

超越错误发现率：用于分组变量选择的逐步降阶组 SLOPE 方法

Xuelin Zhang[1], Jingxuan Liang[1], Xinyue Liu[1], Hong Chen[1], Biqin Song[1]

[1]华中农业大学信息学院，中国湖北省武汉市，430070。

贡献作者：zhangxuelin@webmail.hzau.edu.cn；
1908103270@qq.com；lxy36@webmail.hzau.edu.cn；
chenh@mail.hzau.edu.cn；biqin.song@mail.hzau.edu.cn；

摘要

高维特征选择通常需要在统计功效与对多重错误指标（如 k-族错误率（k-FWER）和假发现比例（FDP））的严格控制之间取得平衡。然而，一些现有框架仅着眼于控制期望错误发现率（FDR），且无法利用协变量的组结构，例如排序 L1 惩罚估计（SLOPE）。我们提出分组逐步 SLOPE，这是一种统一优化过程，能够将 Lehmann–Romano 逐步降阶规则嵌入 SLOPE，从而在 k-FWER 和 FDP 阈值下实现有限样本保证。具体而言，我们推导出适用于正交设计的闭式正则化序列，可在理论上将 k-FWER 和 FDP 控制在用户指定的水平；并通过 gk-SLOPE 和 gF-SLOPE 将这些结果推广至分组情形，以控制相应的组级错误 gk-FWER 和 gFDP。对于非正交一般设计，我们提供一种受高斯近似和蒙特卡罗校正启发、经过校准的数据驱动序列，同时保留凸性和可扩展性。我们针对稀疏、相关及组结构情形开展了广泛模拟。实证结果证实了我们的理论发现：所提方法能够实现名义错误控制，且检验效能显著高于竞争性的逐步下降程序，从而验证了这些理论进展的实际价值。

关键词：受控特征选择，逐步下降程序，组 SLOPE，组结构

1 引言

在高维信号处理的具体统计应用中，只有一部分特征对结果预测至关重要；相关变量可以分组，每个子组中的所有变量将被一并选中或舍弃。因此，从稀疏数据中识别关键特征，对于提高计算效率并提供可解释的预测至关重要。 [1, 2]针对单个回归变量或组级别的特征选择，旨在从高维实测数据中识别信息特征或特征子集，这是机器学习领域的一个重要研究方向。典型的特征选择方法包括稀疏线性模型（例如 Lasso [3], 组 Lasso [4], Sparse Group Lasso [5]）、稀疏加性模型（例如 SpAM [6], Group-SAM [7], SpMAM [8] 和 NAMs [9–11]）、基于树的模型（例如随机森林 [12]）以及稀疏神经网络（例如 LassoNet [13], Deep feature selection [14],Deep-Pink [15], 多尺度特征选择网络 [16]）。然而，尽管这些方法中的大多数在特征选择方面表现出良好的实证效果，却往往缺乏关于错误发现率（FDR）的理论保证，而这至关重要。

沿着这一研究方向，受控特征选择在保证较低错误发现率（FDR）的同时进一步提升了选择质量。由于其应用广泛，例如在生物信息学和生物医学处理领域，近年来受到了越来越多的关注 [17, 18]。受控特征选择的学习系统主要有三类：多重假设检验 [19–22], 、knockoffs滤波器 [23–26], 以及排序 L1 惩罚估计（SLOPE） [27–29]。作为一种经典的特征选择策略，Benjamini-Hochberg（BH）程序通过联合考虑多重假设检验中的 p 值来构建 [19]。尽管该程序在错误发现率控制方面具有良好的理论性质，但在非线性和复杂的回归估计中可能面临计算挑战 [30]。作为一种新颖的特征筛选方案，knockoff推断具有坚实的理论基础，并在实际应用中展现出具有竞争力的表现 [18, 23, 25, 31]。具体而言， [31] 提出了一种基于错误的 knockoffs 推断框架，从发生 k 个或更多错误拒绝的概率（k-族错误率）以及错误发现比例（FDP）的角度，进一步实现受控特征选择。与借助 knockoff特征筛除活跃特征不同，SLOPE 专注于稀疏特征选择的正则化设计，会自适应地对排序后的 ℓ_1 惩罚项施加一个非递增的调节参数序列 [27, 29, 32]。此方法进一步扩展为利用组级稀疏正则化选择特征子集，称为 g-SLOPE [29]。尽管 SLOPE 的优化算法 [27, 29] 及其理论性质 [28], 已取得迅速进展，但现有关于 SLOPE 和 g-SLOPE 的大多数研究仅限于错误发现率控制。受这一局限性的启发，我们进一步考察分组结构，并提出新的组 SLOPE 方法，以在 k-族错误率和错误发现比例标准下实现受控组级特征选择。此外，我们针对不同配置的正交或高斯设计、不同稀疏程度以及不同数量的相关特征（和组）开展了充分的实验，以验证所提出的新组级特征选择方法的有效性。

为弥补这一空白，我们提出了新的 SLOPE 和 g-SLOPE 方法，即 k-SLOPE、F-SLOPE、gk-SLOPE 和 gF-SLOPE，可实现

表1：算法性质（$\surd$ - 具有给定信息， $\times$ - 不具有给定信息）。

算法 \ 性质	错误发现率[1] 控制	错误发现比例[1] 控制	k-族错误率[1] 控制	组级
SLOPE[27, 28, 33–35]	$\surd$	$\times$	$\times$	$\times$
g-SLOPE [29, 36–38]	$\surd$	$\times$	$\times$	$\surd$
逐步下降法（错误发现比例） [21]	$\times$	$\surd$	$\times$	$\times$
逐步下降法（k-族错误率） [21]	$\times$	$\times$	$\surd$	$\times$
F-SLOPE（我们的方法）	$\surd$	$\surd$	$\times$	$\times$
k-SLOPE（本文方法）	$\surd$	$\times$	$\surd$	$\times$
gF-SLOPE（本文方法）	$\surd$	$\surd$	$\times$	$\surd$
gk-SLOPE（本文方法）	$\surd$	$\times$	$\surd$	$\surd$

针对单个回归变量或分组特征进行特征选择，并分别控制 k-FWER 和错误发现比例。与以往依赖 BH 程序的方法不同， [19], 本文提出的 SLOPE 和 g-SLOPE 方法依赖逐步下降程序， [21], 该程序具有更强的可行性和适应性 [27——29] 如表 1所示。本文的主要贡献概述如下:

- 用于 *k-FWER* 和错误发现比例控制的新型 *SLOPE* 方法。我们以统一的方式整合 SLOPE[27] 与逐步下降程序 [21] ，用于 k-FWER 和错误发现比例控制，并构建相应的凸优化问题。与 [31], 中灵活的 knockoff变量推断类似，我们的方法也可以避免复杂的 p值计算，并且易于实施。
- 用于 *g-SLOPE* 的新方法k，用于 *g-FWER* 和 *gFDP* 控制。 我们将 k-SLOPE 和 F-SLOPE 的思想扩展到如下情形：研究者旨在选择具有分组结构的解释变量的整个子集，而非单个回归变量 [29]。
- 理论保证与实证效果。在正交设计条件下，所提出的 SLOPE 方法和 g-SLOPE 方法能够分别将 k-FWER 和错误发现比例控制在预先指定的水平。大量仿真实验验证了逐步下降 SLOPE 方法在 k-FWER 和错误发现比例控制方面，以及分组逐步 SLOPE 方法在 gk-FWER 和 gFDP 控制方面的有效性与灵活性，从而证实了我们的理论发现。

2 相关工作

为了更好地突出所提方法的新颖性，我们回顾相关的 SLOPE [27]和组 SLOPE [29] 方法，以及 FDR、k-族错误率和错误发现比例标准之间的关系。

SLOPE 和组 SLOPE 方法。SLOPE [27] 可视为 Lasso [3],的一种自然扩展，其回归系数会根据秩进行惩罚。正则化序列 $\{\lambda_i\}$ 的一种显著选择是 BH [19] 临界值 $\lambda_{\mathrm{BH}}(i) = \Phi^{-1}(1 - \frac{iq}{2m})$，其中 $q \in (0, 1)$ 是期望的错误发现率水平，m 是特征数， $\Phi(\cdot)$ 是标准正态分布的累积分布函数。SLOPE 的主要动机是提供

回归估计和错误发现率控制的有限样本保证，其中错误发现率是所有选定预测变量中无关回归变量所占比例的期望值。当 X 是正交矩阵时，理论上，使用 λ_{BH} 的 SLOPE 可在期望水平上控制错误发现率。此外，SLOPE 的一个显著特性是不需要知道稀疏度，却能在广泛的 ℓ_0-稀疏类别中自动实现最优总平方误差。

为提高计算效率，研究者提出了一种基于稀疏半光滑牛顿法的增广拉格朗日方法，用于求解更一般的 SLOPE 模型[39]。此外，为提高 SLOPE 的计算效率，尤其是在估计完整正则化路径时，研究者首次提出了一种基于 Lasso 严格规则的 SLOPE 启发式筛选规则 [40]。Larsson 等人还提出了一种求解 SLOPE 优化问题的新型快速算法，该算法结合了近端梯度下降和近端坐标下降步骤 [41]。除了上述有关算法优化的研究外，还有大量研究探讨了 SLOPE 的性质 [28, 42, 43], 、模型改进 [29, 32, 44, 45]和应用 [46, 47]。Liang 等人 [48] （本文的早期会议版本）通过控制其他指标（如 k-族错误率和错误发现比例）改进了经典 SLOPE，并展示了其经验表现和理论保证。此外，为了对相关特征可分组的变量进行选择，Brzyski 等人 [29] 开发了一种先进的组 SLOPE（g- SLOPE），以实现受控的组级特征选择。然而，据我们所知，此前尚无研究探讨基于组级 SLOPE 的特征选择，并提供 k-族错误率或错误发现比例控制保证。

统计指标：错误发现率、 k-族错误率与错误发现比例。Benjamini 和 Hochberg 提出了 BH 程序，以控制错误发现比例的期望，也称为错误发现率控制[19]。随后，Lehmann 和 Romano 提出了单步程序和逐步下降程序，以确保 k-族错误率控制 [21]。他们还考虑了错误发现比例控制，并给出了两种逐步下降程序，用于在温和条件下控制错误发现比例，无论对 p 值结构是否作依赖性假设。借助逐步下降程序 [21], ，已有多项研究利用 k- 族错误率控制 [22, 31, 49, 50] 和错误发现比例控制 [22, 31, 49, 51]进行特征选择。然而，这些程序大多可能依赖 p值来评估每个特征的重要性或底层结构的有效性。此外，传统的 p值计算通常依赖于大样本渐近理论，而这在高维有限样本中可能不再成立 [24]。

此外，解释 FDR、错误发现比例和 k-族错误率之间的关系至关重要。给定 γ， $\alpha \in (0,1)$，错误发现比例控制意味着在水平 α 下，Prob(FDP $> \gamma$)。回想一下，错误发现比例关注的是

$$\mathrm{Prob}\{\mathrm{FDP} > \gamma\} < \alpha, \tag{1}$$

而 FDR 是 FDP 的期望，即 FDR $= \mathbb{E}$(FDP)。接下来，根据 [27, 28], ，我们可以验证

$$\mathrm{FDR} \le \gamma \mathrm{Prob}\{\mathrm{FDP} \le \gamma\} + \mathrm{Prob}\{\mathrm{FDP} > \gamma\}, \tag{2}$$

and

$$\frac{\mathrm{FDR} - \gamma}{1-\gamma} \le \mathrm{Prob}\{\mathrm{FDP} > \gamma\} \le \frac{\mathrm{FDR}}{\gamma}, \tag{3}$$

其中，最后一个不等式由马尔可夫不等式推出。如果某种方法在 q 水平上控制错误发现率， $\leq q/\gamma$，那么它也能控制错误发现比例 $> \gamma$。反过来，如果错误发现比例受到控制，即 Prob(FDP $> \gamma$) $< \alpha$，那么错误发现率可由 $(1-\gamma)\alpha + \gamma$ 所界定。因此，控制错误发现比例的方法通常也能控制错误发现率 [52]。此外，Farcomeni 指出，与错误发现率控制相比， k-族错误率控制更可取，因为在需要更高检验效能的选择场景中，可以获得更高检验效能的选择结果 [53] 。

3 预备知识

本节回顾本文所涉及的一些必要背景知识，例如判据定义（针对错误发现比例、错误发现率和 k-族错误率）、经典稀疏回归量（包括 Lasso [3], 组 Lasso [54], SLOPE [27]）、组 SLOPE [29] 以及逐步下降程序 [21]。

3.1 问题表述

设 $\mathcal{X} \subset \mathbb{R}^m$ 和 $\mathcal{Y} \subset \mathbb{R}$分别为紧输入空间及其对应的输出空间。考虑样本 $\{(x_i, y_i)\}_{i=1}^n$ ，这些样本独立地取自 $\mathcal{X} \times \mathcal{Y}$上的某个未知分布。记

$$X := (X_1, X_2, \cdots, X_n)^T \subset \mathbb{R}^{n \times m} \tag{4}$$

and

$$y = (y_1, y_2, \cdots, y_n)^T \in \mathbb{R}^n. \tag{5}$$

X 的每个列向量的模长均为 1。以下多元线性回归模型生成输出向量 y:

$$y = X\beta + \epsilon, \tag{6}$$

其中 $\beta \in \mathbb{R}^m$ 表示系数向量，且 $\epsilon \sim N(0, \sigma^2 I_n)$。

在稀疏高维回归中，我们通常假设 β 满足稀疏结构。设 V 为错误选定特征的数量， R 为已识别特征的总数。FDP、FDR 和 k-FWER 分别定义为

$$\mathrm{FDP} = \frac{V}{\max\{R, 1\}}, \quad \mathrm{FDR} = \mathbb{E}(\mathrm{FDP}), \tag{7}$$

and

$$k\text{-FWER} = \mathrm{Prob}\{V \geq k\}. \tag{8}$$

3.2 Lasso 和 SLOPE

Tibshirani 等人提出了 Lasso 模型 [3] ，用于从高维数据中去除无关、冗余或高度相关的特征，从而筛选出其余的重要特征以构建模型。这种方法在

机器学习和统计建模中发挥着重要作用。Lasso回归的数学学习对象定义如下:

$$\arg\min_{\beta\in\mathbb{R}^m}\left\{\frac{1}{2}||y-X\beta||^2+\lambda_L\|\beta\|_1\right\}, \tag{9}$$

其中 β 是回归系数。 λ_L 是稀疏 l_1 范数的固定惩罚系数，用于权衡模型拟合与稀疏效应。正如 Bogdan 等人 λ_L 所发现的，合理设定 [27] 是一项颇具挑战的任务。具体而言，使用交叉验证策略寻找合适的λ_L时，Lasso 在特征选择方面表现不佳，导致高维数据中的错误发现率相对较高。

将（ℓ_1）中的 9惩罚项替换为排序后的 ℓ_1 惩罚项后，Bogdan 等人提出了 SLOPE 方法 [27]，用于高维稀疏情形下的受控特征选择。SLOPE 的学习方案可表述为

$$\arg\min_{\beta\in\mathbb{R}^m}\left\{\frac{1}{2}||y-X\beta||^2+J_\lambda(\beta)\right\}, \tag{10}$$

其中 $J_\lambda(\beta) = \sum_{i=1}^m \lambda_i|\beta|_{(i)}$ 是排序后的 l_1 范数，正则化参数 $\lambda_1 \geq \cdots \geq \lambda_m \geq 0$和回归系数 $|\beta|_{(1)} \geq |\beta|_{(2)} \geq \cdots \geq |\beta|_{(m)}$均为非负且非递减序列。当 $\lambda_1 = \lambda_2 = \cdots = \lambda_m$时，优化方案（10）显然退化为 Lasso（9）。

给定目标水平 q，SLOPE 利用参数序列控制错误发现率

$$\lambda_{\mathrm{BH}} = \{\lambda_{\mathrm{BH}}(1), \lambda_{\mathrm{BH}}(2), \cdots, \lambda_{\mathrm{BH}}(m)\}, \tag{11}$$

with

$$\lambda_{\mathrm{BH}}(i) = \sigma\cdot\Phi^{-1}(1-\frac{iq}{2m}), \tag{12}$$

其中， $\Phi(\cdot)$ 表示正交设计下标准正态分布的累积分布函数。

引理 1 [27] 在线性模型中，采用正交设计 X 且 $\epsilon\sim N(0,\sigma^2 I_n)$ 时，采用正则化参数序列 (10) 的 SLOPE（12）满足错误发现率 $\leq \frac{m_0 q}{m}$，其中 m_0 是真实零假设的数量， q 是目标错误发现率水平。

引理 1 阐明了采用由 BH 程序确定的 λ_{BH} 时，SLOPE 在理论上能够保证控制错误发现率 [19]。本文提出的新型 SLOPE 及 gSLOPE 不仅限于错误发现率控制，还将其扩展至 k-族错误率和错误发现比例控制，具体方法是以逐步下降程序取代 BH 程序 [21]。

从计算可行性的角度来看，SLOPE 的优化目标函数（10）是凸但非光滑的，可通过近端梯度下降算法高效求解 [27]。为完整起见，我们在算法 1中给出了 SLOPE 的计算步骤，该步骤同样适用于 SLOPE 的变体。这里，$J_\lambda = \sum_{i=1}^m \lambda_i|\beta|_{(i)}$ 步长通过回溯线搜索选取，且

算法 1： SLOPE 的加速近端梯度算法（10）

输入：训练集 $X \in \mathbb{R}^{n\times m}$ 和 $y \in \mathbb{R}^n$ 以及参数 $\lambda = (\lambda_1, \lambda_2, \ldots\ldots \lambda_m)$。

初始化：$a^0 \in \mathbb{R}^m$、$b^0 = a^0$ 和 $\theta_0 = 1$。

对于k $= 0,1,\cdots$ **执行**

$b^{k+1} = \text{近端算子}_{t_k J_\lambda}\left(a^k - t_k X^T\left(Xa^k - y\right)\right)$

$\theta_{k+1}^{-1} = \frac{1}{2}\left(1 + \sqrt{1 + 4/\theta_k^2}\right)$

$a^{k+1} = b^{k+1} + \theta_{k+1}\left(\theta_k^{-1} - 1\right)\left(b^{k+1} - b^k\right)$

结束 for

输出：a满足停止准则。

满足 $t_k < 2/||X||^2$ [55, 56]。此外，Bogdan 等人还通过对偶理论推导出了具体的停止准则 [27]。

3.3 组 Lasso 与组 SLOPE

在实际应用中，高维数据中的特征往往以组的形式发挥作用，仅考虑单回归器无法实现有效的特征选择。为解决此类问题，组 Lasso 将特征集划分为若干组，并采用组稀疏正则化 [4, 5, 57]。

与组 Lasso 的概念类似，组 SLOPE（g-SLOPE） [29] 将经典 SLOPE [27] 加以扩展，用以解决希望选择整组解释变量而非单回归器的情形。此类组可由对强相关预测变量进行聚类而形成，也可由对应于同一定性预测变量不同水平的虚拟变量组构成。

假设 $I = \{I_1 \cup I_2 \ldots \cup I_t\}$ 是集合 $\{1, \ldots, n\}$的某种划分，即 I_i均为非空集合，$I_i \cap I_j = \emptyset$ 对于 $i \neq j$。具有 t 个组的线性回归数据生成机制为

$$y = \sum_{i=1}^{t} X_{I_i}\beta_{I_i} + \epsilon, \tag{13}$$

其中 X_{I_i} 是由索引为 I_i 的列组成的 X 的子矩阵，β_{I_i} 是 β 在集合 I_i中索引上的限制。此外，X 可以是任意矩阵，尤其允许子矩阵 X_{I_i} 内部存在任意线性关系。假设$\epsilon \sim N(0, \sigma^2 I_n)$，其中 σ 表示标准差。令记号 $l_1, \ldots, l_t$表示子矩阵 $X_{I_1}, \ldots, X_{I_t}$的秩。

组 Lasso 的常用形式定义为

$$\underset{\beta\in\mathbb{R}^m}{\arg\min}\left\{\frac{1}{2}||y - X\beta||^2 + \sigma\lambda_{GL}\sum_{i=1}^{t}\sqrt{|I_i|}\,\|\beta_{I_i}\|_2\right\}, \tag{14}$$

其中，λ_{GL} 是组 Lasso 惩罚项的固定系数。$|I_i|$表示集合 I_i中的元素个数，X_{I_i} 是由 X 中索引为 I_i 的列组成的子矩阵，而β_{I_i} 是回归系数 β 在索引 I_i上的限制。组 Lasso 的局限性与 Lasso 类似，包括在高维

数据中错误发现率较高。因此，将 SLOPE 的概念与分组特征选择相结合至关重要。

类似于组 Lasso 的定义和 SLOPE 的概念。定义排序后的 ℓ_1-诱导惩罚 J_λ 为 $J_\lambda(v) = \sum_{i=1}^{t} \lambda_i v_{(i)}$，其中 $v_{(1)} \geq v_{(2)} \geq \cdots \geq v_{(t)}$。则 g-SLOPE [29] 的形式可表述为

$$\beta^{gS} \in \arg\min_{\beta \in \mathbb{R}^m} \frac{1}{2}||y - X\beta||^2 + \sigma J_\lambda(W\|\beta\|_{X,I}), \tag{15}$$

其中 λ_1，$\cdots$，λ_m 是给定的非负调节参数的非递增序列。稀疏性诱导惩罚 $J_\lambda(\beta) = \sum_{i=1}^{m} \lambda_i |\beta|_{(i)}$ 已在式（10）中定义。W 是一个对角矩阵，其元素为 $W_{i,\,i} = w_i$，其余为 0，其中 w_i 为组权重，实际中根据组大小的平方根 $w_i = \sqrt{|I_i|}$ [29]~~]~~或 $w_i = 1/\sqrt{|I_i|}$设定。

所定义的组正则化 $\|\beta\|_{X,\,I}$ 意味着

$$\|\beta\|_{X,I} = \left(\|X_{I_1}\beta_{I_1}\|_2, \|X_{I_2}\beta_{I_2}\|_2, \ldots, \|X_{I_t}\beta_{I_t}\|_2\right)^T. \tag{16}$$

由于所要解决的问题已经改变，因此错误发现率不再适用。因此，Brzyski 等人定义了组假发现率（gFDR）[29]。设 g-SLOPE 选出的所有组的数量为 Rg，g-SLOPE 错误发现的组的数量为 Vg。因此有

$$Rg := |\{i : ||X_{I_i}\beta_{I_i}^{gS}||_2 \neq 0\}, \tag{17}$$

其中 $\beta_{I_i}^{gS}$ 第 i 组的定义见式（15）和（16），并且还有

$$Vg := |\{i : ||X_{I_i}\beta_{I_i}||_2 = 0, ||X_{I_i}\beta_{I_i}^{gS}||_2\}| \neq 0\}, \tag{18}$$

其中，$\|\beta\|_{X,\,I}$ 的估计由$\|\beta^{gS}\|_{X,\,I}$中非零项对应的索引定义。gFDR 定义为

$$\text{gFDR} := \mathbb{E}\left(\frac{Vg}{max\{Rg, 1\}}\right). \tag{19}$$

目标是确定 g-SLOPE 的正则化序列，使组假发现率 gFDR 能在任意水平 $\alpha \in (0,\ 1)$ 下得到控制。下面给出正交情形和一般情形下正则化参数形式的展开。当设计矩阵 X 为正交矩阵时，正则化参数序列 λ_{max} =(λ_1^{max}，λ_2^{max}，.....，λ_m^{max})^{T}定义为

$$\lambda_i^{max} := \max_{j=1,\ldots,m} \frac{1}{w_j} F_{\chi_{l_j}}^{-1}(1 - \frac{qi}{m}), \tag{20}$$

其中， $F_{\mathcal{X}_{l_j}}$ 是自由度为 l_j 的 $\mathcal{X}$ 分布的累积分布函数。于是有

$$\mathrm{gFDR} \leq \alpha \cdot \frac{m_0}{m}. \tag{21}$$

此外， [29] 证明了所得程序能够适应未知稀疏度，并且在估计不同组的回归变量所解释的响应变量方差比例时具有渐近极小极大性质。

3.4 逐步下降程序

逐步下降程序 [21]旨在控制 k-族错误率和错误发现比例，即给定阈值 α， $\gamma \in (0, 1)$，我们旨在控制

$$k\text{-FWER} \leq \alpha \tag{22}$$

and

$$\mathrm{Prob}\{\mathrm{FDP} > \gamma\} \leq \alpha. \tag{23}$$

设有 m 个单个检验 H_1，……， H_m，其对应的 p 值为 $\hat{}p_1$，……， $\hat{}p_m$。令 $\hat{}p_{(1)} \leq \hat{}p_{(2)} \leq ... \leq \hat{}p_{(m)}$ 为排序后的p值，并令非负且非递减序列 $\alpha_1 \leq \alpha_2 ... \leq \alpha_m$ 为 k-族错误率阈值。与排序后的 p 值对应的假设定义为 $H_{(1)}$， $H_{(2)}$...， $H_{(m)}$。那么，逐步下降程序可按以下步骤定义：

步骤 0：令 $i = 0$。
步骤 1：如果 $\hat{}p_{(i+1)} \geq \alpha_{i+1}$，则转到**步骤 2**。否则，令 $i = i+1$ 并重复**步骤 1**。
步骤 2：对 $j \leq k$ 拒绝 $H_{(j)}$，并对 $j > k$接受 $H_{(j)}$。

换句话说，如果 $p_{(1)} > \alpha_1$，则不拒绝任何原假设。否则，如果$H_{(1)}$， $H_{(2)}$...， $H_{(r)}$ 被拒绝，则最大的 r 满足

$$p_{(1)} \leq \alpha_1, p_{(2)} \leq \alpha_2, ..., p_{(r)} \leq \alpha_r. \tag{24}$$

基于逐步下降程序，Lehmann和Romano提供了两个不同的阈值，分别用于确保 k-族错误率控制和错误发现比例控制 [21]。

引理 2 [21] 对于检验 H_i， $i = 1$，……， m，给定 k 和 $\alpha \in (0,1)$，式（24）中描述的逐步下降程序满足

$$\alpha_i = \begin{cases} \frac{k\alpha}{m}, & i \leq k \\ \frac{k\alpha}{m+k-i}, & i > k \end{cases} \tag{25}$$

控制 k-族错误率，则式（22）成立 .

引理 3 [21] 对于检验 H_i， $i = 1$，……， m，给定 α， $\gamma \in (0,1)$，如果错误原假设对应的p值与正确原假设对应的p值相互独立，则式（24）中描述的逐步下降程序满足

$$\alpha_i = \frac{(\lfloor \gamma i \rfloor + 1)\alpha}{m + \lfloor \gamma i \rfloor + 1 - i} \tag{26}$$

按照（23）的意义控制错误发现比例 .

引理2和3表明， k-族错误率控制和错误发现比例控制在精心选择 α_i的情况下，逐步下降程序具有理论保证。事实上，逐步下降程序的这些理论性质促使我们设计新的 SLOPE 和 g-SLOPE 算法。

4 方法

本节将逐步下降程序 [21] 融入经典 SLOPE [27] 和 g-SLOPE [29]，构建我们基于逐步下降的全新方法，以控制特征选择，同时确保 k-族错误率控制和错误发现比例控制。首先，我们给出正交设计下用于控制 k-族错误率和错误发现比例的调谐参数序列。此外，我们还对一般设定下如何选择正则化参数进行了直观的理论分析。

4.1 正交设计

文献 [27]已阐明，在正交设计下，SLOPE 的多重检验与模型选择之间存在联系。沿着这一思路，我们假设 X是一个 $n \times m$维正交矩阵，即 $X^T X = I_m$，且 $\epsilon \sim N(0, \sigma^2 I_n)$ 是一个已知方差的n维列向量。然后，将线性回归模型$y = X\beta + \epsilon$ 进一步转化为

$$\tilde{y} = X^T y = \beta + X^T \epsilon \sim N(\beta, \sigma^2 I_p). \tag{27}$$

众所周知，有效特征的选择可归结为一个多重假设检验问题。将 m 个假设记为 $H_i: \ \beta_i = 0, \ 1 \le i \le m$。如果拒绝 H_i，则认为 β_i 是一个信息特征，反之亦然。Bogdan 等人针对 SLOPE [27] 的正则化参数，通过 BH 程序 [19] 给出了选择机制。在.....下。为简洁起见，我们分别将所提出的方法称为 k-SLOPE 和 F-SLOPE，分别对应族错误率和错误发现比例。

k-SLOPE 的正则化方案表述如下

$$\arg\min_{\beta \in \mathbb{R}^m} \frac{1}{2}\|y - X\beta\|_{l_2}^2 + \sigma \cdot \sum_{i=1}^{m} \lambda_{k\text{-FWER}}(i)|\beta|_{(i)}, \tag{28}$$

其中

$$\lambda_{k\text{-FWER}}(i) = \begin{cases} \Phi^{-1}(1 - k\alpha/2m), & i \le k \\ \Phi^{-1}(1 - k\alpha/2(m + k - i)), & i > k. \end{cases} \tag{29}$$

为刻画所提出的 k-SLOPE 的理论性质，我们引入以下引理，见 [27, 58]。

引理 4 [27] 设 H_i 为一个零假设，且设 $r \ge 1$。那么我们可以推导出

$$\{y : H_i \text{ is rejected and } R = r\} = \{y : |y_i| > \lambda_r \text{ and } R = r\}. \tag{30}$$

Remark 1 引理 1 表明，当选定特征的数量 R 等于 r ($r \geq 1$) 时，拒绝 H_i 的条件是其对应的 $|y_i|$ 大于 λ_r。

引理 5 [27] 考虑对 $\tilde{y} = (y_1, ..., y_{i-1}, y_{i+1}, ...y_p)$应用SLOPE程序，权重为 $\tilde{\lambda} = (\lambda_2, ..., \lambda_p)$，令 $\tilde{R}$ 表示该程序的拒绝数。那么，当 $r > 1$时，有

$$\{y : |y_i| > \lambda_r \text{ and } R = r\} \subset \{y : |y_i| > \lambda_r \text{ and } \tilde{R} = r - 1\}. \tag{31}$$

Remark 2 引理2指出，如果随机移除 y_i ，则选定特征的数量会多于前一个集合中的数量。

引理 6 [58] 令$s^{[k]}$为集合 $S = \{s_1, ..., s_n\}$的前k个元素，例如 $s_{[1]} \geq s_{[2]} \geq ... \geq s_{[n]}$。 $\sum_{i=1}^{k} s_{[i]}$ 是$(s_1, ..., s_n)$的凸函数。此外，对于 $s_i \geq 0$ 和 $i = 1, ..., n$，有 $\sum_{i=1}^{k} s_{[i]} = \min_{\lambda \geq 0}\{k\lambda + \sum_{i=1}^{n}[s_i - \lambda]_+\}$，其中 $[a]_+ = \max\{0, a\}$ 是铰链函数。

Remark 3 引理3给出了前 k 个非负值之和的等价形式。

所述 k-SLOPE 配备（29）后具有以下理论性质。

定理 1 *In the linear model (6) with the orthogonal matrix X and noise $\epsilon \sim N(0, \sigma^2 I_n)$, given k and $\alpha \in (0, 1)$ the k-FWER of the k-SLOPE model (28) satisfies (22).*

Proof X $\sigma^2 = 1$ $\tilde{y} = X^T y \sim$假设 N β I_m m H_i $\beta_i = 0$ $i = 1$ ... m m_0 该矩阵为正交设计矩阵，且 . 现在我们有 (，) 和原假设 : ,,,, 其中前几个假设为真实零假设。接着，将对应于前 m_0个真实零假设的 y 的绝对值排序，记为

$$|\hat{q}|_{(1)} \geq |\hat{q}|_{(2)} \geq \cdots \geq |\hat{q}|_{(m_0)}. \tag{32}$$

令 k为在可容忍范围内错误选定特征的最大数量，或满足 $\beta_i \neq 0$的索引在 $\{1, ..., m_0\}$内的数量。假设 $m_0 \geq k$，否则便无事可证。于是我们可以推导出，

$$k\text{-FWER} = \text{Prob}(V \geq k) = \sum_{r=1}^{m} \text{Prob}(V \geq k \text{ and } R = r). \tag{33}$$

通过逐步下降程序 [21] 和引理 30 [27]， k-SLOPE 至少会产生 k个错误拒绝，当且仅当

$$|\hat{q}|_{(1)} \geq \lambda_r, |\hat{q}|_{(2)} \geq \lambda_r, ..., |\hat{q}|_{(k)} \geq \lambda_r, \tag{34}$$

当选定特征的数量 $R = r$时。于是我们有

$$\begin{aligned} \text{Prob}(V \geq k \text{ and } R = r) =& \text{Prob}(|\hat{q}|_{(1)} \geq \lambda_r, ..., |\hat{q}|_{(k)} \geq \lambda_r \text{ and } R = r) \\ \leq& \text{Prob}(|\hat{q}|_{(k)} \geq \lambda_r \text{ and } R = r). \end{aligned} \tag{35}$$

由于引理 2 [27] 以及 $|\hat{q}|_{(k)}$ 与 $\tilde{y}$相互独立，我们有

$$\text{Prob}(V \geq k \text{ and } R = r) \leq \text{Prob}(|\hat{q}|_{(k)} \geq \lambda_r)\text{Prob}(\tilde{R} = r - 1) \tag{36}$$

不难发现，$\text{Prob}(|\hat{q}|_{(i)} \geq \lambda_r)$ 是非递增的，因此

$$\text{Prob}(|\hat{q}|_{(k)} \geq \lambda_r) \leq \frac{1}{k}\sum_{i=1}^{k}\text{Prob}(|\hat{q}|_{(i)} \geq \lambda_r) \tag{37}$$

接下来，结合引理 6 [58], ，我们可以推导出以下结果

$$\begin{aligned}\frac{1}{k}\sum_{i=1}^{k}\text{Prob}(|\hat{q}|_{(i)} \geq \lambda_r) =& \frac{1}{k}\min_{t\geq 0}\{kt + \sum_{i=1}^{m_0}[\text{Prob}(|\hat{q}_i| \geq \lambda_r) - t]_+\} \\ \leq& \min_{t_0 \leq t \leq \alpha}\{t + \frac{m_0}{k}[t_0 - t]_+\},\end{aligned} \tag{38}$$

其中 $t_0 = \text{Prob}(|\hat{q}_i| \geq \lambda_r) = k\alpha/(p + k - r)$。

将上述不等式代入（33）可得

$$\begin{aligned}k\text{-FWER} =& \sum_{r=1}^{m}\text{Prob}(|\hat{q}|_{(1)} \geq \lambda_r, ..., |\hat{q}|_{(k)} \geq \lambda \text{ and } R = r) \\ \leq& \sum_{r=1}^{m}\min_{t_0 \leq t \leq \alpha}\{t + \frac{p_0}{k}[t_0 - t]_+\}\text{Prob}(\tilde{R} = r - 1) \\ \leq& \sum_{r=1}^{m}\frac{k\alpha}{p + k - r}\text{Prob}(\tilde{R} = r - 1) \leq \alpha.\end{aligned} \tag{39}$$

证明至此完成 . □

Remark 4 定理 1 表明，在正交设计下， k-SLOPE 控制 k-族错误率。尽管 $\lambda_{k\text{-FWER}}(i)$ 是参照 [21], （28）选取的，但这与上述逐步下降程序并不等价。我们还在第 5 节的不同实验设置下，通过实证验证了这一理论发现。

一般来说，人们愿意接受的错误选择特征数量与已识别特征的数量成正比。因此，我们可能不再关注 k-族错误率，而是关注错误发现比例。类似于（28），F-SLOPE 的凸优化问题表述为

$$\arg\min_{\beta \in \mathbb{R}^m}\frac{1}{2}\|y - X\beta\|_{l_2}^2 + \sigma \cdot \sum_{i=1}^{m}\lambda_{\text{FDP}}(i)|\beta|_{(i)}, \tag{40}$$

其中

$$\lambda_{\text{FDP}}(i) = \Phi^{-1}(1 - \frac{(\lfloor \gamma i \rfloor + 1)\alpha}{2(m + \lfloor \gamma i \rfloor + 1 - i)}). \tag{41}$$

此外，所提出的正则化参数选择算法也给出了如下理论保证。

定理 2 *In the linear model (6) with the orthogonal matrix X and noise $\epsilon \sim N(0, \sigma^2 I_n)$ given α $\gamma \in$ 0, 1 ,() the FDP of the F-SLOPE model (40) satisfies (23).*

Proof 设真实假设的个数非零，即 $m_0 > 0$；否则无需再次证明。给定 $\gamma \in (0,1)$，在与定理4相同的假设下，我们有如下结论

$$\text{Prob}(\text{FDP} > \gamma) = \sum_{r=1}^{m} \text{Prob}(\frac{V}{R} > \gamma \text{ and } R = r) = \sum_{r=1}^{m} \text{Prob}(V \geq k(R) \text{ and } R = r), \quad (42)$$

其中 $k(R) = \lfloor \gamma R \rfloor + 1$。我们观察到，式（42）中的错误发现比例控制与式（ k-族错误率（33）相似，二者唯一的区别在于 k 的值是否受选定特征数量的影响。根据引理 30 [27] 和逐步下降程序 [21], ，我们可以推导出

$$\begin{aligned} \text{Prob}(V \geq k(R) \text{ and } R = r) =& \text{Prob}(|\hat{q}|_{(1)} \geq \lambda_r, ..., |\hat{q}|_{(k(R))} \geq \lambda_r \text{ and } R = r) \\ \leq& \text{Prob}(|\hat{q}|_{(k)} \geq \lambda_r \text{ and } R = r). \end{aligned} \quad (43)$$

类比式（36）和式（37），可得

$$\begin{aligned} \text{Prob}(V \geq k(R) \text{ and } R = r) \leq& \text{Prob}(|\hat{q}|_{(k(R))} \geq \lambda_r) \cdot \text{Prob}(\tilde{R} = r) \\ \leq& \frac{1}{k(R)} \sum_{i=1}^{k(R)} \text{Prob}(|\hat{q}|_{(i)} \geq \lambda_r) \cdot \text{Prob}(\tilde{R} = r - 1). \end{aligned} \quad (44)$$

再结合引理 6 [58], ，可得

$$\begin{aligned} \frac{1}{k(R)} \sum_{i=1}^{k(R)} \text{Prob}(|\hat{q}|_{(i)} \geq \lambda_r) =& \frac{1}{k(R)} \min_{t \geq 0} \{k(R) \cdot t + \sum_{i=1}^{m_0} [\text{Prob}(|\hat{q}_i| \geq \lambda_r) - t]_+\} \\ \leq& \min_{t_0 \leq t \leq \alpha} \{t + \frac{p_0}{k(R)} [t_0 - t]_+\}, \end{aligned} \quad (45)$$

其中 $t_0 = (k(r) + 1)q/(p + k(r) + 1 - r)$。

将上述不等式代入式（42）可得

$$\begin{aligned} \text{Prob}(\text{FDP} > \gamma) =& \sum_{r=1}^{m} P(V \geq k(R) \text{ and } R = r) \\ \leq& \sum_{r=1}^{m} \min_{t_0 \leq t \leq q} \{t + \frac{m_0}{k(R)} [t_0 - t]_+\} \text{Prob}(\tilde{R} = r - 1) \\ \leq& \sum_{r=1}^{m} \frac{k(r)q}{p + k(r) - r} \text{Prob}(\tilde{R} = r - 1) \leq \alpha, \end{aligned} \quad (46)$$

由此完成证明 . □

Remark 5 定理 1 保证了 F-SLOPE 在正交实验下进行错误发现比例控制的能力。后者的正交实验也支持结论（40），而 k- SLOPE 模型（28）则作为惩罚参数序列的选择机制。后者的正交实验也支持这一结论。此外， k-SLOPE 和 F-SLOPE 的优化算法与 SLOPE 相同，因为它们都是凸且非光滑的。更多优化细节见算法 1。

4.2 一般设定

通常，在非正交设定下，SLOPE 很难为错误发现率控制建立坚实的理论保证 [27]。因此， k-SLOPE 和 F-SLOPE 也可能

在这种一般设定下面临性能下降。幸运的是，Bogdan 等人利用其定性见解，对正则化参数序列进行了直观调整，并证明了其经验有效性 [27]。与 SLOPE 类似，我们通过理论分析给出 k-SLOPE 和 F-SLOPE 在一般设定下的正则化参数形式。

假设 k-SLOPE和 F-SLOPE 能正确检测这些特征，并准确估计回归系数的符号。令 X_S 和 β_S 分别表示与 $\beta_i \neq 0$ 相关联的变量子集及其系数值。非零分量对应的估计量近似为

$$\hat{\beta}_S \approx (X_S^T X_S)^{-1}(X_S^T y - \lambda_S) = \hat{\beta}_{\mathrm{LSE}} - (X_S^T X_S)^{-1}\lambda_S, \tag{47}$$

其中，$\lambda_S = (\lambda_1, ..., \lambda_{|S|})^T$，且$\hat{} \beta_{\mathrm{LSE}}$是 β_S的最小二乘估计量。

受 [27], 启发，我们计算 $X_i^T X_S(\beta_S - \hat{} \beta_S)$的分布，以确定 k-SLOPE 和 F-SLOPE 的正则化参数的具体形式。根据式(47)，我们可以推导出

$$\mathbb{E}(\beta_S - \hat{\beta}_S) \approx (X_S^T X_S)^{-1}\lambda_S, \tag{48}$$

and

$$\mathbb{E}X_i^T X_S(\beta_S - \hat{\beta}_S) \approx \mathbb{E}X_i^T X_S(X_S^T X_S)^{-1}\lambda_S. \tag{49}$$

在高斯设计下，X 的每个元素均为独立同分布，服从 $\mathcal{N}(0, 1/n)$，我们有如下结果

$$\mathbb{E}(X_i^T X_S(X_S^T X_S)^{-1}\lambda_S)^2 = \frac{1}{n}\lambda_S^T \mathbb{E}(X_S^T X_S)^{-1}\lambda_S = w(|S|) \cdot ||\lambda_S||^2 \tag{50}$$

and

$$w(|S|) = \frac{1}{n - |S| - 1}, \tag{51}$$

其中，$|S|$ 是 S中的元素个数，$i \notin S$ 而（51）依赖于以下事实：逆 $|S| \times |S|$ Wishart矩阵的期望，当其自由度为 n 时，等于$I_{|S|}/(n - |S| - 1)$ [59]。

k-SLOPE 从 $\lambda_{k\mathrm{G}} = \lambda_{k\text{-FWER}}(1)$ 开始。考虑到方差略有增加，我们有

$$\lambda_{k\mathrm{G}}(2) = \lambda_{k\text{-FWER}}(2)\sqrt{1 + w(2)\lambda_{k\mathrm{G}}(1)^2}. \tag{52}$$

因此，$\lambda_{k\mathrm{G}}$ 的序列可表示为

$$\lambda_{k\mathrm{G}}(i) = \lambda_{k\text{-FWER}}(i)\sqrt{1 + w(i-1)\sum_{j<i}\lambda_{k\mathrm{G}}(i)^2}. \tag{53}$$

F-SLOPE 与 k-SLOPE 之间唯一的区别在于惩罚项系数序列的选择。与（53）类似，F-SLOPE 从 $\lambda_{\mathrm{FG}} =$

$\lambda_{\mathrm{FDP}}(1)$，然后

$$\lambda_{\mathrm{FG}}(i)=\lambda_{\mathrm{FDP}}(i)\sqrt{1+w(i-1)\sum_{j<i}\lambda_{\mathrm{FG}}(i)^2}. \tag{54}$$

当惩罚项的系数序列递增时， k-SLOPE 和 F-SLOPE 不再是凸优化问题。将 $k^*:=k(n,\ m,\ \alpha)$ 记作全局最小值的下标， k-SLOPE 和 F-SLOPE 分别采用

$$\lambda_{\mathrm{kG}^\star}(i)=\begin{cases}\lambda_{\mathrm{kG}}(i), & i\le k^\star,\\ \lambda_{kG}\left(k^\star\right), & i>k^\star,\end{cases} \tag{55}$$

其中 $\lambda_{k\mathrm{G}}(i)$ 如（53）所示，且

$$\lambda_{\mathrm{FG}}(i)=\begin{cases}\lambda_{\mathrm{FG}}(i), & i\le k^\star,\\ \lambda_{FG}\left(k^\star\right), & i>k^\star,\end{cases} \tag{56}$$

其中 $\lambda_{\mathrm{FG}}(i)$ 在（54）中定义。当设计矩阵不服从高斯分布或其列不独立时，我们可以在公式（53）和（54）中使用校正项 [60]，而非 $w(i-1)\sum_{j<i}\lambda(i)^2$。

4.3 分组逐步 SLOPE

受组 SLOPE [29], 的启发，我们进一步将逐步下降 SLOPE 的思想扩展到具有组结构的信息特征子集的整体选择，而非单个回归变量。我们为新提出的分组逐步 SLOPE 方法构建相应的凸优化问题，包括分组 k-SLOPE（gk-SLOPE）和分组 F-SLOPE（gF-SLOPE）。值得注意的是，gk-SLOPE 和 gF-SLOPE 方法分别能够控制 gk-族错误率和 gFDP。

gk-族错误率和 gFDP 的控制定义如下。

定义 1 给定 $\alpha,\ \gamma\in(0,\ 1)$，g$k$-族错误率和 gFDP 控制定义为

$$\mathrm{g}k-\mathrm{FWER}=\mathrm{Prob}(Vg\ge k)<\alpha, \tag{57}$$

and

$$P(\mathrm{gFDP}>\gamma)=\mathrm{Prob}(\frac{Vg}{max\{Rg,1\}}>\gamma)<\alpha. \tag{58}$$

假设 m维响应向量 y 实际上由（13）中的线性模型生成。我们将 $||X_{I_i}\beta_{I_i}||_2$ 的值记为对第 i组响应变量的影响。我们认为，当且仅当 $||X_{I_i}\beta_{I_i}||_2>0$时，组 i 密切相关。因此，识别相关组的任务可以转化为找出$\|\beta\|_{X,I}=\left(||X_{I_1}\beta_{I_1}||_2,...,||X_{I_t}\beta_{I_t}||_2\right)^T$中哪些元素大于0。为估计$\|\beta\|_{X,I}$的非零系数，我们构建了两个分组逐步 SLOPE 模型，即 gk-SLOPE 和 gF-SLOPE，分别用于控制 gk- 族错误率与 gFDP。

关于 gk-族错误率的凸优化目标函数为

$$\arg\min_{\beta\in\mathbb{R}^m}\frac{1}{2}||y-X\beta||^2+\sigma J_{\lambda_{gk}}(W\|b\|_{X,I}), \tag{59}$$

其中， X 是任意矩阵。考虑 t仅包含单个变量的组（即单元素组的情形）、所有权重均为 1，且目标函数（59）简化为（28）。令 $\tilde{m}=l_1+l_2+...+l_t$ 并考虑 $\mathbb{I}=\{\mathbb{I}_1,...,\mathbb{I}_t\}$ 对 $\{1,...,t\}$的如下划分，我们可以推导出

$$\mathbb{I}_1:=\{1,...,l_1\},\mathbb{I}_2:=\{l_1+1,...,l_1+l_2\},...,\text{and}\ \ \mathbb{I}_t:=\{\sum_{i=1}^{t-1}l_i+1,...,\sum_{i=1}^{t}l_i\}. \tag{60}$$

我们可以发现， X_{I_i} 可以分解为 $X_{I_i}=U_iR_i$，其中 U_i 是一个矩阵，其 l_i 列相互正交且 l_2 范数为 1，其张成空间与 X_{I_i} 的列所张成的空间一致，而 R_i 是相应的满行秩矩阵。˜ ˜将 n 记为 l 矩阵，将 X 记为 $X_{\mathbb{I}_i}:=U_i$，并将 $c_{\mathbb{I}_i}$ 记为 $=R_ib_{I_i}$，其中 $i=1$、.....、 t。于是有

$$Xb=\sum_{i=1}^{t}X_{I_i}b_{I_i}=\sum_{i=1}^{t}U_iR_ib_{I_i}=\sum_{i=1}^{t}\tilde{X}_{\mathbb{I}_i}c_{\mathbb{I}_i}=\tilde{X}b, \tag{61}$$

and

$$(b_{X,I})_i=||X_{I_i}b_{I_i}||_2=||R_ib_i||_2=||c_{\mathbb{I}_i}||_2. \tag{62}$$

那么，问题（59）等价于

$$\begin{cases}c^{\text{gs}}:=\arg\min\limits_{c}\left\{\frac{1}{2}\|y-\widetilde{X}c\|_2^2+\sigma J_{\lambda_{gk}}\left(W\|c\|_{\mathbb{I}}\right)\right\}\\ c^{\text{gs}}_{\mathbb{I}_i}:=R_i\beta^{\text{gs}}_{I_i},i=1,\ldots,m\end{cases}, \tag{63}$$

其中， $\|c\|_{\mathbb{I}}:=\left(\|c_{\mathbb{I}_1}\|_2,.....,\|c_{\mathbb{I}_t}\|_2\right)^{\top}$。而（63）可视为（59）的标准形式。

下面，我们对所提出的两种 g-SLOPE 方法——gk-SLOPE 和 gF-SLOPE——进行直观的理论分析。在正交组情形或更一般的高斯情形下，它们分别控制 gk-FWER 和 gF-FWER 标准。

在陈述关于 gk-FWER 控制的主要分析之前，我们先回顾 $\mathcal{X}$ 分布的定义，并定义一种缩放后的 $\mathcal{X}$ 分布。

定义 2 [29] 当随机变量 X_1 可表示为 $X_1=\sqrt{X_2}$，其中 X_2 服从具有 l 个自由度的 $\mathcal{X}^2$ 分布时，我们称 X_1 服从具有 l 个自由度的 $\mathcal{X}$ 分布。此外，当随机变量X_1 可表示为 $X_1:=S\cdot X_2$，其中 X_2 服从具有 l 个自由度的 $\mathcal{X}$ 分布时，我们称其服从具有 l 个自由度、尺度参数为 S 的缩放 $\mathcal{X}$ 分布。为简便起见，我们使用记号 $X_1\sim S\mathcal{X}_l$。

算法 2： gk-SLOPE 的算法

input:$q \in (0,1), w_1, ..., w_t > 0, p, n, m, l_1, ..., l_m \in \mathbb{N}$; $\lambda_i := \overline{F}^{-1}(1 - \frac{qi}{m})$ for $\overline{F}(x) := \frac{1}{m}\sum_{i=1}^{m} F_{w_i^{-1}\mathbf{x}_{l_i}}(x)$;

for $i \in \{2, ..., m\}$ **do**

$\lambda^S := (\lambda_1, ..., \lambda_{i-1})^T$

$\mathcal{S}_j := \sqrt{\frac{n - l_j(i-1)}{n} + \frac{w_j^2 \|\lambda^S\|_2^2}{n - l_j(i-1) - 1}}, \quad$ for $j \in \{1, \dots, m\}$

if $i \le k, \lambda_i^* := \bar{F}_{\mathcal{S}}^{-1}\left(1 - \frac{k\alpha}{2m}\right),$

else $\lambda_i^* := \bar{F}_{\mathcal{S}}^{-1}\left(1 - \frac{k\alpha}{2(m+k-i)}\right), \quad$ where $\bar{F}_{\mathcal{S}}(x) := \frac{1}{m}\sum_{j=1}^{m} F_{\mathcal{S}_j w_i^{-1}\chi_{l_j}}(x)$

if $\lambda_i^* <= \lambda_{i-1}$, let $\lambda_i := \lambda_i^*$.

Otherwise, **end for** and $\lambda_j := \lambda_{i-1}$ for $j \ge i$

end for

命题 3 *In the linear model (13) with t disjoint groups and the orthogonal matrix* X ... ,

i.e. $X_{I_i}^T X_{I_j} = 0$ *for* $i \neq j$. *Define* w_1, , w_t *as the sizes of each group and given* $k \in \{1, 2, ..., t\}$ *the sequence of regularizing parameters can be expressed as* ,

$$\lambda_{gk}(i) = \begin{cases} \max_{j=1,...,t} \frac{1}{w_i} F_{X_{l_j}}^{-1}(1 - \frac{k\alpha}{2m}), & i \le k \\ \max_{j=1,...,t} \frac{1}{w_i} F_{X_{l_j}}^{-1}(1 - \frac{k\alpha}{2(m+k-i)}), & i > k. \end{cases} \tag{64}$$

where $F_{X_{l_j}}$ *is a cumulative distribution function of* X *distribution with* l_j *degrees of freedom.*

Then any solution β^{gS} *w.r.t. problem gk-SLOPE (59) generates the same vector* $\|\beta^{gS}\|_{X\ I} \cdot ,,0$ *Given* $\alpha \in 0,1$ *the gk-FWER control (57) holds.*

命题 3 表明，当正则化参数值满足式（64）时，gk-SLOPE 能够控制 gk-族错误率，这一点在后续数据实验中得到了验证。然而，在实际应用中使数据组正交非常具有挑战性。因此，我们将该模型推广到更一般的情形。与 k-SLOPE 类似，gk-SLOPE 也面临同样的问题，这可以通过对 λ_{gk} 在一般设定下进行启发式修正来解决。该修正序列是在解释变量从高斯分布中随机抽样这一假设下计算得到的。将这一想法与 [29], 相结合，我们提出算法 2 ，用于计算不同组中变量相互独立时的调优参数序列。

人们愿意接受的错误选定特征数量，与识别出的特征数量成正比。因此，我们可能不再关注 gk-族错误率，而是关注 gFDP。gF-SLOPE 的目标函数为

$$\arg\min_{\beta \in \mathbb{R}^m} \frac{1}{2}||y - X\beta||^2 + \sigma J_{\lambda_{gF}}(W\|b\|_{X,I}). \tag{65}$$

与（63）类似，问题（65）可以等价地表示为

$$\begin{cases} c^{\mathrm{gs}} := \arg\min\limits_{c} \left\{ \frac{1}{2}\|y - \tilde{X}c\|_2^2 + \sigma J_{\lambda_{gF}}\left(W\|c\|_{\mathbb{I}}\right) \right\}, \\ c^{\mathrm{gs}}_{\mathbb{I}_i} := R_i \beta^{\mathrm{gs}}_{I_i}, i = 1, \ldots, m. \end{cases} \tag{66}$$

我们可以将 (63) 视为 (59) 的标准版本。接下来，我们将给出两个序列；当不同组中的变量相互正交或服从高斯分布时，这两个序列可以保证控制 gk-FWER。

命题 **4** *In the linear model (13) with the orthogonal matrix* X *i.e.* $X_{I_i}^T X_{I_j} = 0$ *,* ,,, *for* $i \neq j$*. Define* w_1 ... w_t *are positive numbers and given* α $\gamma \in$ $0,1$ *the sequence of* ()*regularizing parameter can be expressed as*

$$\lambda_{gF}(i)^{\max} := \max_{j=1,\ldots,m} \left\{ \frac{1}{w_j} F^{-1}_{X^{l_j}} \left(1 - \frac{(\lfloor \gamma i \rfloor + 1)\alpha}{2(m + \lfloor \gamma i \rfloor + 1 - i)} \right) \right\}, \tag{67}$$

where $F_{X_{l_j}}$ *is a cumulative distribution function of* X *distribution with* l_j *degrees of freedom.*

Then any solution (β^{gS}) *w.r.t. problem gF-SLOPE (65) generates the same vector* $\|\beta^{gS}\|_{X\ I\cdot}$ *,* ( ) *Thus for given* $\alpha \in$ $0,1$ *the gFDP control (58) holds.*

命题 4 说明，当正则化参数的取值满足公式 (67) 时，gF-SLOPE 可以控制 gF-FWER，这一点在后续数据实验中得到了验证。然而，在实际应用中，要使数据组正交极具挑战性。因此，我们将该模型扩展到更一般的情形。与 F-SLOPE 类似，gF-SLOPE 也面临这些问题，可通过在一般设定下对 λ_{gF} 进行启发式修改来解决。该修改后的序列是在假设解释变量从高斯分布中随机抽样的条件下计算得到的。基于这一策略，我们设计了算法 3，用于在不同组中的变量相互独立时计算调节参数序列。注意，命题 3 和4 的证明是将我们的证明策略应用于定理1、2、3、1 和2 后 [29], 得到的，因此此处略去。

5 实证验证

本节验证我们提出的逐步下降 SLOPE 算法（如 F-SLOPE、 k-SLOPE、gF-SLOPE 和 gk-SLOPE）的错误控制性能。所有实验均使用 R 和 Python，在配备 Intel(R) Xeon(R) Platinum 8175M CPU 和 NVIDIA RTX A6000 GPU 的 PC 上进行。报告的结果为每项实验重复 100 次后得到的平均值。

5.1 正交设计实验（逐步下降 SLOPE）

受 [27, 29], 启发，我们构造设计矩阵 $X = I_n$ ，其中 $n = 1000$。随后，我们根据线性模型模拟响应变量

$$y = X\beta + \epsilon, \quad \epsilon \sim N(0, I_n). \tag{68}$$

算法 3： gF-SLOPE 的算法

输入：$\alpha, \gamma \in (0,\ 1)$ $w_1, ..., w_t > 0, p, n, m,\quad l_1, ..., l_m \in \mathbb{N}$;

$\lambda_i :=\overline{F}^{-1}(1 - \frac{(\lfloor \gamma i \rfloor + 1)\alpha}{2(m+\lfloor \gamma i \rfloor + 1 - i)})$ 对于 $\overline{F}(x) :=\ \frac{1}{m}\sum_{i=1}^{m} F_{w_i^{-1}\mathbf{X}_{l_i}}(x)$;

对于 $i \in \{2, ..., m\}$ **执行**

$\lambda^S :=\ (\lambda_1,\ .....,\ \lambda_{i-1})^T$

$\mathcal{S}_j :=\ \sqrt{\frac{n - l_j\ (i-1)\ n}{}} + \frac{w_j^2 \|\lambda^S\|_2^2}{n - l_j(i-1) - 1},$ 对于 $j \in \{1, \ldots, m\}$

$\lambda_i^* :=\bar{F}_{\mathcal{S}}^{-1}\Big(1 - \frac{(\lfloor \gamma i \rfloor + 1)2}{m + \lfloor \gamma i \rfloor + 1 - i)}\Big),$ for $=\ ^{F_{\mathcal{S}}(x):}\ \frac{1}{m}\sum_{j=1}^{m} F_{\mathcal{S}_j w_i^{-1}\chi_{l_j}}(x)$

若 $\lambda_i^* <= \lambda_{i-1}$，令 $\lambda_i := \lambda_i^*$。

否则，**结束循环**并且 $\lambda_j := \lambda_{i-1}$ 对于 $j \geq i$

结束循环

表 2：正交设计下受控特征选择的结果（不同t 和固定的 $k = 5$）。

t	k-SLOPE 错误发现比例概率 $> \gamma$	FDR	检验效能	F-SLOPE 错误发现比例概率 $> \gamma$	FDR	检验效能
50	0.001	0.006	1.000	0.003	0.007	1.000
100	0.000	0.002	0.998	0.002	0.005	1.000
200	0.001	0.001	1.000	0.000	0.007	1.000
300	0.002	0.001	1.000	0.000	0.005	0.995
400	0.000	0.001	0.995	0.001	0.004	0.994
500	0.000	0.001	0.997	0.000	0.005	0.997

相关特征的数量在 $t \in \{50, 100, 200, 300, 400, 500\}$ 之间变化，非零回归系数等于 $3\sqrt{2\log\ n}$。我们将目标假发现率水平设为 $\alpha = 0.1$ 和 $\gamma = 0.1$，用于 F-SLOPE，并将 $k \in \{5, 10, 15, 20, 25, 30\}$ 和 $\alpha = 0.1$ 设为k，用于2 -SLOPE。表格报告了 100 次重复试验中 FDR、Prob{错误发现比例 $\geq \gamma$} 和检验效能的估计结果。图 1 汇总了这些试验中 SLOPE、 k-SLOPE 和 F-SLOPE 的结果。这些结果表明，我们提出的基于逐步降阶的 SLOPE 方法能够灵活地控制错误发现比例、FDR 和 k-FWER，而经典 SLOPE 方法主要侧重于控制 FDR 标准。同时， k-SLOPE 和 F-SLOPE 在大多数经验配置下也表现出优异性能。此外，这些经验结果验证了定理4和定理5的有效性。

5.2 基于相关统计量的多重均值检验（逐步下降 SLOPE）

我们以相关检验统计量下的典型多重检验问题为例，说明所提出方法的性质。与 [27], 中的设置类似，我们考虑以下实验设计。研究人员开展 $n = 1000$ 项实验

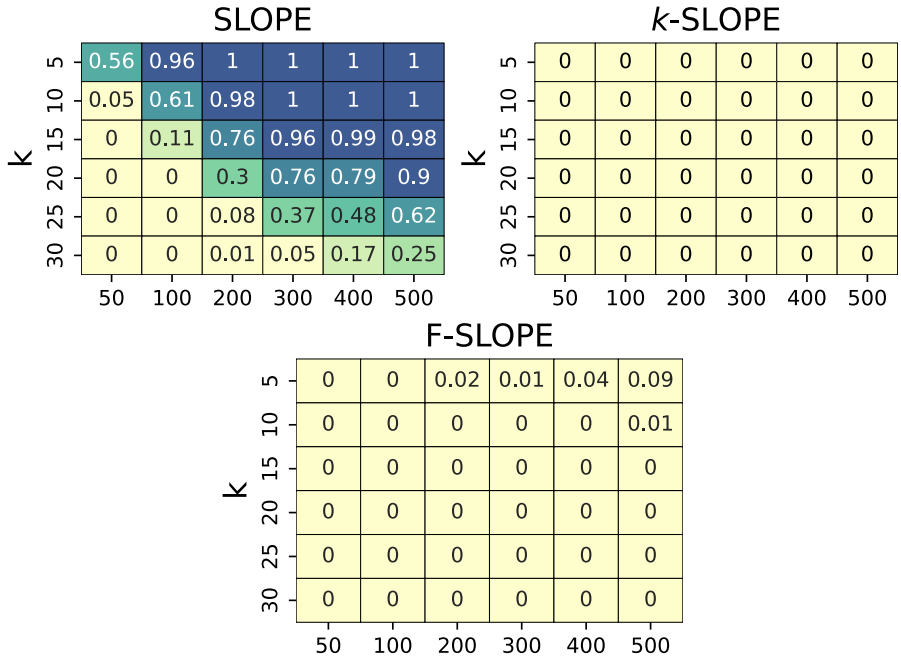

图 1：不同方法在正交设计下进行受控特征选择时的 k-族错误率（使用不同的 k 和 t）。小方框中的数值是 k-族错误率的大小。颜色越深，k-族错误率越高，反之亦然。

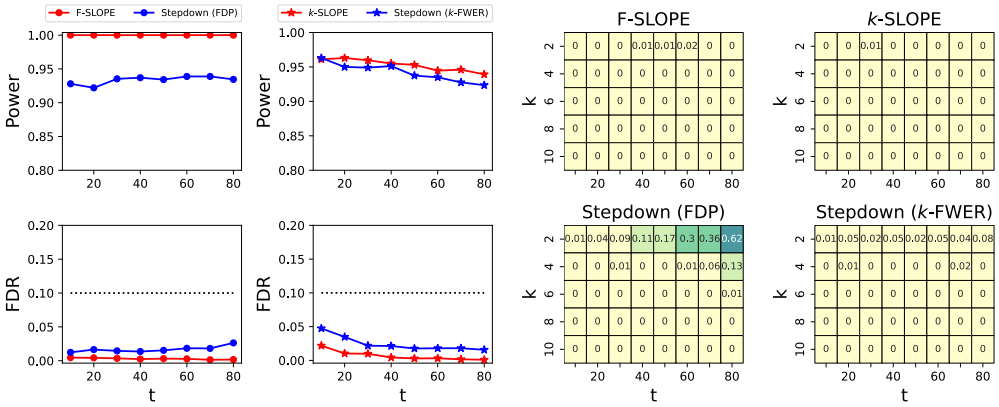

图 2：在模拟数据上进行受控特征选择的结果。黑色虚线表示目标假发现率水平。第二列（从左到右）中 k-SLOPE 的常数为 $k = 6$。第三列和第四列（从左到右）中，小方框内的数值是 k-族错误率的大小。颜色越深，k-族错误率越高，反之亦然。

表3：模拟数据上多重均值检验（$k=6$）的概率（FDP $>\gamma$）结果

t	F-SLOPE	k-SLOPE	标准差（FDP）	标准差（k-FWER）
10	0.00	0.04	0.02	0.08
20	0.03	0.00	0.00	0.03
30	0.00	0.01	0.01	0.01
40	0.00	0.00	0.00	0.00
50	0.01	0.00	0.00	0.00
60	0.00	0.00	0.00	0.00
70	0.00	0.00	0.00	0.00
80	0.00	0.00	0.00	0.00

表4：弱信号下模拟数据中 k-SLOPE（$m=2n$）的 k-FWER

k/t	10	20	30	40	50	60	70	80
2	0.02	0.00	0.02	0.00	0.04	0.04	0.07	0.07
4	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01

表5：弱信号下模拟数据中 F-SLOPE 的错误发现率（$m=2n$）

k/t	10	20	30	40	50	60	70	80
2	0.00	0.00	0.05	0.03	0.10	0.12	0.07	0.13
4	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.06
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
8	0.00	0.00	0.00	0.01	0.00	0.02	0.00	0.00

在随机选取的 $p=5$ 个实验室中。观测结果建模为

$$y_{i,j}=\mu_i+\tau_j+z_{i,j},\quad 1\le i\le n, 1\le j\le p, \tag{69}$$

其中，$\tau_j\sim N(0,\sigma_\tau^2)$ 为实验室影响因子，$z_{i,j}\sim N(0,\sigma_z^2)$ 为误差，且彼此独立。我们的目标是检验 μ_i 是否等于 0，即 $H_i:\ \mu_i=0, i=1,2,...,n$。对 5 个实验室的观测值取平均，得到结果的均值

$$\bar{y}_i=\mu_i+\bar{\tau}+\bar{z}_i,\quad 1\le i\le n, \tag{70}$$

其中 $\bar{y}=(\bar{y}_1,\ ...,\ \bar{y}_n)^T$独立抽取自 $\mathcal{N}(\mu,\ \Sigma)$，其中 $\Sigma_{i,\ i}=\frac{1}{5}\sigma_\tau^2=\rho$ 以及 $\Sigma_{i,\ j}=\frac{1}{5}(\sigma_\tau^2+\sigma_z^2)=\sigma^2$对于 $i\neq j$ [27]。关键问题是确定多元高斯相关向量的边际均值是否收敛于零。一种经典解法是使用 $\bar{y}$ 统计量进行边际检验，该方法依赖于逐步下降程序来控制 k-族错误率或错误发现比例 [21]。换言之，我们对$\bar{y}$序列按 $|\bar{y}|_{(1)}\ge|\bar{y}|_{(2)}\ge\cdots|\bar{y}|_{(m)}$进行排序。然后我们使用逐步下降程序，并配合

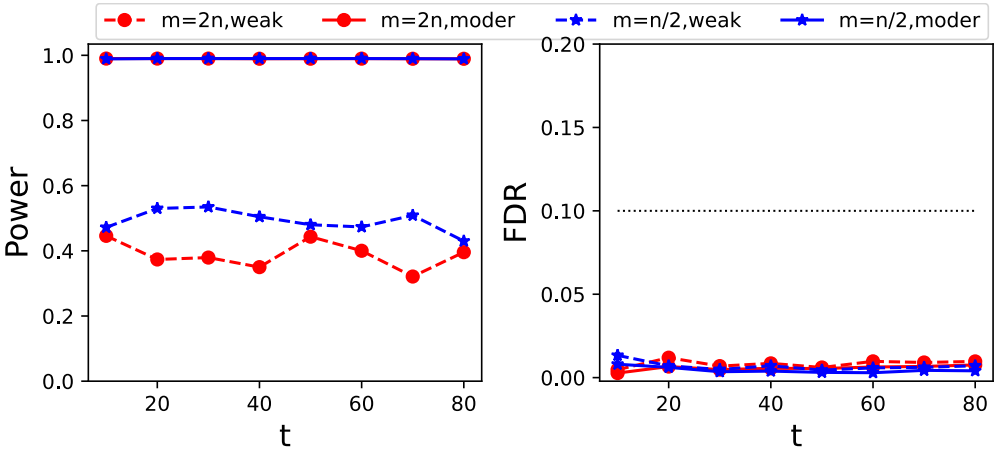

图 3：高斯设计下 F-SLOPE 的检验效能和错误发现率（不同 t）。黑色虚线表示目标错误发现率水平。

k-族错误率或错误发现比例临界值。另一种解法是“白化噪声”，即回归方程被简化为

$$\tilde{y} = \Sigma^{-1/2}\bar{y} = \Sigma^{-1/2}\mu + \epsilon, \tag{71}$$

其中 $\epsilon \sim N(0, I_p)$， $\Sigma^{-1/2}$ 是回归设计矩阵。假设 $\Sigma^{-1/2}$ 接近正交矩阵。在这种情况下，多重检验问题将转化为近似正交设计下的特征选择问题，此时 k-SLOPE 和 F-SLOPE 能够取得更好的表现。

按照 [27]，中的配置，我们令 $\sigma_\tau^2 = \sigma_z^2 = 2.5$，并考虑一个稀疏情形，其中相关特征的数量 $t \in \{10, 20, 30, 40$,50、60、70、80、90，$100\}$。非零均值设为满足 $2\sqrt{2}\log p/c$，其中 c等于 $\Sigma^{-1/2}$每列的欧几里得范数。对于所有错误发现比例受控的方法，我们设定 $\alpha = \gamma = 0.1$；对于 k-FWER 受控方法，我们设定 $k \in \{2, 4, 6, 8, 10\}$ 和 $\alpha = 0.1$。图 2 展示了 F-SLOPE、 k-SLOPE、错误发现比例控制的逐步下降程序（Sd(FDP)）以及 k-FWER 控制的逐步下降程序（Sd(k-FWER)）所达到的错误发现率、 k-FWER 和检验效能。表 3 展示了 F-SLOPE、 k-SLOPE 和逐步下降程序的 Prob(FDP $> \gamma$)。这些实验结果表明，我们提出的方法能够同时控制错误发现比例、错误发现率和k-FWER，而变体 Sd (FDP)（或 Sd (k-FWER)）则侧重于控制错误发现比例（或 k-FWER）和错误发现率。此外，所提出的 SLOPE 方法——F-SLOPE 和 k-SLOPE——比逐步下降程序具有更高的检验效能。因此，在多重检验中，我们提出的方法优于经典的逐步下降程序。

5.3 高斯设计实验（逐步下降 SLOPE）

我们在一般设定下研究 k-SLOPE 和 F-SLOPE 的性能。按照 [27]，中的策略，令设计矩阵 X 的各元素独立同分布地抽取自

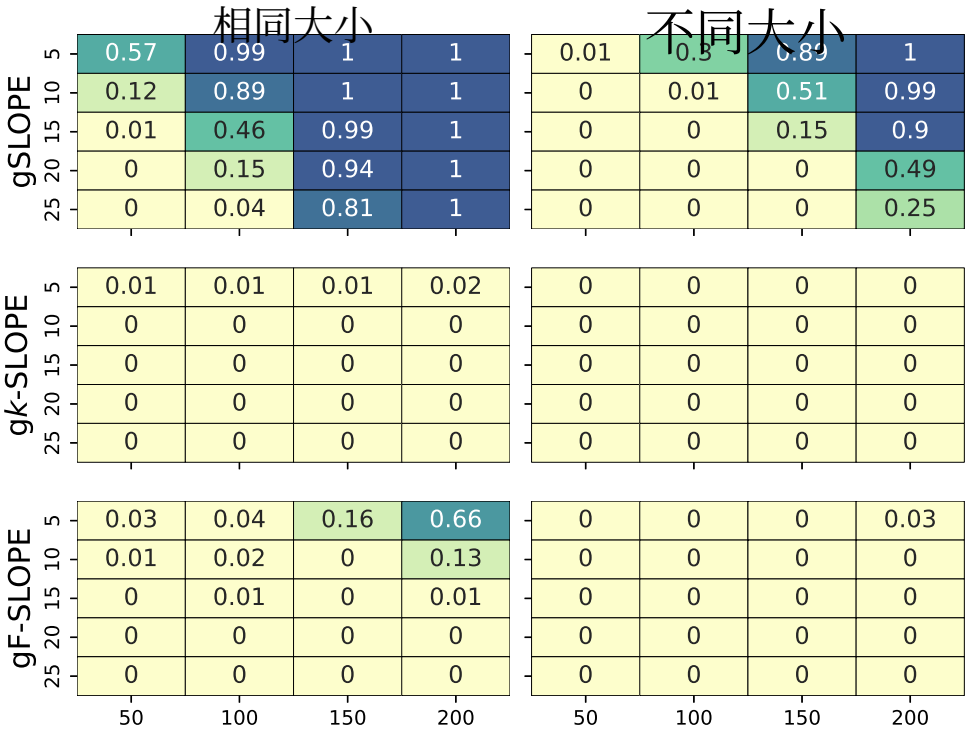

图4：正交设计下，采用不同方法进行受控特征选择时的 gk-族错误率。小方框中的数值为 gk-族错误率的大小。颜色越深，gk-族错误率越大，反之亦然。

分布 $\mathcal{N}(0,\ 1/n)$，其中 $n = 5000$。相关特征的数量 t 在 $\{10, 20, 30, 40, 50, 60, 70, 80\}$ 范围内变化。中等信号的非零系数相应设为 $2\sqrt{2\log\, m}$，而弱信号则设为 $\sqrt{2}\log\, m$。对于 F-SLOPE，我们设定$\alpha = \gamma = 0.1$；对于 k-SLOPE，则设定 $k = \{2, 4, 6, 8, 10\}$ 和 $\alpha = 0.1$。

接下来，我们考虑两种情景：(1) $m = 2n$；(2) $m = n/2$。表5和图3表明，在这两种情景（$m = 2n$和 $m = n/2$）下，无论信号是弱还是中等，F-SLOPE 都能将 Prob(FDP $> \gamma$) 和 FDR 控制在名义水平以下。同时，图3还显示，在弱信号下，F-SLOPE（$m = n/2$）的检验效能高于 F-SLOPE（$m = 2n$）；而在中等信号下，F-SLOPE（$m = 2n$）和 F-SLOPE（$m = n/2$）的检验效能相近。如表4所示，在这两种情景（$m = 2n$ 和 $m = n/2$）下，k-SLOPE 控制 k-FWER。此外，k-SLOPE 在中等信号下的检验效能也表现良好。此外，实验结果验证了 k-SLOPE 在 $\lambda_{k\mathrm{G}^*}$ 下以及 F-SLOPE 在 λ_{FG^*}下的有效性。

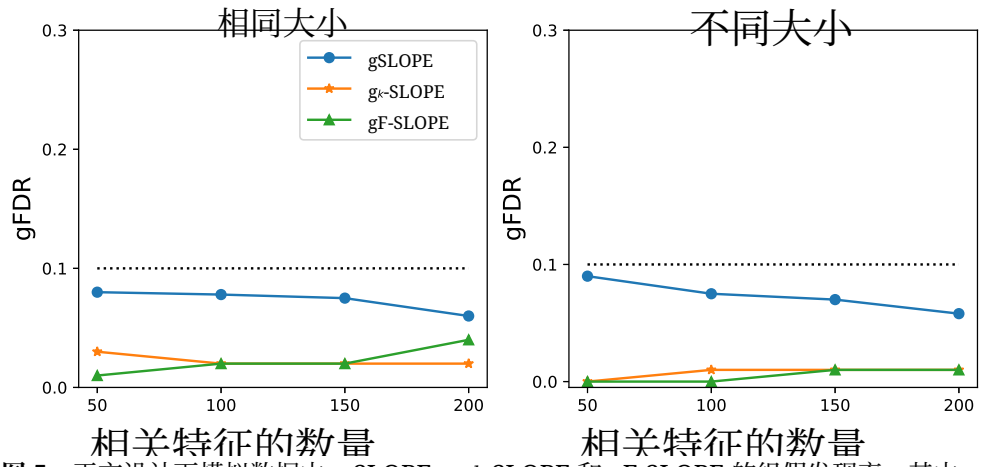

图 5：正交设计下模拟数据中 g-SLOPE、gk-SLOPE 和 gF-SLOPE 的组假发现率，其中 $k = 15$。

表6：弱信号和中等信号下，F-SLOPE 在模拟数据中的 Prob(FDP$>\gamma$)（不同 t）。

t	$m = 2n$		$m = n/2$	
	weak	中等	weak	中等
10	0.03	0.00	0.05	0.00
20	0.07	0.00	0.03	0.01
30	0.01	0.00	0.00	0.00
40	0.00	0.01	0.00	0.00
50	0.00	0.00	0.00	0.00
60	0.00	0.00	0.00	0.00
70	0.01	0.00	0.00	0.01
80	0.00	0.02	0.00	0.00

表7：g-SLOPE、gk-SLOPE 和 gF-SLOPE 在模拟数据中的 Prob(FDP$>\gamma$)。

数量	50	100	150	200
g-SLOPE	0.51	0.73	0.98	1.00
gk-SLOPE	0.01	0.00	0.00	0.01
gFSLOPE	0.01	0.00	0.00	0.00

5.4 正交设计实验（组间逐步下降 SLOPE）

受.....启发， [29], 我们在正交设置下设计了以下实验。令设计矩阵 X 为 p 单位矩阵，即 $X = I_p$ 其中 $p = 5000$。

表 8：正交设计中 g-SLOPE、gk-SLOPE 和 gF-SLOPE 的功效值。

数量	50	100	150	200	250
g-SLOPE	0.8	0.9	0.93	0.95	0.96
gk-SLOPE	0.99	0.98	1	0.99	1
gF-SLOPE	1	0.99	1	1	0.98

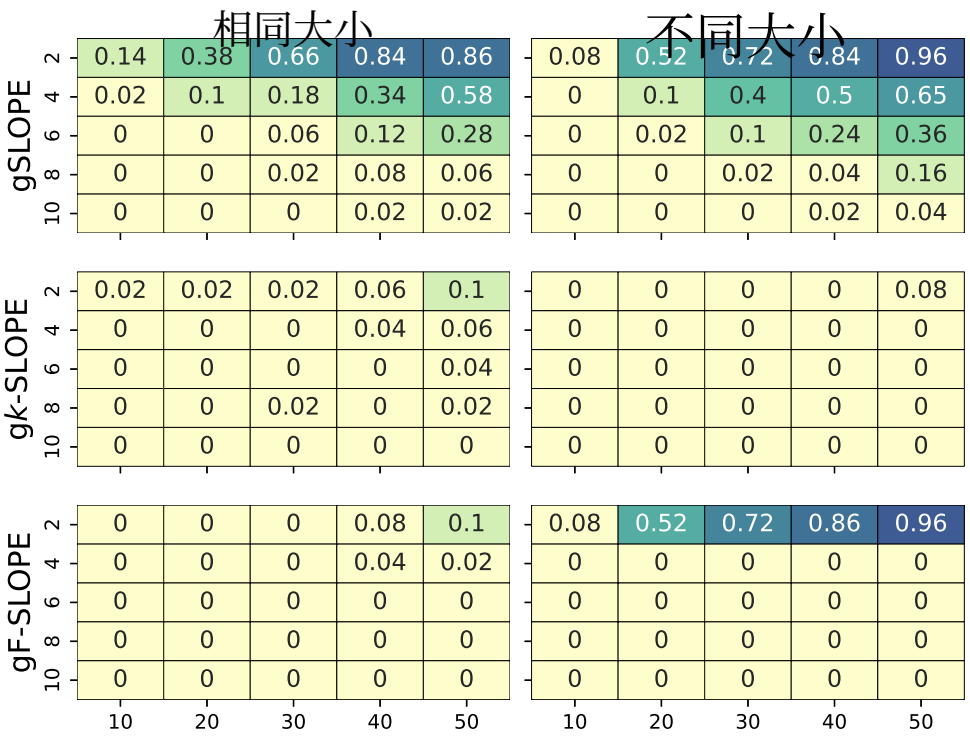

图6：正交设计下不同受控特征选择方法所提供的gk-族错误率。小方格中的数值为gk-族错误率的大小。颜色越深，gk-族错误率越大，反之亦然。

对于每个重要特征组，β_{I_i} 中的每个元素均按照 $U[0.1, 1.1]$ 上的均匀分布独立生成，并对 β_{I_i} 进行缩放，即 $(\beta_{X,\ I_i}) = a\sqrt{l_i}$，其中 a 需满足

$$\frac{1}{t}\sum_{i=1}^{t} a\sqrt{l_i} = \frac{1}{t}\sum_{i=1}^{t}\sqrt{4\ln(t)/\left(1-t^{-2/l}\right)-l}.$$

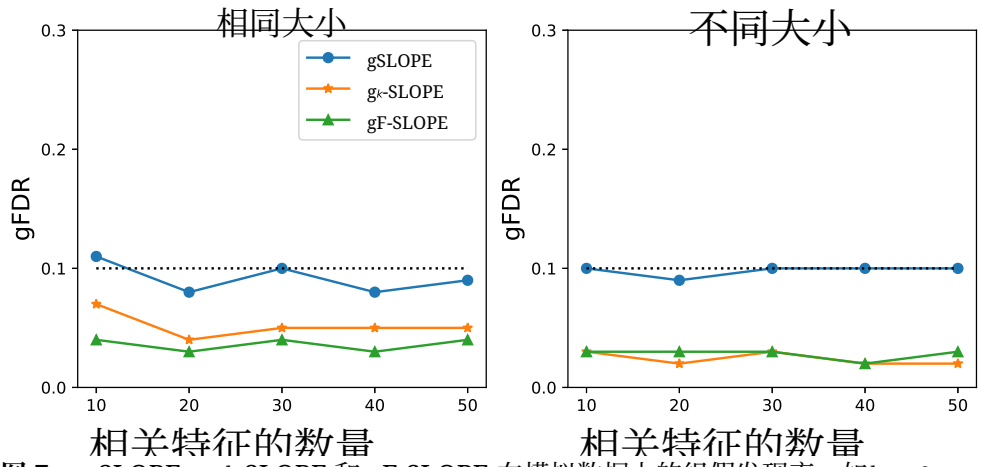

图 7：g-SLOPE、gk-SLOPE 和 gF-SLOPE 在模拟数据上的组假发现率，如$k = 6$。

将相关特征的数量设定为在 $\{50, 100, 200, 300, 400, 500\}$ 范围内变化。然后，我们根据线性模型模拟响应变量

$$y = \sum_{i=1}^{t} X_{I_i}\beta_{I_i} + \epsilon, \quad \epsilon \sim N(0, I_p). \tag{72}$$

对于分组设置，我们考虑两种情景：

（1）存在 1000 组解释变量，其确切组大小为$l = 5$;

（2）存在 1000 个组，它们被平均分为五类，具有不同的组大小 $\{3, 4, 5, 6, 7\}$。

我们在正交条件下比较组 SLOPE、gk-SLOPE 和 gF-SLOPE 这三种方法。我们将 gF-SLOPE 的目标假发现率水平设为 $\alpha = 0.1$ 和 $\gamma = 0.1$，将 gk-SLOPE 的目标假发现率水平设为 $k = \{5, 10, 15, 20, 25, 30\}$ 和 $\alpha = 0.1$。

图4 和 5 表明，尽管 g-SLOPE、gk-SLOPE 和 F-SLOPE 都能成功控制组假发现率，但 gk-SLOPE 在调控 gk-族错误率准则方面优于其他方法，而 g-SLOPE 明显不具备这一能力。表 7表明，gk-SLOPE 和 gFSLOPE 在控制 gFDP 方面优于 g-SLOPE，当$k = 15$时。在表 9中显然可见，无论关键特征的数量如何，k-SLOPE 和 F-SLOPE 的检验效能均高于 g-SLOPE。简而言之，我们提出的方法在控制 gk-族错误率、gFDP 和检验效能方面具有显著优势，从而最大限度地减少错误特征选择，并最大限度地选择真实特征，尤其是在正交情形下。

表9：模拟数据上 Prob(FDP> γ) 的结果。不同分组配置下的数值列于括号中。

Size	10	20	30	40	50
g-SLOPE	0.42(0.32)	0.28(0.38)	0.46(0.48)	0.24(0.44)	0.48(0.56)
gk-SLOPE	0.16(0.00)	0.08(0.00)	0.12(0.02)	0.02(0.00)	0.1(0.00)
gF-SLOPE	0.16(0.06)	0.08(0.00)	0.12(0.00)	0.02(0.00)	0.1(0.00)

Ta表 10：一般设计下的功效值　　设计。

Size	50	100	150	200	250
g-SLOPE	0.68	0.87	0.9	0.92	0.93
gk-SLOPE	1	1	1	1	1
gF-SLOPE	1	1	1	1	1

5.5 高斯设计实验（分组逐步 SLOPE）

为评估 g-SLOPE、gk-SLOPE 和 F-SLOPE 在高斯设计下的有效性，我们进行了样本量为 $n = 5000$ 和 $t = 1000$ 的模拟。对于每个稀疏度水平，以及组假发现率、gk-族错误率和 gFDP 水平为 0.1 的情形，在每次迭代中，我们使用高斯分布 $\mathcal{N}(0, 1/n)$ 生成设计矩阵的元素。随后，我们对 X 进行了标准化，并通过模型（72）利用 $\sigma = 1$生成了响应变量的值。信号的生成方式与正交设计中所呈现的模拟相同。相关特征的数量在$\{10, 20, 30, 40, 50, 60\}$内变化。分组情况与正交设计下相同。我们在高斯条件下比较组 SLOPE、gk-SLOPE 和 gF-SLOPE 这三种方法。

图 6 和 7 表明，尽管 g-SLOPE、gk-SLOPE 和 F-SLOPE 均能有效控制组假发现率，gk-SLOPE 和 gF-SLOPE 在控制 gk-族错误率标准方面优于其他方法，而 g-SLOPE 明显缺乏这种能力。表 9显示，当$k = 6$时，gk-SLOPE 和 gFSLOPE 在控制 gFDP 方面优于 g-SLOPE。表 10则清楚表明，无论关键特征的数量如何，k-SLOPE 和 F-SLOPE 的检验效能都高于 g-SLOPE。简言之，所提出的新型 g-SLOPE 方法，包括 gk-SLOPE 和 gF-SLOPE，在控制 gk-族错误率、gFDP 和检验效能方面具有显著优势，从而减少错误特征选择并最大限度地选择真实特征，尤其是在正交情形下。

5.6 在真实世界 ADNI 数据上的应用

我们在阿尔茨海默病神经影像学倡议（ADNI）的临床数据上评估 gk-SLOPE 和 gF-SLOPE（数据下载自 http://adni.loni.usc.edu）。这些数据包含关于组结构和关键特征的先验专家知识。为构建组建模框架，我们根据变量的结构类别或先验领域知识定义特征组。对于 ADNI 数据集，特征

表 11：真实 ADNI 数据上的结果。

方法	均方误差 ($\times 10^{-2}$)	时间（秒）	gk-FWER	概率(gFDP≥0.1)
g-SLOPE	2.84 ± 0.21	2.34 ± 0.82	0.82	0.48
组 Lasso	2.79 ± 0.19	0.82 ± 0.02	0.91	0.63
gk-SLOPE	2.61 ± 0.15	1.51 ± 0.43	0.12	0.17
gF-SLOPE	2.60 ± 0.14	1.59 ± 0.47	0.19	0.11

依据已知的先验组结构进行划分 [61–63]。该数据集包含 116 个变量组：其中 46 组各含一个协变量（皮层下体积，SV），其余 70 组各含四个协变量：皮层厚度均值（TA）、厚度标准差（TS）、表面积（SA）和皮层体积（CV） [64]。这种固有的分组结构为我们的分析方法提供了有价值的框架，并构成了本研究所用神经影像数据的基本组织原则。

表11 验证了我们提出的方法在实际应用中的未知分布下的有效性和灵活性。总体而言，实证结果支持理论结论。所提出的逐步正则化方案能够严格控制 g k-FWER 和 gFDP，同时避免了保守的多重检验校正通常伴随的明显检验效能下降。

6 结论

本文基于 SLOPE 技术构建了新的特征选择模型[27]。不同于现有关于错误发现率控制的研究，本文模型聚焦于特征选择中的 k-FWER 控制和错误发现比例控制。此外，我们将所提出的方法扩展到分组特征选择，从而进一步改进了经典的 g-SLOPE，并为 k-FWER 和错误发现比例控制提供了保证。借助逐步下降程序 [21]，我们在正交设计下确立了这些方法的理论保证。这项工作显著拓展了基于 SLOPE 的方法在多种错误控制准则和分组结构约束下实现严格特征选择的适用范围和实用价值。

参考文献

[1] Wu, L., Su, H., Zhang, X., Tai, Y., Li, H., Hu, W.：基于可解释图神经网络的 LUT 硬件木马自动检测。载于：2023 年 IEEE/ACM 国际计算机辅助设计会议（ICCAD），第 1–9 页（2023）。IEEE[2] Su, H., Hu, W., Zhang, X., Zhu, D., Wu, L.：面向 LUT 级硬件木马的精确且可解释定位。《IEEE 集成电路与系统计算机辅助设计汇刊》44(7)，2817–2821（2025）[3] Tibshirani, R.：基于 LASSO 的回归收缩与变量选择。《皇家统计学会杂志：B辑（方法论）》58(1)，267–288（1996）

[4] Yuan, M., Lin, Y.：含分组变量的回归模型选择与估计。皇家统计学会会刊：B辑（统计方法学）**68(1)**，49–67（2006）[5] Ida, Y., Fujiwara, Y., Kashima, H.：快速 Sparse Group Lasso。神经信息处理系统进展**32**（2019）[6] Ravikumar, P., Lafferty, J., Liu, H., Wasserman, L.：稀疏加性模型。《皇家统计学会会刊：B辑（统计方法学）》**71(5)**，1009–1030（2009）[7] Chen,H., Wang,X., Deng,C., Huang, H.：组稀疏加性机器。载于：第 31 届神经信息处理系统国际会议论文集，第 198–208 页（2017）[8] Chen, H., Wang, Y., Zheng, F., Deng, C., Huang, H.：稀疏模态加性模型。《IEEE神经网络与学习系统汇刊》**32(6)**，2373–2387（2021）[9] Zhang, W., Barr, B., Paisley, J.：高斯过程神经加性模型。载于：AAAI 人工智能会议论文集，第 38 卷，第 16865–16872 页（2024）[10] Zhu, L., Li, H., Zhang, X., Wu, L., Chen, H.：神经部分线性加性模型。《计算机科学前沿》**18(6)**，186334（2024）[11] Zhang, X., Chen, H., Wu, L.：适用于含标签噪声和类别不平衡数据集的可解释元加权稀疏神经加性网络。载于：第 34 届 ACM信息与知识管理国际会议论文集，第 5505–5509 页（2025）[12] Breiman, L.：随机森林。《机器学习》**45(1)**，5–32（2001）[13] Lemhadri, I., Ruan, F., Tibshirani, R.：LassoNet：具有特征稀疏性的神经网络。载于：人工智能与统计国际会议，第 10–18 页（2021）。PMLR[14]Taherkhani, A., Cosma, G., McGinnity, T.M.：Deep-fs：一种用于深度玻尔兹曼机的特征选择算法。《神经计算》**322**，22–37（2018）[15] Lu, Y., Fan, Y., Lv, J., Stafford Noble, W.：DeepPink：深度神经网络中可复现的特征选择。神经信息处理系统进展**31** （2018）[16] Li, M., Zhao, Y., Zhang, F., Luo, B., Yang, C., Gui, W., Chang, K.：用于轻量级图像超分辨率的多尺度特征选择网络。《神经网络》**169**，352–364（2024）

[17] Aggarwal, S., Yadav, A.K.：蛋白质组学中的错误发现率估计。载于：蛋白质组学统计分析，第 119–128 页。Springer，???（2016）[18] Yu, L., Kaufmann, T., Lederer, J.：生物网络中的错误发现率。载于：人工智能与统计国际会议，第 163–171 页（2021）。PMLR[19] Benjamini, Y., Hochberg, Y.：控制错误发现率：一种实用且有效的多重检验方法。《皇家统计学会杂志：B辑（方法论）》**57**(1), 289–300（1995）[20] Ferreira, J., Zwinderman, A.：论 Benjamini–Hochberg方法。《The Annals of Statistics》**34**(4), 1827–1849（2006）[21] Lehmann, E., Romano, J.P.：族系错误率的推广。《The Annals of Statistics》，1138–1154（2005）[22] Romano, J.P., Shaikh, A.M.：用于控制族系错误率推广形式的逐步递增程序。《The Annals of Statistics》**34**(4), 1850–1873（2006）[23] Barber, R.F., Cand`es, E.J.：通过 knockoff变量控制错误发现率。《The Annals of Statistics》**43**(5), 2055–2085（2015）[24] Cand`es, E., Fan, Y., Janson, L., Lv, J.：淘金：用于高维受控变量选择的“model-x” knockoff变量。《皇家统计学会会刊：B辑（统计方法学）》**80**(3), 551–577（2018）[25] Barber, R.F., Cand`es, E.J., Samworth, R.J.：基于 knockoff变量的稳健推断。《The Annals of Statistics》**48**(3), 1409–1431（2020）[26] Romano, Y., Sesia, M., Cand`es, E.：深度knockoff变量。《美国统计学会杂志》**115**(532), 1861–1872（2020）[27] Bogdan, M., Van Den Berg, E., Sabatti, C., Su, W., Cand`es,E.J.：SLOPE——通过凸优化实现自适应变量选择。《应用统计年刊》**9**(3), 1103（2015）[28] Su, W., Cand`es, E.：SLOPE 能适应未知稀疏度，且具有渐近极小极大性质。《The Annals of Statistics》**44**(3), 1038–1068（2016）[29] Brzyski, D., Gossmann, A., Su, W., Bogdan, M.：组 SLOPE——预测变量组的自适应选择。《美国统计学会杂志》**114**(525), 419–433（2019）[30] Javanmard, A., Javadi, H.：通过去偏Lasso控制错误发现率。《电子统计学期刊》**13**(1), 1212–1253（2019）

[31] Zhao, X., Chen, H., Wang, Y., Li, W., Gong, T., Wang, Y., Zheng, F.：用于受控特征选择的基于误差的 knockoffs 推断。载于：AAAI 人工智能会议论文集，第 36 卷，第 9190–9198 页（2022）[32] Jiang, W., Bogdan, M., Josse, J., Majewski, S., Miasojedow, B., Ro˘ckov´a, V., Group, T.：自适应贝叶斯 SLOPE：不完整数据下的模型选择。《Journal of Computational and Graphical Statistics》**31**(1), 113–137（2022）[33] Larsson,J.,Bogdan, M.,Wallin, J.：SLOPE 的强筛选规则。《神经信息处理系统进展》**33**, 14592–14603（2020）[34] Dupuis, X., Tardivel, P.：SLOPE 的解路径。载于：人工智能与统计国际会议，第 775–783 页（2024）。PMLR[35] Minsker, S., Ndaoud, M., Wang, L.：通过平方根 SLOPE 实现稳健且无需调参的稀疏线性回归。SIAM 数据科学数学杂志**6**(2), 428–453（2024）[36]Wei, Q., Zhang, Y., Zhao, Z.：采用预测变量组自适应选择的稀疏降秩回归。载于：2021 年第55届阿西洛马会议：信号、系统与计算机，第 145–149 页（2021）。IEEE[37] Chen, Y., Luo, Z.：组 SLOPE 惩罚低秩张量回归。《机器学习研究杂志》**24**(352), 1–30（2023）[38] Feser, F., Evangelou, M.：稀疏组 SLOPE：具有错误发现率控制的自适应双层选择。arXiv 预印本 arXiv:2305.09467（2023）[39] Luo, Z., Sun, D., Toh, K.-C., Xiu, N.：使用半光滑牛顿增广拉格朗日方法求解 OSCAR 和 SLOPE 模型。《机器学习研究杂志》**20**(106), 1–25（2019）[40] Larsson,J.,Bogdan, M.,Wallin, J.：SLOPE 的强筛选规则。《神经信息处理系统进展》**33**, 14592–14603（2020）[41] Larsson, J., Klopfenstein, Q., Massias, M., Wallin, J.：SLOPE 的坐标下降。arXiv 预印本 arXiv:2210.14780（2022）[42]Bellec, P.C., Lecu´e, G., Tsybakov, A.B.：SLOPE 遇上 Lasso：改进的预言界与最优性。《The Annals of Statistics》**46**(6B), 3603–3642（2018）[43] Kos, M., Bogdan, M.：论 SLOPE 的渐近性质。Sankhya A **82**(2), 499–532（2020）[44] Lee, S., Sobczyk, P., Bogdan, M.：在错误发现率控制下学习高斯马尔可夫随机场的结构。《Sy mmetry》**11**(10), 1311（2019）[45]Riccobello, R., Bogdan, M., Bonaccolto, G., Kremer, P.J., Paterlini, S.,

Sobczyk, P.：通过排序 l_1-范数进行稀疏图模型。arXiv 预印本 arXiv:2204.10403（2022）[46] Brzyski, D., Peterson, C.B., Sobczyk, P., Cand`es, E.J., Bogdan, M., Sabatti, C.：控制 GWAS 假发现率。Genetics **205**(1), 61–75（2017）[47] Kremer, P.J., Lee, S., Bogdan, M., Paterlini, S.：通过排序 l_1-范数进行稀疏投资组合选择。Journal of Banking & Finance **110**, 105687（2020）[48] Liang, J., Zhang, X., Chen, H., Li, W., Tang, X.：用于受控特征选择的逐步下降 SLOPE。载于：AAAI 人工智能会议论文集，第 37 卷，第 8728–8736 页（2023）[49] Romano, J.P., Wolf, M.：多重检验中广义错误率的控制。The Annals of Statistics **35**(4), 1378–1408（2007）[50]Alem´an, X., Duryea, S., Guerra, N.G., McEwan, P.J., Mu˜noz, R., Stampini, M., Williamson, A.A.：音乐训练对儿童发展的影响：委内瑞拉 El Sistema 的一项随机试验。Prevention Science **18**(7), 865–878（2017）[51] Delattre, S., Roquain, E.：通过 Romano–Wolf 启发式方法控制假发现比例的新程序。The Annals of Statistics **43**(3), 1141–1177（2015）[52] Laan, M.J., Dudoit, S., Pollard, K.S.：用于控制广义族错误率和假阳性比例尾概率的增广程序。Statistical Applications in Genetics and Molecular Biology **3**(1)（2004）[53] Farcomeni, A.：现代多重假设检验综述，特别关注假发现比例。Statistical Methods in Medical Research **17**(4), 347–388（2008）[54] Bakin, S.：数据挖掘问题中的自适应回归与模型选择。Spectrochimica Acta Part A Molecular Biomolecular Spectroscopy（1999）[55] Beck, A., Teboulle, M.：用于线性逆问题的快速迭代收缩阈值算法。SIAM Journal on Imaging Sciences **2**(1), 183–202（2009）[56] Becker,S.R., Cand`es, E.J., Grant, M.C.：凸锥问题的模板及其在稀疏信号恢复中的应用。Mathematical Programming Computation**3**(3), 165–218（2011）[57] Simon, N., Friedman, J., Hastie, T., Tibshirani, R.：稀疏组套索。Journal of Computational and Graphical Statistics **22**(2), 231–245（2013）[58] Fan, Y., Lyu, S., Ying, Y., Hu, B.-G.：使用平均 top-k 损失进行学习。载于：第 31 届神经信息处理国际会议论文集

Systems，第497–505页（2017）

[59] Nydick, S.W.：威沙特分布与逆威沙特分布。《电子统计学期刊》**6(1-19)**（2012）[60] Hammersley, J.M., Morton, K.W.：穷人的蒙特卡洛方法。《皇家统计学会杂志：B辑（方法论）》**16(1)**, 23–38（1954）[61] Liu, X.,Cao, P., Wang, J., Kong, J., Zhao, D.：融合组套索正则化的多任务特征学习及其在阿尔茨海默病认知表现预测中的应用。《神经信息学》**17(2)**, 271–294（2019）[62] Zhang, X., Wang, Y., Zhu, L., Chen, H., Li, H., Wu, L.：基于倾斜经验风险最小化的稳健变量结构发现。《应用智能》**53(14)**, 17865–17886（2023）[63] Zhou, W., Xu, C., Peng, L., Zhang, X.：适用于噪声标签和类别不平衡数据集的可解释双层加性泰勒模型。载于：2025 IEEE 第 31 届并行与分布式系统国际会议（ICPADS），第 1–8 页（2025）。IEEE[64] Chen, J., Zhang, X., Yuan, P., Lan, R., Chen, H.：最大似然神经加性模型。《信息科学》，123104（2026）