arXiv:2305.04182v2 [math.ST] 9 Dec 2024

一种用于高维双重稀疏线性回归的极小极大最优方法

Yanhang Zhang[1], Zhifan Li[2], Shixiang Liu[1], 和Jianxin Yin[1, 3]

[1]中国人民大学统计学院[2]北京应用数学研究院[3]中国人民大学应用统计中心与统计学院

摘要

本文重点研究高维双重稀疏线性回归问题，即逐元素稀疏性与逐组稀疏性的结合。为解决这一问题，我们提出了一种 IHT 风格的（迭代硬阈值）程序，该程序在每一步动态更新阈值。我们建立了参数估计的匹配上界和下界，证明了所提出方法在极小极大意义下的最优性。更重要的是，我们提出了一种完全自适应最优程序，旨在处理未知的稀疏度和噪声水平。我们的自适应程序具有快速收敛的特点，并展现出最优的统计精度。此外，我们阐明了逐元素稀疏度 s_0 作为 IHT 与组 IHT 之间权衡的意义，凸显了我们的方法相较于这两种方法的优越性能。借助 beta-min 条件，我们证明了所提出的 IHT 风格程序能够达到预言机估计速率，并在元素层面和组层面实现对真实支撑集的近乎完全恢复。最后，通过在合成数据集和真实世界数据集上与多种最先进的算法进行比较，我们进一步证明了所提出方法的优越性。

关键词： 双重稀疏性、迭代硬阈值、极小极大最优性、完全自适应过程、预言机估计速率。

1 引言

过去十年间，高维数据的迅速增长引起了众多科学领域对稀疏学习的广泛关注，并在算法、理论和应用方面取得了许多卓越成果。其中一个研究充分的问题

是稀疏约束线性回归，也称为最佳子集选择。我们考虑一个线性模型

$$y = X\beta^* + \xi,$$

其中，$y \in \mathbb{R}^n$ 是响应向量，$X \in \mathbb{R}^{n\times p}$ 是设计矩阵，$\beta^* \in \mathbb{R}^p$ 是潜在的回归系数，$\xi \in \mathbb{R}^n$ 是尺度参数为 σ^2 的次高斯随机误差。在高维框架下，我们关注 $p \gg n$ 且系数 β^* 是稀疏的情形，也就是说，只有少数协变量对模型具有重要作用。传统上，逐元素 ℓ_0 稀疏问题考虑的参数空间

$$\beta^* \in \{\beta \in \mathbb{R}^p : \sum_{i=1}^{p} \mathrm{I}(\beta_i \neq 0) \leq s'\},$$

其中，β_i 是 β 的第 i 个条目，s' 是某个正整数，用于控制稀疏约束线性回归问题的稀疏度。最佳子集选择是一个著名的 NP 难问题 [Natarajan，1995]，，并且已在统计学和机器学习领域得到广泛研究 [Bertsimas 等，2016，Yuan 等，2018，Huang 等人，2018，Zhu 等，2020]。

近年来，越来越多关于高维变量选择的研究聚焦于结构化稀疏性的概念。这些研究假设重要变量形成特定的结构或模式，其中逐组稀疏性是最突出的示例之一。逐组 ℓ_0 稀疏性考虑参数空间

$$\beta^* \in \{\beta \in \mathbb{R}^p : \sum_{j=1}^{m} \mathrm{I}(\beta_{G_j} \neq 0) \leq s\},$$

其中，$\{G_j\}_{j=1}^m$ 是 m 不重叠组的索引，满足 $\cup_{j=1}^m G_j = \{1, \ldots, p\}$。此处正整数 s 控制模型中非零组的数量。组稀疏性意味着，在一个组内，系数要么全部为零，要么至少有一个非零系数。特别地，当 $|G_1| = \ldots = |G_m| = 1$ 时，组选择问题归结为标准最佳子集选择。迄今为止，已经探索并研究了多种实用算法，以执行组 ℓ_0 选择 [Eldar 等人，2010年，Huang 等人，2011年，Hazimeh 等人，2023年，Zhang 等人，2023]。

在考虑每个已选中的组时，人们普遍认为，组成该组的变量中实际上只有少数是显著的。我们将这一思想称为双重稀疏性，并定义如下：

定义 1（双重稀疏性）回归系数 $\beta^* \in \mathbb{R}^p$ 称为（s, s_0）-稀疏，如果

$$\|\beta^*\|_{0,2} \coloneqq \sum_{j=1}^{m} \mathrm{I}(\beta^*_{G_j} \neq 0) \leq s \quad and \quad \|\beta^*\|_0 \coloneqq \sum_{i=1}^{p} \mathrm{I}(\beta^*_i \neq 0) \leq ss_0. \tag{1}$$

双重稀疏性同时促进组内和组间的稀疏性。具体而言，它将模型中非零组的数量限制为 s，并且在这 s 个组中，非零元素的数量不得超过 ss_0。直观地说， s_0 可以理解为 s 个选定组内的平均稀疏度，从而揭示非零组内的稀疏度水平。

1.1 相关工作

近年来，稀疏组选择已成为高维结构化稀疏学习的重要研究领域。为解决这一问题，通常会考虑结合两种惩罚方法。为了实现稀疏组选择，Friedman 等人[2010] 和 Simon 等人 [2013] 提出了稀疏组Lasso（SGLasso），将 Lasso 惩罚项 [Tibshirani、1996] 与分组套索惩罚项 [Yuan 和 Lin 2006]结合起来。许多研究致力于加速 SGLasso 的收敛 [Ida 等人，2019年，Zhang 等人， 2020]。

关于双重稀疏性的理论研究始于 Cai 等人 [2022], ，他们建立了双重稀疏线性回归估计误差的极小极大下界，并在不可表示条件下获得了 SGLasso 估计误差的近最优上界。此外，他们还为 SGLasso 的样本复杂度和估计误差提供了理论保证。Li等人 [2024] 研究了具有双重稀疏结构的高斯位置模型。他们建立了 $\ell_u(\ell_q)$ 混合范数下 $u, q \in [0, 1]$估计误差的极小极大速率。尽管取得了这些进展，但具有最优理论保证的方法仍然十分匮乏。

传统的基于凸松弛的方法（如 SGLasso）会不可避免地使系数产生估计偏差，尤其是在较大的系数受到显著收缩时。此外，Bellec [2018] 证明，即使满足 beta-min 条件，凸估计量（如 Lasso 型估计量）也无法达到预言机估计速率 $O(\sigma\sqrt{\frac{ss_0}{n}})$。这一现象促使我们开发计算上可行的非凸算法，其中迭代硬阈值（IHT，Blumensath 和 Davies [2009]）就是一个典型示例。IHT 及其变体因其在解决各类高维统计推断问题方面的有效性而日益受到关注 [Blumensath 和 Davies，2010年，Jain 等，2014年，Yuan 等，2020年，Hao 等，2021]。给定稀疏度 s'，IHT 对参数 β执行梯度下降步骤，然后在每个后续步骤中选取绝对值最大的 s' 个分量。在受限凸性/光滑性条件下，Jain 等人 [2014] 证明，只要稀疏度足够大，IHT 就能为高维 M 估计获得极小极大最优估计量。Yuan 等人 [2018] 研究了 RIP 型条件下 IHT 对 $s = s^*$ 和 $s \gg s^*$ 的参数估计与支撑集恢复。Giraud [2021] 在线性回归的组稀疏性情形中采用了 IHT过程，并建立了参数估计的最优上界。然而，大多数相关研究

将已知稀疏度 s' 作为先验信息，因此在未知 s'的情况下，难以从非渐近角度分析理论保证。为解决这一问题，Ndaoud [2020] 提出了一种完全自适应的 IHT 风格的程序，在未知 s'的情况下也能实现参数估计的最优速率。

1.2 主要结果与贡献

本文旨在为双重稀疏线性回归构造可行的方法，这些方法不仅高效，而且具有保证的最优统计性质。据我们所知，本文首次针对高维双重稀疏线性回归，在 s、 s_0 和 σ 未知的情况下，提出了完全自适应最优程序。

在 Cai 等人 [2022], Li等人 [2024] 之前，如何处理满足双重稀疏假设的信号一直是一个尚未解决的难题。Cai 等人 [2022] 采用的方法依赖于次梯度和对偶证书的构造，仅适用于 ℓ_1 型惩罚。Li等人 [2024]较早提出了一种 IHT式算法，用于检测具有双重稀疏结构的信号。他们证明了所提出算法在参数估计方面具有极小极大最优性。然而，该算法并不实用，因为它依赖于未知参数 s, s_0 和 σ。值得注意的是，实现双重稀疏性的自适应性，远比实现逐元素稀疏性或逐组稀疏性的自适应性更具挑战性。一种自然的方法是采用网格搜索技术来调整未知参数 s 和 s_0 ，例如 Cai 等人 [2022] 所采用的方法。然而，网格搜索方法在计算上不可行，而且从理论角度也难以建立最优保证。受逐元素稀疏性的自适应框架 [Verzelen, 2012, Ndaoud, 2020], 启发，我们提出了一种两步自适应程序，用于双重稀疏线性回归中的参数估计和变量选择。

重要的是，我们的程序并非经典 IHT [Ndaoud, 2020]与组 IHT [Giraud, 2021]的简单组合。我们的两个 IHT 算子的序列至关重要，顺序不可互换。具体而言，调换这两个步骤的顺序可能会破坏反证法的逻辑框架。

与稀疏组Lasso等凸优化方法相比，双重稀疏 IHT 的优势显而易见。我们的理论完全基于 RIP型条件，而稀疏组Lasso 的理论（参见 Cai 等人 [2022]）则依赖更强的不可表示条件。我们进一步证明，在 beta-min 条件下，我们的算法能够达到预言机估计速率 $O(\sigma\sqrt{\frac{ss_0}{n}})$，彰显了我们的算法相较于稀疏组Lasso 的优越性。此外，据我们所知，尚未有研究在温和假设下确立稀疏组Lasso 的支撑集恢复结果，而我们则在逐元素和逐组层面都实现了近乎完全恢复 [Butucea等人， 2018] 。合成数据和真实世界数据分析进一步支持了这一结论。

总之，本文的主要贡献总结如下:

- 我们提出了一种新颖的双重稀疏 IHT 算子，可同时确保逐元素和

逐组稀疏性。该算子由两个步骤组成，能够高效地控制模型复杂度。在双重稀疏 IHT 算子的基础上，我们提出了一种新颖的 IHT 式程序，可在每次迭代时动态更新阈值。我们分析了所提方法的估计误差上界，并建立了相匹配的估计误差极小极大下界 $O\left(\sqrt{\frac{\sigma^2}{n}(ss_0 \log \frac{ed}{s_0} + s \log \frac{em}{s})}\right)$，从而确证了所提方法的最优性。

- 我们提出了一种完全自适应最优程序，可处理未知稀疏度s，s_0 和噪声水平 σ。通过研究，我们证明了自适应程序得到的估计量在极小极大意义下达到了最优性能。据我们所知，这是首个针对双重稀疏线性回归的极小极大自适应程序。此外，我们发现元素稀疏度 s_0 在 IHT 与组 IHT 之间的权衡中发挥着关键作用，这凸显了我们的方法相较于两者的优越性能。我们已将提出的方法实现为一个名为 `ADSIHT` 的开源 R 软件包。

- 在逐元素和逐组 beta-min 条件下，我们证明了我们的算法达到了预言机估计速率 $O(\sigma\sqrt{\frac{ss_0}{n}})$。这一结果表明，在给定真实支撑集的情况下，我们的程序能够达到与普通最小二乘估计量相当的性能。这从理论上凸显了我们的 DSIHT 程序相较于稀疏组 Lasso 等凸方法的优越性。此外，我们证明了我们的程序能够在元素和组层面实现对真实支撑集的近乎完全恢复。

- 我们将提出的方法应用于合成数据集和真实世界数据集，并与几种最先进方法进行全面的实证比较。结果表明，我们的方法在多种指标上均表现出优势。此外，真实世界数据集的计算结果显示，我们的方法使用更少的变量和组，仍能实现更高的预测准确性。

1.3 章节安排

本文其余部分安排如下。本节末尾将介绍全文使用的符号。在第 2 节中，我们提出一种收敛速度快的 IHT 式程序，并为估计误差建立相匹配的上界和下界。在第 3 节中，我们首先提出一种新颖的信息准则，用于确定最优停止时间，并开发一种自适应程序，在未知 s 和 σ 的情况下进行稀疏组选择。随后，我们阐明本文工作与 IHT 及组 IHT 之间的联系。我们还提出一种极小极大自适应程序，用于选择 s_0 的最优值，从而使我们的方法成为完全自适应最优程序。在第4节中，我们证明 DSIHT 算法能够达到预言机估计速率，并实现

在 beta-min 条件下近乎完全恢复。在第 5 节中，我们通过合成数据集和真实世界数据集上的数值实验，将我们的方法与几种最先进的方法进行比较。最后，在第 6 节中，我们总结本研究，并在附录中给出主要结果的详细证明。

1.4 记号

对于给定的序列 a_n 和 b_n，当某个正常数 c 使得 $a_n \leq cb_n$ （相应地， $a_n \geq cb_n$）时，我们称 $a_n = O(b_n)$ 或 $a_n \lesssim b_n$ （相应地， $a_n =\Omega(b_n)$ 或 $a_n \gtrsim b_n$）。如果 $a_n = O(b_n)$ 且 $a_n = \Omega(b_n)$，则记为 $a_n \asymp b_n$。令 $d = \max_{1\leq j\leq m} |G_j|$ 表示最大的组大小。记 $[m]$ 为集合 $\{1, 2, \ldots, m\}$，并将 $\mathrm{I}(\cdot)$ 记为示性函数。令 $x \vee y$ 为 x 和 y 的最大值，而 $x \wedge y$ 为 x 和 y 的最小值。记 $S^* = \{i : \beta_i^* \neq 0\} \subseteq [p]$ 为 β^* 的支撑集。类似地，令 $G^* = \{j : \beta_{G_j}^* \neq 0、 G_j \subseteq [p]$，以及 $G_j \cap G_{j'} = \emptyset、 \forall j \neq j'\} \subseteq [m]$ 为 β^* 的逐组支撑集。令 $S_{G^*} = \cup_{j\in G^*} G_j$ 表示包含于组 G^* 中的所有元素。显然， $S^* \subseteq S_{G^*}$。对于任意基数为 $|S|$ 的集合 S，令 $\beta_S^* = (\beta_j, \ j \in S) \in \mathbb{R}^{|S|}$ 且 $X_S = (X_j, \ j \in S) \in \mathbb{R}^{n\times|S|}$；令 $(X^\top X)_{SS} \in \mathbb{R}^{|S|\times|S|}$ 表示 $X^\top X$ 的子矩阵，其行和列均列于 S 中。对于向量 β，记 $\|\beta\|_2$ 为其欧几里得范数。对于矩阵 A，记 $\|A\|_2$ 为其谱范数，记 $\|A\|_F$ 为其 Frobenius 范数。记 $\mathbb{I}_p$ 为 $p \times p$ 阶单位矩阵。令 $C、 C_0、 C_1、$ .....表示正常数，其实际值可随时变化。记双重稀疏性的参数空间为 $\Theta^{m、 d}(s, \ s_0)$。记 $\mathcal{S}^{m、 d}(s, \ s_0)$ 为由所有 $(s, \ s_0)$-稀疏向量的支撑集组成的空间。值得注意的是，根据双重稀疏性的定义，对于任意正常数 $a_1b_1 = a_2b_2$ 和 $a_1 \leq a_2$，有 $\mathcal{S}^{m、 d}(a_1s, \ b_1s_0) \subseteq \mathcal{S}^{m\, d}(a_2s, \ b_2s_0)$。例如， $\mathcal{S}^{m、 d}(2s, \ 2s_0)$ 是 $\mathcal{S}^{m、 d}(4s, \ s_0)$ 的子空间。为便于计算，我们假设 $\|X_j\|_2 = \sqrt{n}、 \forall j \in [p]$。

2 极小极大最优性分析

在第 2.1 节中，我们介绍双重稀疏迭代硬阈值（DSIHT）算子。具体而言，我们清晰地解释其构造，并提出一种在稀疏度和噪声水平已知时使用的 DSIHT 算法。随后，在第2.2节中，我们分析估计误差的来源。接着，我们在第 2.3 节中建立 DSIHT 算法的参数估计上界。在第 2.4 节中，我们推导双重稀疏线性回归的极小极大下界，由此得出第 2.3 节中的上界是极小极大最优的。

2.1 双重稀疏迭代硬阈值算子

给定 $\lambda, \ s_0 > 0$，我们将双重稀疏迭代硬阈值算子 $\mathcal{T}_{\lambda, \ s_0} : \mathbb{R}^p \to \mathbb{R}^p$ 定义为以下两个步骤：

步骤 1（逐元素条件检查）：定义一个作用于 $\beta \in \mathbb{R}^p$ 上的逐元素硬阈值算子 $\mathcal{T}_\lambda^{(1)}$：$\mathbb{R}^p \to \mathbb{R}^p$。

$$\{\mathcal{T}_\lambda^{(1)}(\beta)\}_j = \beta_j \mathrm{I}(|\beta_j| \geq \lambda), \quad \forall j \in [p].$$

算子 $\mathcal{T}_\lambda^{(1)}$ 保留绝对值大于或等于λ的信号，因此可视为识别重要变量的初步筛选过程。

步骤 2（逐组条件检查）：记为

$$\mathcal{J}_{s_0} \coloneqq \{j \in [m] : \|\beta_{G_j}\|_2^2 \geq s_0 \lambda^2\}.$$

算子 $\mathcal{T}_\lambda^{(2)}{}_{,s_0}$ 的定义：$\mathbb{R}^p \to \mathbb{R}^p$ 是

$$\{\mathcal{T}_{\lambda,s_0}^{(2)}(\beta)\}_{G_j} = \begin{cases} \beta_{G_j}, & \text{if } j \in \mathcal{J}_{s_0}. \\ 0, & \text{if } j \in [m] \backslash \mathcal{J}_{s_0}. \end{cases}$$

算子 $\mathcal{T}_\lambda^{(2)}{}_{,s_0}$ 筛选幅值较大的组，并利用组信息进一步筛选重要变量。算子 $\mathcal{T}_{\lambda,s_0} = \mathcal{T}_{(2)\,\lambda,s_0} \circ \mathcal{T}_{(1)\,\lambda}$ 是这两个步骤的组合。与经典迭代硬阈值程序不同，我们的程序在每一步都会更新 $\mathcal{T}_{\lambda,s_0}$ 中的阈值λ，以实现最优统计精度和快速收敛。给定 $\lambda_0 > \lambda_\infty > 0$ 和 $0 < \kappa < 1$，我们给出序列$\{\lambda_t\}_{t=1}^\infty$ 的形式如下

$$\lambda_t = \kappa^t \lambda_0 \vee \lambda_\infty, \ t = 0, 1, 2, \ldots \tag{2}$$

对于给定的 s_0 和阈值序列 $\{\lambda_t\}_{t=1}^\infty$，我们将估计量 $\{\beta^t\}_{t=1}^\infty$ 定义为

$$\beta^t = \mathcal{T}_{\lambda_t,s_0}\left(\beta^{t-1} + \frac{1}{n} X^\top (y - X\beta^{t-1})\right), \ t = 1, 2, \ldots. \tag{3}$$

此外，我们将 $\{\beta^t\}_{t=1}^\infty$ 的对应支撑集记为 $\{S^t\}_{t=1}^\infty$。在变量选择研究中，误判真实支撑集 S^*，即 $S^t \cap (S^*)^c$，称为第一类错误；而遗漏 S^*，即 $(S^t)^c \cap S^*$，称为第二类错误。我们的程序总结如下：

算法1 Double Sparse **I**HT (DSIHT)算法，其中已知 s、 s_0 和 σ。

要求： X， y， $\{G_j\}_{j=1}^m$， κ， λ_0， s_0， s， σ。

1: 初始化 $t = 0$、 $\beta^t = 0$ 和 $\lambda_\infty = 4\sqrt{\frac{\sigma^2}{n}(\log \frac{ed}{s_0} + \frac{1}{s_0}\log \frac{em}{s})}$。 2: **当** $\lambda_t \geq \lambda_\infty$时，**执行**3: $\beta^{t+1} = \mathcal{T}_{\lambda_t, s_0}\left(\beta^t + \frac{1}{n}X^\top (y - X\beta^t)\right)$。 4: $\lambda_{t+1} = \kappa\lambda_t$。 5: $t = t + 1$。

6: **结束循环 确保：** $\hat{} \beta = \beta^t$。

这里我们直观地解释 λ_t的选择。较大的 λ_t 会促进估计量 β^t的稀疏性，从而通过防止将伪变量纳入模型，显著降低第一类错误。然而，过度的稀疏性可能因遗漏过多真实变量而导致较高的第二类错误。如第2.2节所示，这会导致较高的估计误差，因为 β^* 的幅度会被大幅压缩至零。相反，较小的 λ_t 可以通过增加模型的复杂性来降低第二类错误。不过，这会使过多的伪变量进入模型，从而导致较高的第一类错误。这一直观解释促使我们通过平衡这两类错误来选择序列$\{\lambda_t\}_{t=1}^\infty$ 的具体形式。

在我们的程序中，我们采用递减序列 (2)，而不是直接将阈值设为这一阶。原因是，如此小的阈值可能会导致初始步骤选入过多不重要的变量。稀疏性的缺失使我们的程序难以利用 DSRIP 条件的收缩性质，并且无法在迭代过程中良好地控制估计误差。相比之下，足够大的 λ_0 能够识别出一组较小的变量，从而有效控制初始解中的假发现。随着阈值逐渐降低，我们在不损失稀疏性的前提下，通过迭代沿着适当的方向优化解。我们程序的一项创新在于，它能在每一步隐式地将第一类错误控制在较低水平，并通过迭代降低第二类错误。在定理 1 中，我们选择 $\lambda_\infty \asymp \sqrt{\frac{\sigma^2}{n}(\log \frac{ed}{s_0} + \frac{1}{s_0}\log \frac{em}{s})}$，并证明了其在极小极大意义下的最优性。

2.2 估计误差分析

为了进行 t th理论分析中，我们将迭代项分解为三部分：

$$\begin{aligned} H^{t+1} :=& \beta^t + \frac{1}{n}X^\top (y - X\beta^t) \\ =& \beta^* + \left(\frac{1}{n}X^\top X - \mathbb{I}_p\right)(\beta^* - \beta^t) + \frac{1}{n}X^\top \xi \\ =& \beta^* + \Phi(\beta^* - \beta^t) + \Xi, \end{aligned} \tag{4}$$

其中 $\Phi := \frac{1}{n} X^\top X - \mathbb{I}_p$ 和 $\Xi := \frac{1}{n} X^\top \xi$。方程（4）表明，估计误差来自三个方面：

- 真实参数 β^* 被错误地缩小了。
- 优化误差 β^t 近似于 β^*。

误差 ξ 导致的随机性。•

在这三个来源中，优化误差对应于迭代过程，而我们提出的程序中的随机性主要来自第三项 Ξ。接下来，我们将详细说明如何准确地给出后两种误差来源的上界。首先，我们介绍设计矩阵 X 的一个关键条件，以确保优化误差收缩。

假设 1（DSRIP 条件）如果 $X \in \mathbb{R}^{n \times p}$ 满足双重稀疏受限等距性质 $DSRIP(s, s_0, \delta)$，常数为 $0 < \delta < 1$，那么对于 $\forall S \in \mathcal{S}^{m,\ d}(s, s_0)$ 和 $\forall u \neq 0, u \in \mathbb{R}^{|S|}$，有：

$$1 - \delta \leq \frac{\|X_S u\|_2^2}{n\|u\|_2^2} \leq 1 + \delta.$$

注 1双重稀疏受限等距性质（*DSRIP*）是普通受限等距性质条件的一种自然扩展 *[Candes and Tao, 2005]* 适用于双重稀疏线性回归。对于次高斯设计，考虑 p维、 ss_0 稀疏结构时，我们要求样本量为 $n = \Omega(ss_0 \log \frac{ep}{ss_0})$，以确保 *RIP* 条件以高概率成立。然而，要满足 *DSRIP* 条件，我们只需要$n = \Omega(ss_0 \log \frac{ed}{s_0} + s\log \frac{em}{s})$。值得注意的是，在给定 $p = m \times d$的情况下，相较于 *RIP* 条件，*DSRIP* 条件可以用更小的样本量满足。更多详情请参见附录 *C*。

DSRIP 是分析高维双重稀疏线性回归 [Li等人， 2024] 的重要组成部分。它的条件比普通 RIP 更宽松。在元素级稀疏性相同的情况下，DSRIP 只要求满足至多包含 s 组的 ss_0-稀疏向量子集，而 RIP 则要求所有 ss_0-稀疏向量都满足条件。若设计矩阵 X 满足 DSRIP(s， s_0， δ)，则有 $\|\Phi\|_2 \leq \delta < 1$，这表明 Φ 是所有 (s， s_0)-稀疏向量的收缩因子。因此，利用 DSRIP 和信号的稀疏结构，收缩因子 Φ 能够逐步降低优化误差。

接下来，我们分析随机误差项 Ξ。为了给这一误差来源确定上界，我们需要刻画噪声项的复杂度。

引理 1 假设 X 满足 ***DSRIP***(s, s_0, δ)。那么，存在一个常数 $C > 0$，使得事件

$$\mathcal{E} \coloneqq \left\{ \forall S \in \mathcal{S}^{m,d}(s, s_0) : \sum_{i \in S} \Xi_i^2 \leq \frac{4\sigma^2}{n} \left(s s_0 \log \frac{ed}{s_0} + s \log \frac{em}{s} \right) \right\}$$

以至少 $1 - \exp\left\{ -C(s s_0 \log \frac{ed}{s_0} + s \log \frac{em}{s}) \right\}$ 的概率成立。

引理 1 以高概率给出了随机误差项的一致上界。现在我们详细分析随机项 Ξ，并将随机误差的来源分解为两部分:

- 附着于真实支撑集 S^* 的随机误差 Ξ。
- 由第一类错误（即对真实参数的误识别）导致的随机误差 Ξ β^*。更具体地说，一些随机误差逃逸出了算子 $\mathcal{T}_{\lambda, s_0}$，我们在下文将这些误差称为纯误差。

随机误差 Ξ 导致的误差可归因于两个来源：分别对应于 S^* 和 $(S^*)^c$ 的随机误差。由于 $S^* \in \mathcal{S}^{m,d}(s, s_0)$ 具有稀疏先验，附着于 S^* 的随机误差以高概率可由事件 $\mathcal{E}$ 良好地界定。然而，由于纯误差的数量尚未确定，很难找到纯误差的上界。因此，算子 $\mathcal{T}_{\lambda, s_0}$ 要解决的核心问题是界定纯误差的支撑集。直观地说，我们希望将纯误差收集到属于 $\mathcal{S}^{m,d}(s, s_0)$ 的某些子集中。这样，纯误差的大小就可以由事件 $\mathcal{E}$ 给出上界。

我们考虑将$\mathcal{T}_{\lambda, s_0}$ 直接应用于纯误差，并证明如果纯误差溢出 $\mathcal{S}^{m,d}(s, s_0)$，那么它将以高概率与 $\mathcal{E}$ 矛盾。根据 $\mathcal{S}^{m,d}(s, s_0)$的结构，我们将讨论分为两种情形:

情形1: 假设$\mathcal{T}_{\lambda, s_0}$ 选出的集合至多属于 s 个组，但其数量超过 ss_0。逐元素条件检查确保所有选出的条目都大于 λ。因此，对于该集合中任意一个基数为ss_0的(s, s_0)形状子集，这些子集的总量超过 $ss_0\lambda^2$。取 $\lambda \geq 2\sqrt{\frac{\sigma^2}{n}(\log \frac{ed}{s_0} + \frac{1}{s_0} \log \frac{em}{s})}$，可得 $ss_0\lambda^2 \geq 4\frac{\sigma^2}{n}(ss_0 \log \frac{ed}{s_0} + s \log \frac{em}{s})$，这以高概率与事件 $\mathcal{E}$ 矛盾。我们在图 1 中给出一个说明性示例。

情形 2: 假设由 $\mathcal{T}_{\lambda, s_0}$ 选定的集合位于多于 s 个组中，但在任意 s 个选定组内，选定条目的数量不超过 ss_0。逐组条件检查表明，每个选定组的规模都大于$s_0\lambda^2$。因此，由任意 s 个选定组构成的 (s, s_0) 形子集满足其总规模超过 $ss_0\lambda^2$。对于 $\lambda \geq 2\sqrt{\frac{\sigma^2}{n}(\log \frac{ed}{s_0} + \frac{1}{s_0} \log \frac{em}{s})}$，它

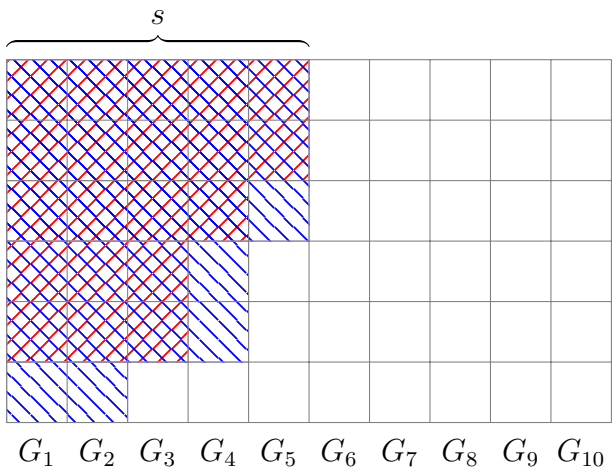

图 1：情形1的示意示例。共有 10 个组，每组大小相同，为 $d = 6$；我们将组结构重塑为一个 6×10 矩阵，每一列代表一个组。这里 $s = 5$ 和 $s_0 = 4$。蓝色区域表示选定集，红色区域表示一个（s， s_0）形状的子集，且其总量超过 $ss_0\lambda^2$。这里，红色区域集合的基数为 $s \times s_0 = 20$。注意，整个支撑集向量被重塑为具有特定组结构的矩阵。

以高概率与事件 $\mathcal{E}$ 相矛盾。我们在图 2 中提供了一个示例。值得注意的是，如果存在 s 个选定组，且其中选定条目的数量超过 ss_0，我们将以类似于**情形 1** 的方式分析这种情形。

总的来说，通过直接应用算子 $\mathcal{T}_{\lambda,\ s_0}$，$\Xi$ 以高概率可缩减为一个（s， s_0）形状的子集。

2.3 估计误差上界

在第2.2节中，我们介绍了控制由优化误差和随机性造成的估计误差的思路。严格来说，估计误差的三个来源可以依次得到界定。接下来，我们分析所提出程序的误差界。理论分析的主要结果由定理 1 给出。

定理 1 假设 β^* 是(s, s_0)-稀疏的，且 X 满足 $DSRIP(3s, \frac{5}{3}s_0, \delta)$。假设 $\delta < 0.11 \wedge \kappa^{10}$，$\|\beta^*\|_2 \leq \sqrt{ss_0}\lambda_0$ 和 $\lambda_\infty \geq 4\sqrt{\frac{\sigma^2}{n}(\log \frac{ed}{s_0} + \frac{1}{s_0}\log \frac{em}{s})}$。运行算法*1*，得到相应的解序列 $\{\beta^t\}, t = 1, 2, \cdots$。那么，以至少 $1 - \exp\left\{-C(ss_0 \log \frac{ed}{s_0} + s \log \frac{em}{s})\right\}$ 的概率，有

(i) 在组 G^* 内，第一类错误可以通过一个形状为 (s, s_0) 的子集来控制，即，

$$S_{G^*} \cap S^t \cap (S^*)^c \in \mathcal{S}^{m,d}(s, s_0). \tag{5}$$

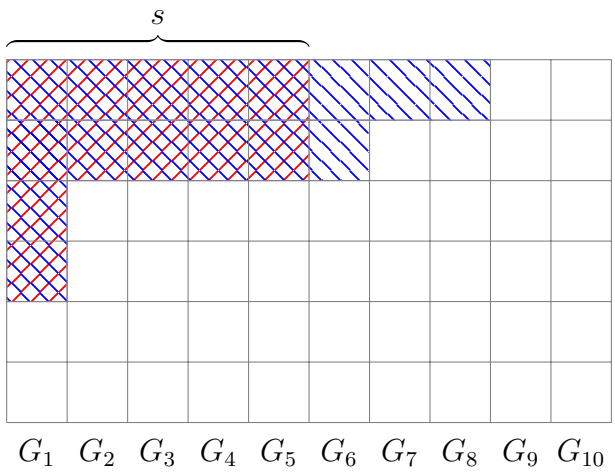

图 2：**情形 2**的示意图。图 2 中的元素与图 1 相同。红色区域中的条目覆盖 $s = 5$ 组，其基数小于$s \times s_0 = 20$。

(ii) 在组 G^* 之外，第一类错误可以由一个(s, s_0)-形子集控制，即，

$$S^c_{G^*} \cap S^t \in \mathcal{S}^{m,d}(s, s_0). \tag{6}$$

(iii) 估计误差的上界为

$$\|\beta^* - \beta^t\|_2 \leq \frac{3}{2}(1 + \sqrt{2})\sqrt{ss_0}\lambda_t. \tag{7}$$

定理 1 的 (i) 部分表明，$\{\beta^t\}$ 在真实组 G^* 内的第一类错误可以控制在一个 $(s,\ s_0)$ 形集合中。定理 1 的 (ii) 部分断言，我们的程序选入模型的错误组少于 s 个，且组外变量至多有 ss_0 个G^*。综合来看，这表明我们的程序生成的解序列 $\{\beta^t\}$ 在每一步都是 $(2s, \frac{3}{2}s_0)$-稀疏的，证实了我们的程序能够有效控制元素层面和组层面的假发现。IHT 风格方法的非凸性可能导致参数估计误差在每一步不一定都减小。为解决这一问题，一种常见做法是构造上界的代理函数，使其按指数速度递减 [Yuan 等，2018，Zhu 等，2020，Zhang 等人，2023]。选择 $\{\lambda_t\}$ 时，(7) 给出了参数估计误差的一个递减上界。值得注意的是，选择 $\lambda_\infty \asymp \sqrt{\frac{\sigma^2}{n}}(\log \frac{ed}{s_0} + \frac{1}{s_0}\log \frac{em}{s})$ 时，上界会几何收敛至 (8) 中的极小极大下界，这证明了我们的程序在极小极大意义下的最优性。

注 2 在上述讨论中，我们已经讨论了构造 $\mathcal{T}_{\lambda,s_0}$ 的思路。

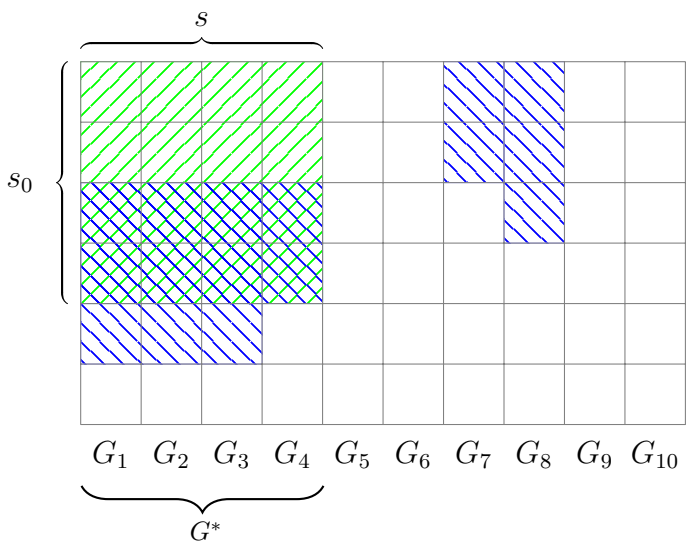

图 3：错误发现的两种情形示意图。这里 $G^* = \{G_1, G_2, G_3, G_4\}$和 $s = s_0 = 4$。绿色区域表示真实支撑集 S^*，蓝色区域表示选定集 S^t。图 3 中其余元素与图 1 相同。

通过将其直接应用于 Ξ。在我们的实际过程中，我们将 $\mathcal{T}_{\lambda,s_0}$ 应用于 H^t，而不是 Ξ。参考上述两种情况，我们可以证明，

(i) 在真实组 G^* 内，如果 $S^t \cap (S^*)^c \notin \mathcal{S}^{m,d}(s, s_0)$，则存在一个形状为 (s, s_0) 的子集 $\tilde{S}_{1,t} \subseteq S^t \cap S_{G^*} \cap (S^*)^c$，使得 $ss_0\lambda_{t+1}^2 \leq \sum_{i\in\tilde{S}_{1,t}} \{\mathcal{T}_{\lambda_{t+1,s_0}}(H^{t+1})\}_i^2$。

(ii) 在真实组 G^* 外，如果 $S^t \cap (S^*)^c \notin \mathcal{S}^{m,d}(s, s_0)$，则存在一个形状为 (s, s_0) 的子集 $\tilde{S}_{2,t} \subseteq S^t \cap S_{G^*}^c$，使得 $ss_0\lambda_{t+1}^2 \leq \sum_{i\in\tilde{S}_{2,t}} \{\mathcal{T}_{\lambda_{t+1,s_0}}(H^{t+1})\}_i^2$。

值得注意的是，我们的证明主要依赖于数学归纳法。假设定理 *1* 中的(5),(6),(7) 在步骤 t 中成立，我们首先利用归纳假设证明 (5) 和 (6) 在步骤 $t+1$ 中成立。随后，我们将归纳假设与步骤 $t+1$ 中的(5) 和(6)相结合，以证明(7)成立，从而完成归纳步骤。

注释 3 这里我们详细说明为何将错误发现的分析分为两种情况。随后，我们给出一个示例，说明在错误发现 $S^t \cap (S^*)^c$中，不存在满足 $\tilde{S}_t \subseteq S_{G^*}^c$ 的子集 $\tilde{S}_t$，使得 $ss_0\lambda_{t+1}^2 \leq \sum_{i\in\tilde{S}_t} \{\mathcal{T}_{\lambda_{t+1,s_0}}(H^{t+1})\}_i^2$。

在图 *3* 中，容易验证 $S^t \cap (S^*)^c$ 有 *8* 个条目，且 $S^t \cap (S^*)^c \notin \mathcal{S}^{m,d}(s, s_0)$，因为它覆盖了 *5* 个组。通过逐组条件检查，对于 $i = 7$，有 $\|\{\mathcal{T}_{\lambda_{t+1,s_0}}(H^{t+1})\}_{G_i}\|_2^2 \geq s_0\lambda_{t+1}^2$，*8*。另一方面，在 G^* 内，$S^t \cap (S^*)^c$ 的每个元素的绝对值都不小于 λ_{t+1}。然而，我们无法找到一个 (s, s_0) 形状的子集，使得 $ss_0\lambda_{t+1}^2 \leq \sum_{i\in\tilde{S}_t} \{\mathcal{T}_{\lambda_{t+1,s_0}}(H^{t+1})\}_i^2$。因此，我们考虑分别用两个 (s, s_0) 形状的子集，覆盖 G^* 内部和 G^* 外部的错误发现。

2.4 双重稀疏线性回归的极小极大下界

以往的研究已对高维稀疏线性回归的极小极大速率进行了深入探讨。许多论文关注逐元素 s-稀疏类[Raskutti 等人，2011年，Verzelen, 2012，Bellec 等人， 2018], ；另有一些研究致力于组稀疏性，例如 Huang 和 Zhang [2010] 以及 Lounici 等人 [2011]。近期，Cai 等人 [2022] 给出了双重稀疏线性回归的非渐近极小极大下界。本文将用一种更简洁的技术证明这一结果。考虑参数空间$\sim\Theta^{m,\ d}(s,\ s_0)$:

$$\widetilde{\Theta}^{m,d}(s,s_0) \coloneqq \{\beta \in \mathbb{R}^p : \|\beta\|_{0,2} \le s \text{ and } \|\beta_{G_j}\|_0 \le s_0, \forall j \in [m]\}.$$

$\widetilde{}$与 $\Theta^{m,d}(s,s_0)$ 不同， $\Theta^{m,d}(s,s_0)$ 对每个组施加半径为s_0的 ℓ_0-球约束。此外， $\widetilde{\Theta}^{m,d}(s,s_0)$ 的总稀疏性限制为 ss_0。容易观察到$\widetilde{}$， $\Theta^{m,d}(s,s_0) \subseteq \Theta^{m,d}(s,s_0)$。因此，

$$\inf_{\hat{\beta}} \sup_{\beta^* \in \Theta^{m,d}(s,s_0)} \mathbf{E}_{\hat{\beta}} \|\hat{\beta} - \beta^*\|_2^2 \ge \inf_{\hat{\beta}} \sup_{\beta^* \in \widetilde{\Theta}^{m,d}(s,s_0)} \mathbf{E}_{\hat{\beta}} \|\hat{\beta} - \beta^*\|_2^2,$$

$\hat{}$ 其中$\mathbf{E}_{\hat{\beta}}$ 表示关于 β 的期望。

定义 2（装填数）集合 $\mathcal{S}$ 关于度量 $\|\cdot\|_\psi$ 的一个 ρ-装填是一个集合族 $\{\beta^1, \ldots, \beta^M\} \subset \mathcal{S}$，满足对于所有不同的 i、$j \in [M]$，均有 $\|\beta^i - \beta^j\|_\psi > \rho$。$\rho$-装填数 $M(\delta;\ \mathcal{S}, \|\cdot\|_\psi)$ 是最大 ρ-装填的基数。

$\widetilde{}$令 $M(\rho;\ \Theta^{m,\ d}(s,\ s_0)\ \|\cdot\|_H)$ 表示参数空间$\widetilde{\Theta}^{m,\ d}(s,\ s_0)$ 中关于汉明度量 $\|\cdot\|_H$ 的 ρ-打包集的基数。关于 $\Theta^{m,\ d}(s,\ s_0)$ 的打包数下界如下所示。

引理 2（打包数的下界 [Li等人， 2024]） $\frac{ss_0}{4}$-打包集的基数 $\widetilde{\Theta}^{m,\ d}(s,\ s_0)$ 的下界为 —

$$\log\left(M(\frac{ss_0}{4}; \widetilde{\Theta}^{m,d}(s,s_0), \|\cdot\|_H)\right) \ge \frac{ss_0 \log\frac{ed}{s_0} + s\log\frac{em}{s}}{4}.$$

Li等人 [2024] 利用双重稀疏性的结构，并结合多元 Gilbert-Varshamov 界 [Gilbert, 1952] 以更简洁的方式构造了 $\widetilde{\Theta}^{m,\ d}(s,\ s_0)$ 的打包集。结合引理 2，我们建立了一个与 Cai 等人 [2022]提出的结果一致的极小极大下界。该结论见以下定理。

定理 2 考虑线性回归模型 $y = X\beta^* + \varepsilon$，其中 $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathrm{I}_n)$。将最大$(2s, 2s_0)$-稀疏特征值记为

$$\vartheta_{\max} = \max_{u \in \Theta^{m,d}(2s,2s_0)} \frac{\|Xu\|_2}{\sqrt{n}\|u\|_2}.$$

假设 $\vartheta_{\max} < \infty$。那么，我们有

$$\inf_{\hat{\beta}} \sup_{\beta^* \in \Theta^{m,d}(s,s_0)} \mathbf{E}_{\hat{\beta}} \|\hat{\beta} - \beta^*\|_2^2 \geq \frac{\sigma^2}{512\vartheta_{\max}^2 n} \left(s s_0 \log \frac{ed}{s_0} + s \log \frac{em}{s} \right). \tag{8}$$

定理 2 建立了 ℓ_2 估计误差的下界，这与 Cai 等人 [2022]的结果一致。 $\beta^{t\infty}$ 的估计误差达到了极小极大下界 (8)，证明了我们提出的 IHT 风格程序具有最优性。

3 一个完全自适应最优程序

第 2 节提出的程序依赖于未知的稀疏度 s， s_0和噪声水平 σ，这给实际应用带来了挑战。为了解决这一问题，我们采用数据驱动方法来确定程序的初始阈值和最优停止时间，从而使其更适用于实际场景。给定 s_0，我们在第 3.1 节提出一种能够适应未知 s 和 σ 的程序。在第 3.2 节中，我们探讨经典 IHT 与组 IHT 在不同 s_0值下的权衡。最后，我们提出一种针对 s_0 的数据自适应调参方法，并证明其最优性，使我们的方法成为完全自适应程序。

3.1 对未知 s 和 σ的自适应

在第 3.1 节的其余部分，我们假设稀疏度 s_0 已知。首先，我们介绍初始阈值 λ_0的自适应选择。定理 1 的假设为 λ_0的选择提供了一个下界。然而，从优化的角度来看，选择过大的 λ_0 值可能会降低算法的效率，因为这可能导致更多冗余迭代。在本文其余部分，记

$$M := \frac{1}{n} X^\top y = \beta^* + \Phi\beta^* + \Xi \quad \text{and} \quad \sigma_t^2 := \frac{1}{n}\|y - X\beta^t\|_2^2.$$

我们给出 λ_0 的显式形式如下

$$\lambda_0 := \frac{100}{9} \sqrt{\frac{\sigma_0^2}{n} \left(\log \frac{ed}{s_0} + \frac{1}{s_0} \log em \right)} \vee \frac{19}{4} \|M\|_\infty, \tag{9}$$

其中 $\|M\|_\infty := \max_i |M_i|$。

定理 3 假设 β^* 是(s, s_0)-稀疏的，且 X 满足 $DSRIP(2s, \frac{3}{2}s_0, \delta)$。假设 $\delta < 0.11$ 且 $n > 105^2(ss_0 \log \frac{ed}{s_0} + s\log em)$。那么，以至少 $1 - \exp\{-C(ss_0 \log \frac{ed}{s_0} + s \log \frac{em}{s})\}$ 的概率，有 $\|\beta^*\|_2 \leq ss_0\lambda_0$。

定理 3 表明，选择式 (9) 可保证以高概率满足定理 1 中的假设。接下来，我们定义三个停止时间 t_∞, t_0 和 $\bar{t}$ 如下

$$
\begin{aligned}
t_\infty &:= \inf\left\{t : \lambda_t \le 4\sqrt{\frac{\sigma^2}{n}\left(\log\frac{ed}{s_0} + \frac{1}{s_0}\log\frac{em}{s}\right)}\right\},\\
t_0 &:= \inf\left\{t : \lambda_t \le 12\sqrt{\frac{\sigma^2}{n}\left(\log\frac{ed}{s_0} + \frac{1}{s_0}\log em\right)}\right\},\\
\bar{t} &:= \inf\left\{t : \lambda_t \le 8\sqrt{\frac{\sigma_t^2}{n}\left(\log\frac{ed}{s_0} + \frac{1}{s_0}\log em\right)}\right\}.
\end{aligned}
\tag{10}
$$

t_∞ 是达到最优阈值 λ_∞的停止时间。显然， t 是一个不依赖于 s 和 σ的可达停止时间。另一方面， t_0 和 t_∞ 是分别对应于未知参数 s 和 σ的理论停止时间。我们在下述定理中说明这三个停止时间之间的关系。

定理 4 假设定理 *1* 中的所有条件均成立，且样本量为 $n > 105^2(ss_0\text{对数}\frac{ed}{s_0} + s\text{对数}\ em)$。那么，以至少$1 -$ 指数$\left\{-C(ss_0\text{对数}\frac{ed}{s_0} + s\,\text{对数}\ \frac{em}{s})\right\}$的概率，我们有

$$
t_0 \le \bar{t} \le t_\infty.
$$

定理 4 表明， t 可以由理论停止时间 t_0 和 t_∞界定。特别是，由于 t 被最优停止时间 t_∞控制，估计误差 $\|\beta^{\bar{t}} - \beta^*\|$ 可以由（7）给出上界。此外，定理 1 意味着 β^{t_0} 在极小极大意义下是次优的。更具体地说，我们可以推得 $\beta^{\bar{t}}$ 达到的统计精度在对数因子范围内是最优的。我们将这一极小极大次优结果表述为推论 1。

推论 1 假设引理 *4* 中的条件成立。那么，我们有

$$
\sup_{S^*\in\mathcal{S}^{m,d}(s,s_0)} P\left(\|\beta^{\bar{t}} - \beta^*\|_2 \ge 50\sqrt{\frac{\sigma^2}{n}\left(ss_0\log\frac{ed}{s_0} + s\log em\right)}\right) \le e^{-C(ss_0\log\frac{ed}{s_0} + s\log\frac{em}{s})}.
$$

推论 1 是定理 4 的直接推论。它表明，在 t 处停止是一种极小极大次优程序。下一个悬而未决的问题是，我们能否改进这一极小极大次优程序，使其达到极小极大最优。下面的分析表明，在某些条件下，答案是肯定的。记

$$
\Omega(\beta) := s_0\|\beta\|_G\log\frac{ed}{s_0} + \|\beta\|_G\log\frac{em}{\|\beta\|_G},
\tag{11}
$$

其中 $\|\beta\|_G := \|\beta\|_{0,\ 2} \vee \frac{\|\beta\|_0}{s_0}$。我们考虑 Birgé-Massart 准则的一个变体 [Birgé 和 Massart, 2001]：

$$\tilde{t} = \arg \min_{t \in [T] \setminus [\bar{t}-1]} \left\{ \frac{1}{n} \left\| y - X\beta^t \right\|_2^2 + \frac{1000\sigma_{\bar{t}}^2 \Omega(\beta^t)}{n} \right\}, \tag{12}$$

其中 $T := \inf\{t:\ \lambda_t \leq 4\frac{\sigma_{\bar{t}}}{\sqrt{n}}\} + 1$. 此处，停止时间 T 取大于 t_∞ 的值，以确保搜索范围足够大。一旦迭代达到次优停止时间 $\bar{t}$，我们便根据 (12) 开始选择最优迭代。现在，我们可以在算法 2 中给出所提出的自适应程序的详细伪代码。

算法 2 Double Sparse IHT (DSIHT) 算法（已知 s_0）

要求： $X,\ y,\ \{G_j\}_{j=1}^m,\ \kappa,\ s_0$。
1：初始化 $t = 0, \beta^0 = 0$ 和 $\lambda_0 = \frac{100}{9}\sqrt{\frac{\sigma_0^2}{n}(\log \frac{ed}{s_0} + \frac{1}{s_0}\log em)}$ $\vee \frac{19}{4}\|M\|_\infty$。2：当 $\lambda_t \geq 8\sqrt{\frac{\sigma_t^2}{n}(\log \frac{ed}{s_0} + \frac{1}{s_0}\log em)}$ 时，执行 3：$\beta^{t+1} = \mathcal{T}_{\lambda_t, s_0}\left(\beta^t + \frac{1}{n}X^\top(y - X\beta^t)\right)$。4：$\lambda_{t+1} = \kappa\lambda_t$。5：$t = t + 1$。6：结束 while 7：计算 $\sigma_{\bar{t}}^2 = \frac{1}{n}\|y - X\beta^t\|_2^2$。8：当 $\lambda_t \geq \frac{4\sigma_{\bar{t}}}{\sqrt{n}}$ 时，执行 9：计算 $\mathsf{C}_t = \frac{1}{n}\|y - X\beta^t\|_2^2 + \frac{1000\sigma_{\bar{t}}^2}{n}\Omega(\beta^t)$。10：$\beta^{t+1} = \mathcal{T}_{\lambda_t, s_0}\left(\beta^t + \frac{1}{n}X^\top(y - X\beta^t)\right)$。11：$\lambda_{t+1} = \kappa\lambda_t$。12：$t = t + 1$。13：结束 while 14：$\tilde{t} = \operatorname{argmin}_{t} \mathsf{C}_t$。确保：$\hat{\beta} = \beta^{\tilde{t}}$。

算法 2 依赖参数 s_0，并消除了对未知值 s 和 σ的依赖。停止时间 t 的最优结果如下所示。

定理 5 假设 β^* 是(s, s_0)-稀疏的，且 X 满足 $DSRIP(5s, s_0, \delta)$。假设 $\delta < 0.11 \wedge \kappa^{10}$ 和 $n > 1000^2(ss_0 \log \frac{ed}{s_0} + s\log em)$。那么，我们有

$$\sup_{S^* \in \mathcal{S}^{m,d}(s,s_0)} P\left(\|\beta^{\tilde{t}} - \beta^*\|_2 \geq 150\sqrt{\frac{\sigma^2}{n}\left(ss_0 \log\frac{ed}{s_0} + s\log\frac{em}{s}\right)} \right) \leq e^{-C(ss_0 \log \frac{ed}{s_0} + s\log\frac{em}{s})},$$

and

$$\sup_{S^* \in \mathcal{S}^{m,d}(s,s_0)} P\left(\|\beta^{\tilde{t}}\|_G \geq 4s \right) \leq e^{-C(ss_0 \log \frac{ed}{s_0} + s\log\frac{em}{s})}.$$

˜定理 5 给出了 β^t 估计误差的上界，表明 β^t 自适应地达到了极小极大最优收敛速率。此外，定理 5˜ 证明了我们的程序能够以高概率保证估计量 β^t 的稀疏性。具体而言，我们可以将模型规模 $\|\beta^t\|_0$ 控制在 $O(ss_0)$˜ 的阶内，并将选定的组数$\|\beta^t\|_{0,\,2}$ 控制在 $O(s)$ 的阶内。

推论2 假设定理 *5* 中的所有条件均成立。对于 *(12)* 中的停止时间 T，我们有

$$\sup_{S^*\in\mathcal{S}^{m,d}(s,s_0)} P\left(T \geq \log\left(6(\frac{\sqrt{n}\|\beta^*\|_2}{\sigma} \vee \sqrt{\log ep})\right)/\log(1/\kappa)+1\right) \leq e^{-C(ss_0\log\frac{ed}{s_0}+s\log\frac{em}{s})}.$$

推论2保证，我们的 IHT 过程能够以较高概率实现最优统计精度和线性收敛，这证明了我们所提出方法的高效性。

3.2 IHT 与组IHT 之间的自适应权衡

在本节中，我们研究 s_0 的错设问题，而这在现实应用中通常是不可观测的。令 $\bar{s}_0$ 为算法 2 中的输入参数。值得注意的是，给定样本 (X, y) 和步长 κ 后，估计量 β 完全由算法 2 中的 $\bar{s}_0$ 决定。因此，我们引入以下基于给定 $\bar{s}_0$ 的算法 2 所导出的统计量：

- $\hat{\beta}(\bar{s}_0)$ 表示给定 $\bar{s}_0$ 时算法 2 的估计量。ˆ
- ˆ$s(\bar{s}_0)$ 表示 $\beta(\bar{s}_0)$ 的选定组数。ˆˆ
- $A(\bar{s}_0)$ 表示 $\beta(\bar{s}_0)$ 的非零元素个数。

根据 (s, s_0)-稀疏性和参数空间 $\mathcal{S}^{m,d}(s, s_0)$ 的定义，我们建立如下关系

$$\mathcal{S}^{m,d}(s,s_0) \subseteq \begin{cases} \mathcal{S}^{m,d}(ss_0/\bar{s}_0, \bar{s}_0), & \bar{s}_0 \leq s_0. \\ \mathcal{S}^{m,d}(s, \bar{s}_0), & \bar{s}_0 > s_0. \end{cases}$$

一方面，当 $\bar{s}_0 > s_0$ 属于算法 2 时，设计矩阵 X 满足 DSRIP$(5s,\ \bar{s}_0,\ \delta)$，且 β^* 是 $(s,\ \bar{s}_0)$-稀疏的。算法 2 可以获得关于参数空间 $\mathcal{S}^{m,\,d}(s,\ \bar{s}_0)$ 的极小极大最优估计量，并保留从定理 1 到推论2 的所有既有理论结果。另一方面，给定算法 2 中的 $\bar{s}_0 \leq s_0$，若设计矩阵 X 满足 DSRIP$(5ss_0/\bar{s}_0,\ \bar{s}_0,\ \delta)$，且 β^* 是 $(ss_0/\bar{s}_0,\ \bar{s}_0)$-稀疏的，则算法 2 可以获得关于参数空间 $\mathcal{S}^{m,\,d}(ss_0/\bar{s}_0,\ \bar{s}_0)$ 的极小极大最优估计量，并保留所有既有理论结果。我们在表1中总结这些结果：

表1： s_0错设模型的性质。

值	参数空间	极小极大速率	支撑集控制
$\bar{s}_0 < s_0$	$\mathcal{S}^{m,\ d}(ss_0/\bar{s}_0,\ \ \bar{s}_0)$	$\sqrt{\frac{\sigma^2}{n}\left(ss_0 \text{对数} \frac{ed}{\bar{s}_0}\right)}$ — — $+\frac{ss_0}{\bar{s}_0}\text{对数}\frac{em\bar{s}_0}{ss_0}\Big)$	$\hat{}A(\bar{s}_0) \lesssim ss_0$
$\bar{s}_0 = s_0$	$\mathcal{S}^{m,\ d}(s,\ \ s_0)$	$\sqrt{\frac{\sigma^2}{n}\left(ss_0 \text{对数} \frac{ed}{s_0}\right)}$ — — $+s\text{对数}\frac{em}{s}\Big)$	$\hat{}A(\bar{s}_0) \lesssim ss_0,$ $\hat{s}(\bar{s}_0) \lesssim s$
$\bar{s}_0 > s_0$	$\mathcal{S}^{m,\ d}(s,\ \ \bar{s}_0)$	$\sqrt{\frac{\sigma^2}{n}\left(s\bar{s}_0 \text{对数} \frac{ed}{\bar{s}_0} + s \text{对数} \frac{em}{s}\right)}$ — — —	$\hat{}s(\bar{s}_0)$ $\lesssim s$

表1表明，在 $\bar{s}_0 > s_0$ 和 $\bar{s}_0 < s_0$ 两种情况下，理论性质存在显著差异。当 $\bar{s}_0 < s_0$ 时，估计误差的上界为 $\sqrt{\frac{\sigma^2}{n}\left(ss_0 \log \frac{ed}{\bar{s}_0} + \frac{ss_0}{\bar{s}_0}\log \frac{em\bar{s}_0}{ss_0}\right)}$，模型规模可以控制在 $O(ss_0)$ 的量级内。在 $\bar{s}_0 > s_0$ 的情况下，上界为 $\sqrt{\frac{\sigma^2}{n}\left(ss_0 \log \frac{ed}{s_0} + s \log \frac{em}{s}\right)}$，选定群组的数量可以控制在 $O(s)$ 的量级内。值得注意的是，无论是 $\bar{s}_0 < s_0$ 还是 $\bar{s}_0 > s_0$，都无法同时控制元素层面和组层面的稀疏性。

我们在图 4 中展示了 s_0 从 1 到 d 变化时的极小极大速率。如图 4 所示，当 $1 \le \bar{s}_0 < s_0$时，极小极大速率趋于呈反比例函数关系。另一方面，当 $s_0 < \bar{s}_0 \le d$时，极小极大速率则呈现近似线性增长趋势。值得注意的是，对于 $\bar{s}_0 = s_0$，极小极大速率在这些值中达到最小值。

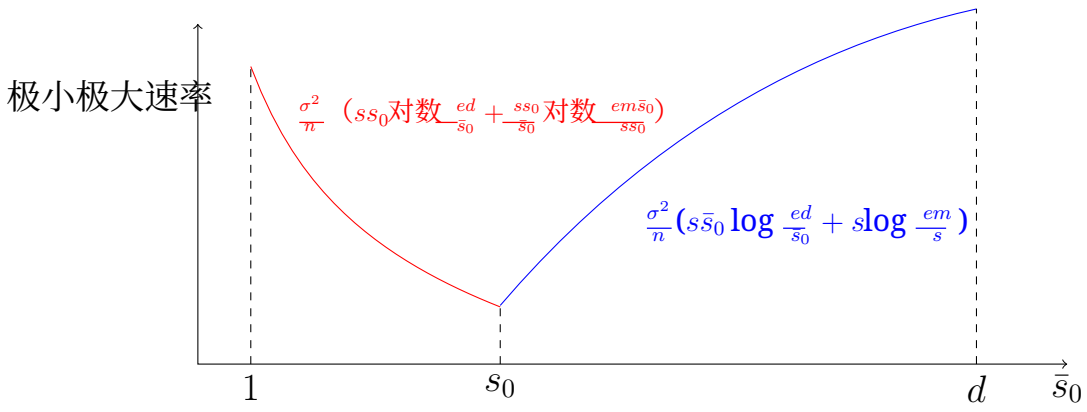

图 4：不同参数空间下度量 $\|\cdot\|_2^2$ 的极小极大速率。

注 4 无论 $\bar{s}_0$取何值，上述结果均给出了算法 *2* 的估计误差上界和稀疏性控制性质。特别地，当$\bar{s}_0 = 1$时，*DSIHT* 算法退化为经典 *IHT* 算法 *[Ndaoud, 2020]*, ，且表*1*中的结果恢复了极小极大速率 $O(\sqrt{\frac{\sigma^2}{n} s s_0 \log \frac{ep}{ss_0}})$ *[Raskutti* 等人*, 2011]*。当 $\bar{s}_0 = d$时，表*1*中的结果恢复了组稀疏性的极小极大速率，即$O(\sqrt{\frac{\sigma^2}{n}(sd + \overline{s \log \frac{em}{s}})})$ *[Lounici* 等人*, 2011]*。因此，*DSIHT* 可视为 *IHT [Ndaoud, 2020]*与组 *IHT [Giraud, 2021]*之间由参数 s_0决定的折衷。

3.3 未知 s_0 的数据自适应调参

前几节介绍了一种自适应程序，用于处理 s 和 σ 未知的情形。本节重点构造一个无需预先了解 s_0、即可达到极小极大最优性的自适应估计量，进一步说明我们的方法（参见算法 3）是一种完全自适应算法。

给定一个序列 $\{s_{0,l}\}_{l=1}^{L}$，确定最优选择的一种直观方法是将 s_0 视为调参参数。具体而言，沿该序列运行 DSIHT 算法，并采用模型选择准则确定最优模型规模。这里，我们采用 Verzelen [2012] 提出的 Birgé-Massart 准则的一种变体。与准则 (12) 中的做法不同，该变体隐式地利用了 σ^2 的信息，而非代入 σ 的同阶估计量。受此启发，我们提出一种新的双重稀疏信息准则（DSIC），如下所示，其中 $A(\bar{s}_0)$ 和 $\hat{s}(\bar{s}_0)$ 在第 3.2 节开头定义：

$$\mathrm{DSIC}(\bar{s}_0) = \log\left(\frac{\|y - X\hat{\beta}(\bar{s}_0)\|_2^2}{n}\right) + \frac{K}{n}\left(\hat{A}(\bar{s}_0)\log ed + \hat{s}(\bar{s}_0)\log\frac{em}{\hat{s}(\bar{s}_0)}\right), \tag{13}$$

其中 K 是一个正常数。使 (13) 最小化的估计量 $\hat{\beta}^{\bar{s}_0}$ 是我们程序的最优解。算法总结如下：

算法 3 自适应 双重 稀疏 IHT （ADSIHT）算法

需要： X, y, $\{G_j\}_{j=1}^{m}$, κ, $\{s_{0,l}\}_{l=1}^{L}$。1：对于 $l = 1, \ldots, L$，执行
2：$\hat{\beta}^{l} =$ 算法 2($X, y, \{G_j\}_{j=1}^{m}, \kappa, s_{0,l}$)。3：计算双重稀疏信息准则 DSIC($s_{0,l}$)。4：结束循环 5： $l^* =$argmin$_{l\in[L]}$\{DSIC($s_{0,l}$)\}。确保：
$\hat{\beta} = \hat{\beta}^{l^*}$。

注释 5 如第 *3.2* 节所述，要实现最优统计性能，需要우우우우우우

这意味着 $\bar{s}_0$ 与 s_0 处于同一数量级。遵循 ***Bellec*** 等人 *[2018],* 的方法，我们将 s_0 的候选值设为指数序列 $\{s_{0,l}\}_{l=1}^{L}=\left\{2^{\frac{l-1}{2}}, 1 \leq l \leq L\right\}$，其中 $L:=$最大 $\left\{l \in \mathbb{N}: 2^{\frac{l-1}{2}} \leq d\right\}$。此设置确保候选集合中包含一个与 s_0 处于同一数量级的值。回顾一下，***Cai*** 等人 *[2022]* 为未知参数 s 和 s_0 引入了候选集合，并采用网格搜索技术进行调优。相比之下，算法 *3* 只需要为 s_0 设置一个含有 $O(\log d)$ 个元素的候选集合，因此是一种计算效率高得多的调优方法。

在给出理论结果之前，我们需要对样本量和设计矩阵作出一些假设。首先，我们假设存在一个区间 $\mathcal{S}_0:=[s_{0,\min}, s_{0,\max}]$，使得 $s_0 \in \mathcal{S}_0$。

假设 2（样本量假设）我们假设样本量 n满足$n \gtrsim \left\{\left(ss_0 \text{对数} ed+\frac{ss_0}{s_0}{}_{,\min} \text{对数} em\right) \vee \left(ss_{0,\max} \text{对数} \overline{ed}+s \text{对数} em\right)\right\}$。

假设 2 是在噪声水平未知时实现极小极大自适应所必需的技术假设（Verzelen, 2012; Giraud 等， 2012]）。此外，我们要求 DSRIP 条件对 $\mathcal{S}_0$中的每个元素均成立。

假设 3（自适应 DSRIP 条件）我们假设设计矩阵 X同时满足 ***DSRIP***$(5s, s_{0,\max}, \delta)$ 和 ***DSRIP***$(5ss_0/s_{0,\min}, s_{0,\min}, \delta)$。

注 6 特别地，当 $s_{0,\min}$ 相对较小时，尤其是对于 $s_{0,\min}=1$，我们观察到 ***DSRIP***$(5ss_0/s_{0,\min}, s_{0,\min}, \delta)$ 退化为经典的 ***RIP*** 条件 ***[Candes and Tao,****2005]*。反之，当 $s_{0,\max}=d$时，***DSRIP***$(5s, s_{0,\max}, \delta)$ 则变为组 ***RIP*** 条件*[****Eldar and Mishali,*** *2009]*。

现在，我们在以下定理中给出极小极大自适应结果：

定理 6 假设 β^* 是(s, s_0)-稀疏的。给定区间 $\mathcal{S}_0$，假设假设 *2* 和 *3* 成立，且 $\delta<0.11 \wedge \kappa^{10}$。令 $\hat{} s_0=\operatorname{argmin}_{\bar{s}_0 \in \mathcal{S}_0}$ ***DSIC***$(\bar{s}_0)$，其中 K 足够大。则以大于 $1-\exp\left\{-C_1(ss_0 \log(ed/s_0)+s \log(em/s))\right\}$ 的概率，有如下结论：

$$\left\|\hat{\beta}(\hat{s}_0)-\beta^*\right\|_2 \leq C_2 \sigma \sqrt{\frac{ss_0 \log ed+s \log(em/s)}{n}}. \tag{14}$$

定理 6 表明，我们的自适应程序，即算法 3，是一种最优的完全自适应过程。重要的是，算法 3 无需知道 s、 s_0 和 σ，即可获得极小极大自适应解。

注 7 适应 s_0 的重要性在于能够在经典 ***IHT*** *[****Ndaoud,*** *2020]* 与组 ***IHT*** *[****Giraud,*** *2021]* 之间实现最优权衡。若假设 *2* 和

3 均成立，则可以达到这种最优权衡。需要强调的是，当 s_0 未知时，无法同时控制逐元素稀疏性和逐组稀疏性。因此，我们推导出了自适应估计量的近乎最优估计误差界。更多细节请参见定理 *6* 的证明。

4 满足 beta-min 条件的预言机估计速率

众所周知，支撑集为真实支撑集 S^* 的普通最小二乘（OLS）估计量可以达到 $O(\sigma\sqrt{\frac{ss_0}{n}})$ 的预言机估计速率。在本节中，我们将在 beta-min 条件下证明，DSIHT 算法同样能够达到预言机估计速率。这意味着，通过 DSIHT 获得的估计量与预言机 OLS 估计量的表现同样出色。此外，在 beta-min 条件下，DSIHT 还展现出近乎完全恢复 [Butucea等人， 2018] 真实支撑集 S^* 的能力。

记

$$\tilde{\lambda}_a := a\sqrt{\frac{8\sigma^2}{n}\left(\log\frac{ed}{s_0}+\frac{1}{s_0}\log\frac{em}{s}\right)},\ a>0. \tag{15}$$

给定初始估计量 $\tilde{\beta}^0$，我们在 DSIHT 算子 $\mathcal{T}_{\tilde{\lambda}_2,s_0}$ 中使用固定阈值 $\tilde{\lambda}_2$ 更新估计量。具体而言，我们按照以下方式更新系数:

$$\tilde{\beta}^{t+1}=\mathcal{T}_{\tilde{\lambda}_2,s_0}\left(\tilde{\beta}^t+\frac{1}{n}X^\top(y-X\tilde{\beta}^t)\right). \tag{16}$$

记 $\tilde{S}^t$ 为 $\tilde{\beta}^t$ 的支撑集。以下定理研究了采用固定阈值的迭代程序的理论保证。

定理 7 假设对于任意常数 $\epsilon>0$，均有 $\min_{i\in S^*}|\beta_i^*|\geq(\sqrt{2}+\epsilon)\tilde{\lambda}_2$ 和 $\min_{j\in G^*}\|\beta^*_{G_j}\|_2\geq(\sqrt{2}+\epsilon)\sqrt{s_0}\tilde{\lambda}_2$。假设 X 满足 $DSRIP(3s,\frac{5}{3}s_0,\delta)$ 和 $\delta\leq\epsilon^4\wedge 0.05$。令 $\tilde{\beta}^0$ 为满足定理 *1* 中 (5)-(7) 的初始估计量。我们运行 (16)，并得到相应的解序列 $\{\tilde{\beta}^t\}$。那么，对于 $\forall t\geq 0$，当 $\min\left\{\log\frac{ed}{s_0}+\frac{1}{s_0}\log\frac{em}{s},\ ss_0\log\frac{ed}{s_0}+\frac{1}{s_0}\log\frac{em}{s}\right\}\to\infty$ 时，以趋于 1[1]，的概率，有

(i) $S^c_{G^*}\cap\tilde{S}^t\in\mathcal{S}^{m,d}(s,s_0)$.

(ii) $S_{G^*}\cap(S^*)^c\cap\tilde{S}^t\in\mathcal{S}^{m,d}(s,s_0)$.

(iii) 估计误差的上界满足

$$\left\|\tilde{\beta}^t-\beta^*\right\|_2<\ 16\left(\frac{3}{4}\right)^t\sqrt{\frac{\sigma^2}{n}\left(ss_0\log\frac{ed}{s_0}+s\log\frac{em}{s}\right)}+16\sqrt{\frac{\sigma^2ss_0}{n}}. \tag{17}$$

[1]具体而言，当 $\Delta:=\frac{1}{s_0}\log(em/s)+\log(ed/s_0)$ 足够大时，该概率大于$1-C_1\exp(-C_2ss_0/\Delta)-C_3\Delta^2\exp(-C_4\Delta)$。定理 8 的尾概率也同样如此。

如定理 7 所示，固定迭代过程保留了错误发现控制的结果。具体而言，在满足 beta-min 条件时，结果 (17) 表明，估计误差的上界可分解为两个部分：一个逐渐减小的优化误差 $16\left(\frac{3}{4}\right)^t\sqrt{\frac{\sigma^2}{n}}(ss_0\log\frac{ed}{s_0}+s\log\frac{em}{s})$，它会随着 $t\to\infty$ 趋于零；以及一个统计误差 $16\sqrt{\frac{\sigma^2 ss_0}{n}}$。当优化误差小于统计误差时，项 $O\left(\sqrt{\frac{\sigma^2 ss_0}{n}}\right)$ 主导估计误差。

作为定理 7 的推论，当 t 足够大时，估计量 β^t 能够达到预言机估计速率，并且几乎能够在元素层面和组层面恢复真实支撑集。为了阐明这一性质，我们将逐元素解码器 $\eta^*\in\{0,1\}^p$记为 $\eta_i^*=\mathbf{I}(\beta_i^*\neq 0)$，并将逐组解码器 $\eta_G^*\in\{0,1\}^m$ 记为 $(\eta_G^*)_j=\mathbf{I}(\beta_{G_j}^*\neq\mathbf{0})$。对于 β^t，记 $\tilde{\eta}^t\in\{0,1\}^p$为 $\tilde{\eta}_i^t=\mathbf{I}(\beta_i^t\neq 0)$，并将逐组解码器 $\tilde{\eta}_G^t\in\{0,1\}^m$ 记为$(\tilde{\eta}_G^t)_j=\mathbf{I}(\beta_{G_j}^t\neq\mathbf{0})$。

定理8 假设定理 *7* 中的所有条件均成立。对于 $\forall t>2\log\left(256(\log\frac{ed}{s_0}+\frac{1}{s_0}\log\frac{em}{s})\right)$，当 $\min\left\{\log\frac{ed}{s_0}+\frac{1}{s_0}\log\frac{em}{s},\ \frac{ss_0\log\frac{ed}{s_0}+\frac{1}{s_0}\log\frac{em}{s}}{}\right\}\to\infty$ 时，以趋于 *1* 的概率，有：

(i) 估计量 β^t 满足

$$\left\|\tilde{\beta}^t-\beta^*\right\|_2\leq 17\sigma\sqrt{\frac{ss_0}{n}}. \tag{18}$$

(ii) 估计量 β^t 实现组级近乎完全恢复，即

$$\|\tilde{\eta}_G^t-\eta_G^*\|_0=o(s). \tag{19}$$

(iii) 估计量 β^t 实现逐元素几乎完全恢复，即 ，

$$\|\tilde{\eta}^t-\eta^*\|_0=o(ss_0). \tag{20}$$

定理 8 表明，经过足够多次迭代后，β^t 可达到预言机估计速率。值得注意的是，Bellec [2018] 证明，即使满足 beta-min 条件，凸估计量也无法达到预言机估计速率。这凸显了我们的 DSIHT 算法优于稀疏组Lasso。此外，beta-min 条件还能确保 [Butucea等人 2018] 在逐元素和逐组层面均实现近乎完全恢复。具体而言，我们可以控制逐元素和逐组层面的第一类错误与第二类错误，其阶分别为 $o(ss_0)$ 和 $o(s)$。

与表1中的方法类似，当 $\bar{s}_0>s_0$ 或 $\bar{s}_0<s_0$时，我们可以采用替代参数空间，即 $\mathcal{S}^{m,d}(s,\bar{s}_0)$ 或 $\mathcal{S}^{m,d}(ss_0/\bar{s}_0,\bar{s}_0)$，并获得相应的预言机

估计率。这些结果如表2所示。

表2：beta-min 条件下 s_0-错误设定模型的性质。

值	参数空间	阶数 $\tilde{\lambda}_2$	预言机估计 Rate	近乎完全恢复
$\bar{s}_0 < s_0$	$\mathcal{S}^{m,d}\left(\frac{ss_0}{\bar{s}_0}, \bar{s}_0\right)$	$\sqrt{\frac{\sigma^2}{n}\left(\text{对数}\frac{ed}{\bar{s}_0}+\frac{1}{\bar{s}_0}\text{对数}\frac{em\bar{s}_0}{ss_0}\right)}$	$\sqrt{\frac{\sigma^2}{n}ss_0}$	逐元素
$\bar{s}_0 = s_0$	$\mathcal{S}^{m,d}(s, s_0)$	$\sqrt{\frac{\sigma^2}{n}\left(\text{对数}\frac{ed}{s_0}+\frac{1}{s_0}\text{对数}\frac{em}{s}\right)}$	$\sqrt{\frac{\sigma^2}{n}ss_0}$	逐元素 and 逐组
$\bar{s}_0 > s_0$	$\mathcal{S}^{m,d}(s, \bar{s}_0)$	$\sqrt{\frac{\sigma^2}{n}\left(\log\frac{ed}{\bar{s}_0}+\frac{1}{\bar{s}_0}\log\frac{em}{s}\right)}$	$\sqrt{\frac{\sigma^2}{n}s\bar{s}_0}$	逐组

表2显示，当 $\bar{s}_0 \leq s_0$时，预言机估计速率为 $\sqrt{\frac{\sigma^2}{n}ss_0}$，表明其对 $\bar{s}_0$的变化不敏感。相反，对于 $\bar{s}_0 > s_0$，预言机估计速率增至 $\sqrt{\frac{\sigma^2}{n}s\bar{s}_0}$，进一步凸显了 s_0 作为 IHT 与组 IHT 之间权衡因素的作用，正如第 3.2 节所述。

注8 当 $\bar{s}_0 = 1$ 时，我们的结果与元素层面 *IHT* *[Ndaoud, 2020]* 的假设和结论一致。尽管 *IHT* 和 *DSIHT* 对于满足 *beta-min* 条件的（$\sqrt{\frac{\sigma^2}{n}ss_0}$） s, s_0）-稀疏向量均达到了预言机估计速率O，定理*8*表明，*DSIHT* 不仅如（20）所示在元素层面实现了近乎完全恢复，而且还如（19）所示在组层面实现了近乎完全恢复。这凸显了 *DSIHT* 相较于 *IHT* 的优越性。

5 数值实验

本节通过合成数据集和真实世界数据集开展数值实验，以考察我们提出的方法在实际中的表现。我们的算法在 R软件包 `ADSIHT`中实现。我们将其与几种最先进方法进行比较：稀疏组Lasso（SGLasso，Simon 等人 [2013]），由 R软件包 `sparseglt`[Liang 等人， 2024], 组桥（GBridge，Huang 等人 [2009]）、组指数Lasso（GEL，Breheny [2015]）和复合极小极大凹惩罚（CMCP，Breheny 和 Huang [2009]），这些方法由 R软件包 `grpreg` [Breheny, 2015]计算。对于 SGLasso，我们通过五折交叉验证确定调参参数。对于其他比较

方法，我们使用扩展贝叶斯信息准则（EBIC） [Chen and Chen, 2008]选择最优解。对于 ADSIHT，我们使用提出的 DSIC，并通过 $K = 5$ 选择最优模型。此外，其余超参数均采用 `sparsegl`和 `grpreg`中的默认值。所有数值实验均在 R 中进行，并在一台个人笔记本电脑上运行（AMD Ryzen 9 5900HX，3.30 GHz，16.00GB 内存）。

5.1 合成数据分析

合成数据集根据潜在模型 $y = X\beta^* + \xi$ 生成，其中 $\beta^* \in \mathbb{R}^p$ 有 m 个组，且组大小相同，即 $p_1 = \cdots = p_m = d$。设计矩阵 X 由多元高斯分布 $\mathcal{MVN}(0, \Sigma)$ 生成。协方差矩阵 Σ 设为自回归结构，即对于 $1 \le i, j \le p$, $\Sigma_{ij} = 0.5^{|i-j|}$。接下来，在以下两种情形下生成系数 β^*：

- 同质信号： β^* 从 $\{1, -1\}$ 中随机选取。
- 异质信号： β^* 从 $\mathcal{N}(0, 1)$ 中随机选取。

最后，随机误差 ξ_i 独立地由 $N(0, \sigma^2)$ 生成，并选择 σ 以达到所需的信噪比（SNR）。所有模拟结果均基于 100 次重复。给定一个输出 (S, β)，我们使用以下指标评估变量选择和参数估计的准确性：

- **稀疏性误差（SE）**： $|\hat{S}| - |S^*|$。
- **组级稀疏性误差（GSE）**： $\|\beta\|_{0,2} - \|\beta^*\|_{0,2}$。
- **马修斯相关系数（MCC）**：

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}},$$

  其中 TP$= S \cap S^*$和 TN$= S^c \cap (S^*)^c$分别表示真正例/真负例。 FP$= S \cap (S^*)^c$和 FN$= S^c \cap S^*$ 分别表示假正例/假负例。

- **估计误差（EE）**： $\|\beta - \beta^*\|_2^2$。

这里，SE 或 GSE 接近零意味着在支撑集上的估计结果更好。MCC 的取值范围为 $[-1, 1]$, 且 MCC 越大，变量选择性能越好。

5.1.1 不同信噪比下的统计性能

在本节中，我们研究模型信噪比变化对 ADSIHT 和其他最先进方法性能的影响。我们考虑生成模型包含 50

个非零系数，均匀分布在 10 个组中。我们设定样本量为 $n = 300$，组大小为 $d = 10$，组数为 $m = 100$。信噪比从 增至 20，增量为 2。图5分别展示了子图 **A** 和**B**中同质场景和异质场景的计算结果。

图5显示，随着信噪比的增加，所有方法的表现均趋于改善。在整个信噪比范围内，我们的方法在所有评价指标上均表现出色。对于同质信号设置，当信噪比较高时，我们的方法能够实现完整支持集恢复。另一方面，尽管即使在较高信噪比下，所有纳入考虑的方法均无法准确识别全部真实变量，但我们的方法在变量选择和参数估计方面仍表现出明显优势。

5.1.2 不同组数下的统计性能

这里我们研究统计指标如何随组数变化而变化。我们考虑生成模型包含50个非零系数，并将其均匀分配到10个组中。设定样本量为 $n = 500$、组大小为 $d = 10$ ，信噪比为 $= 5$。组数从50增加到500，增量为50。结果如图6所示。

从图6可以看出，我们的方法在高维设置下更加稳健。在变量选择和参数估计方面，我们的方法似乎优于其他考虑的方法，且在高维设置下差异最为显著。随着组数增加，其他方法的表现，尤其是 GBridge 的表现，会显著下降。

5.1.3 不同样本量下的统计性能

这里我们研究在其他参数保持不变的情况下，样本量变化对性能的影响。我们考虑包含50个非零系数的生成模型，这些系数均匀分布在5个组中。设组大小为 $d = 20$，组数为$m = 200$ ，信噪比为$= 5$。样本量从300增加到1000，增量为100。

如图7所示，随着样本量增加，所有方法的性能均显著提升。在不同统计指标上，我们的方法明显优于其他方法。对于同质信号设置，当样本量超过800时，我们的方法能够完美恢复支撑集。相比之下，即使样本量足够大，其他方法也无法实现完整支持集恢复。特别是对于两种信号设置，我们的方法都能准确估计系数，这与我们的方法在参数估计意义下具有极小极大最优性相一致。

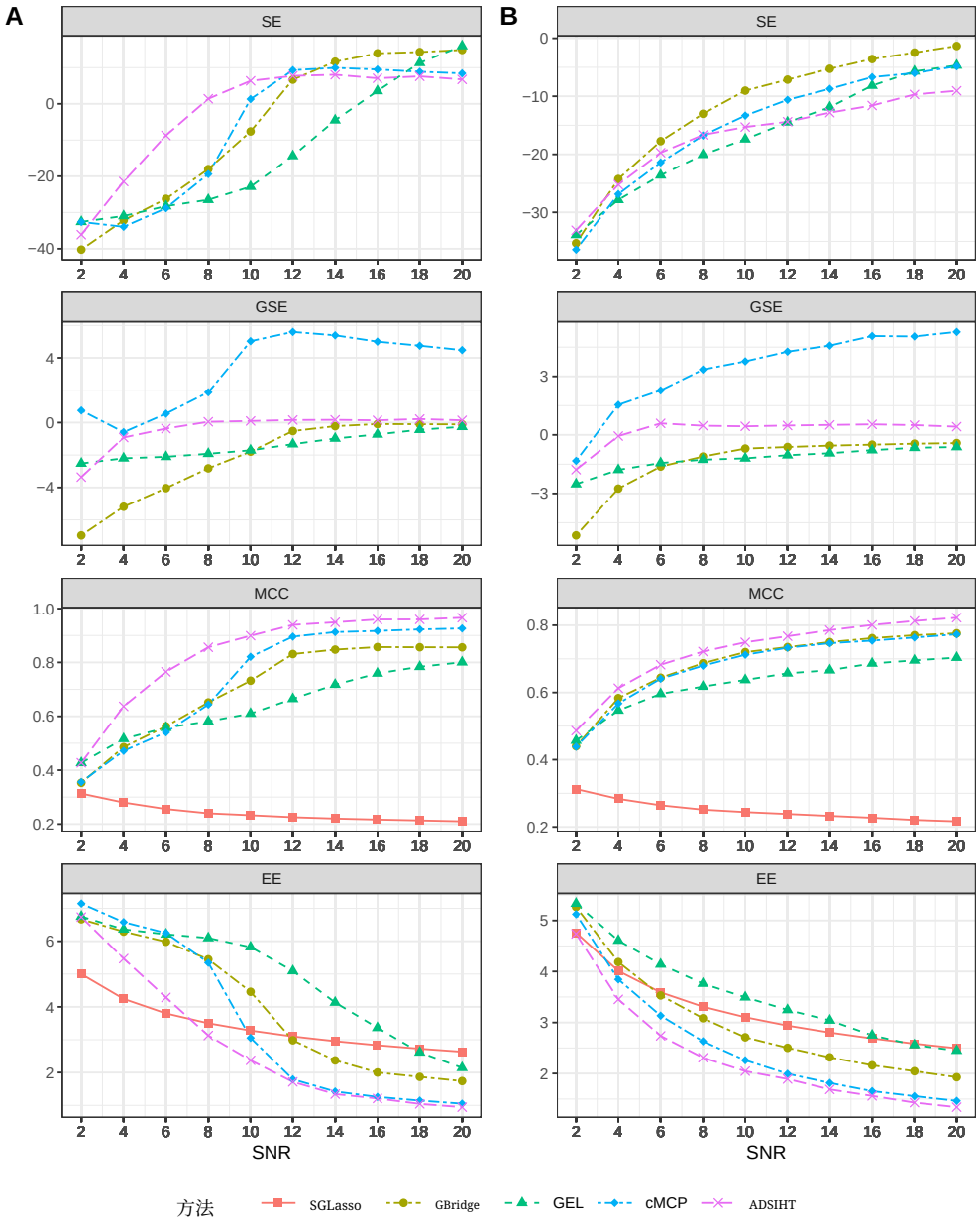

图5：随着信噪比（SNR）从1增加到10的性能指标。(A) 同质信号的计算结果。(B) 异质信号的计算结果。

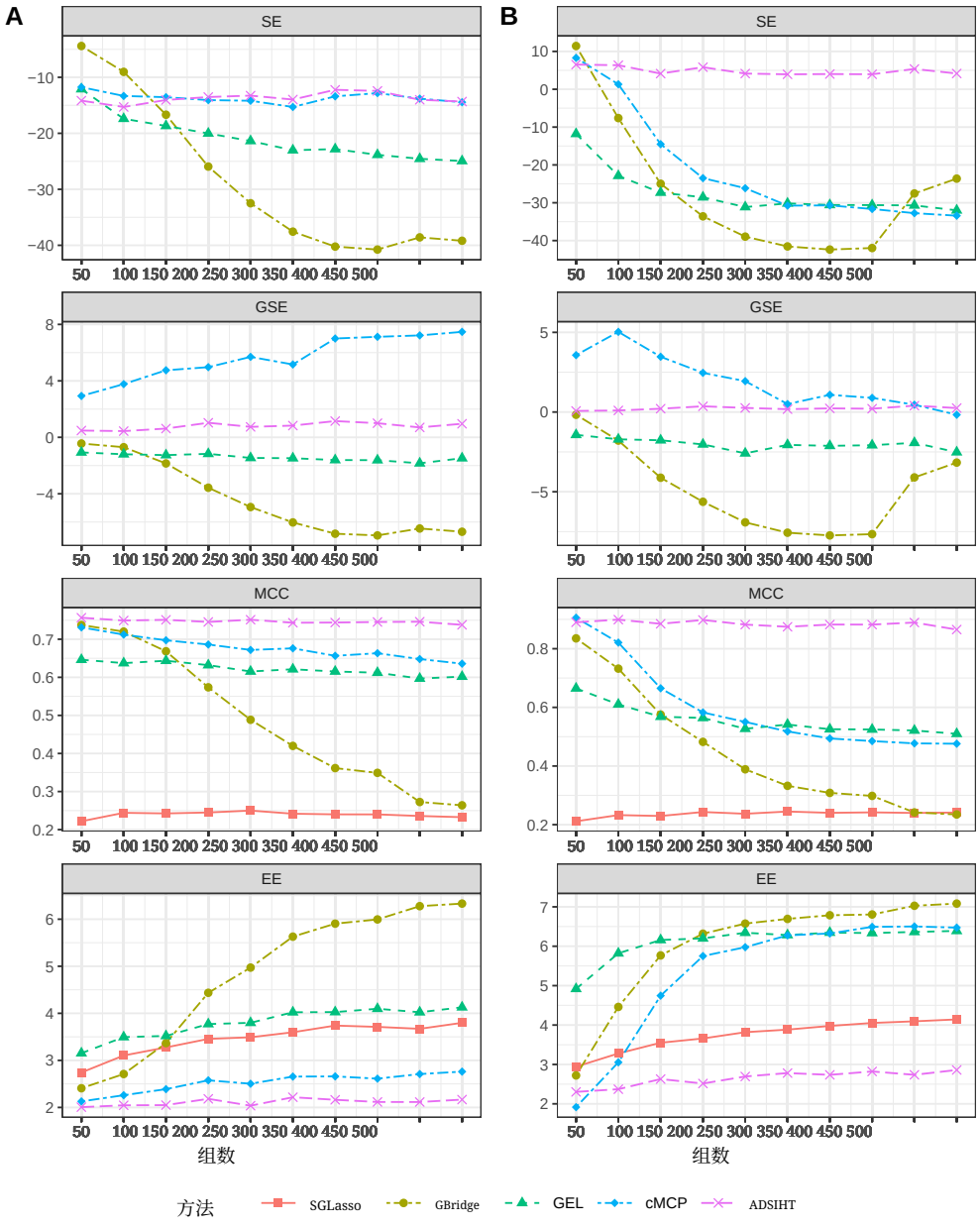

图6：组数从100增加到1200时的性能指标。（A）同质信号下的计算结果。（B）异质信号下的计算结果。

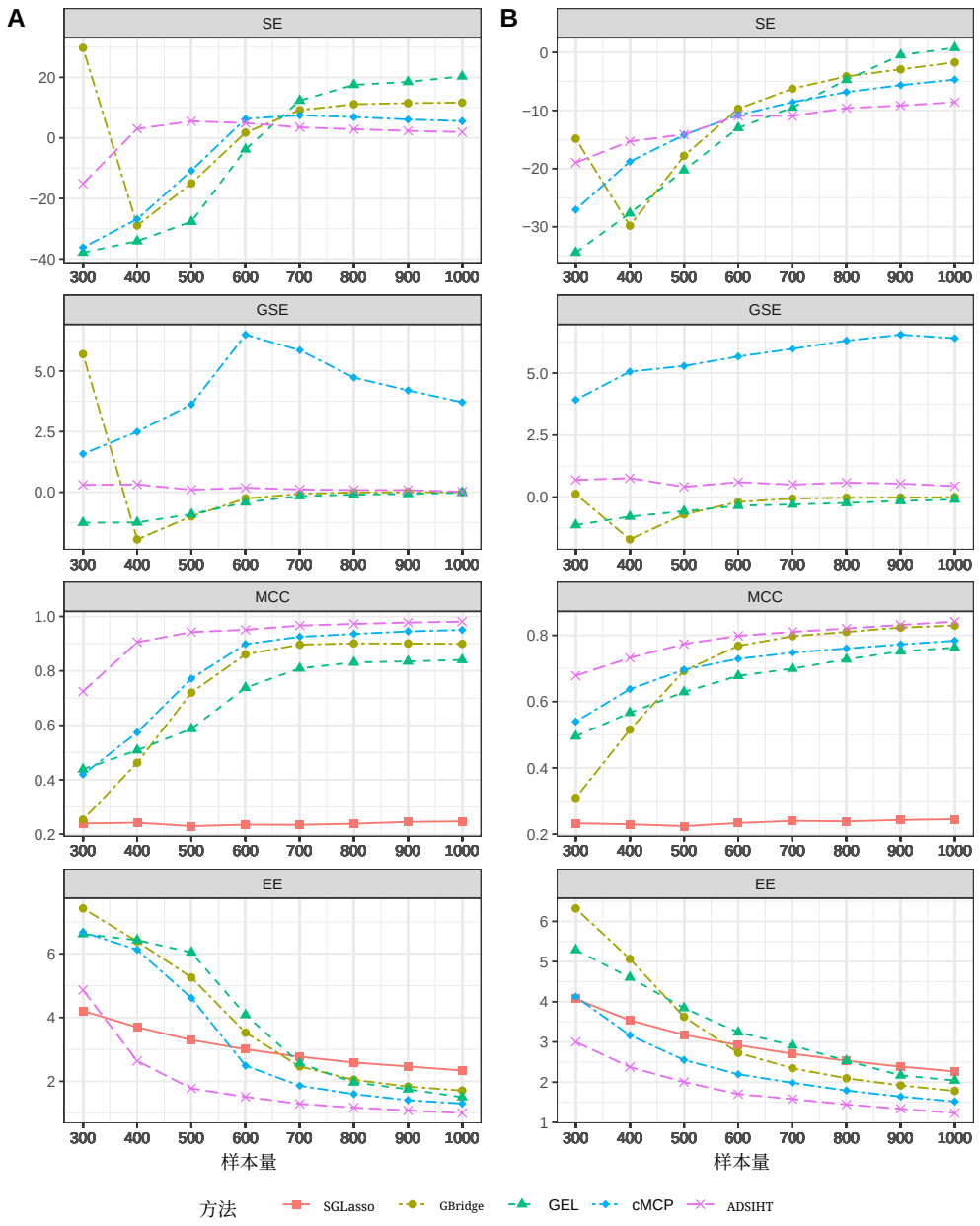

图7：样本量从300增加到1000时的性能指标。(A) 同质信号下的计算结果。(B) 异质信号下的计算结果。

5.2 真实世界数据分析

TRIM32 数据集涉及巴德-比德尔综合征相关基因的基因表达，最初由 Scheetz 等人 [2006] 提出，并已在多项统计学研究中得到广泛研究 [Huang 等人，2010年，Fan 等人，2011年，Zhang 等人， 2023]。本研究采集了 120 只 12 周龄雄性大鼠的眼部组织，并进行微阵列分析。在该数据集中，与导致巴德-比德尔综合征相关的基因 TRIM32 [Chiang 等人， 2006], 被作为响应变量，其余 18,975 个可能影响 TRIM32 表达的基因探针则作为协变量。

本文旨在找出与基因 TRIM32 在统计上显著相关的基因，并构建准确的预测模型。在 18,975 个探针中，我们选取边际球相关最高的前 300 个探针 [Pan 等人， 2019] 。随后，对于每个基因，我们使用包含十项的自然三次样条基展开，构建一个含 10 个变量的组。这项在科学研究中常用的技术[Huang 等人，2010年，Breheny 和 Huang，2015年，Zhang 等人， 2023], 使我们能够更有效地分析数据。完成上述操作后，该问题可描述为一个具有 $n = 120$、 $m = 300$和 $d = 10$的高维变量选择问题。在我们的分析中，将 120 只大鼠随机分为训练集（100 个样本）和测试集（其余 20 个样本）。我们重复这一随机划分过程 200 次，并计算所选变量和组的数量以及测试集预测均方误差（PMSE）的平均值。计算结果及 PMSE 箱线图分别见表 3 和图 8。

表 3：TRIM32 数据集的计算结果。标准差显示在括号内。

方法	变量数	组数	100×预测均方误差
SGLasso	139.26 (68.74)	26.95 (12.45)	1.71 (1.84)
GBridge	2.95 (0.81)	1.04 (0.18)	2.01 (2.00)
GEL	35.78 (28.03)	7.07 (3.69)	2.55 (2.70)
cMCP	21.60 (3.37)	20.95 (3.08)	1.92 (2.12)
ADSIHT	29.06 (11.93)	9.20 (4.09)	**1.70 (1.85)**

表 3 显示，SGLasso 识别出的变量和组显著多于其他方法。然而，这并未带来测试集上最佳的预测表现。另一方面，尽管使用的变量和组更少，我们提出的方法在预测结果时仍展现出最高的统计精度。此外，图 8 表明，我们的方法在预测表现上既准确又稳健，证明了其优于其他方法。

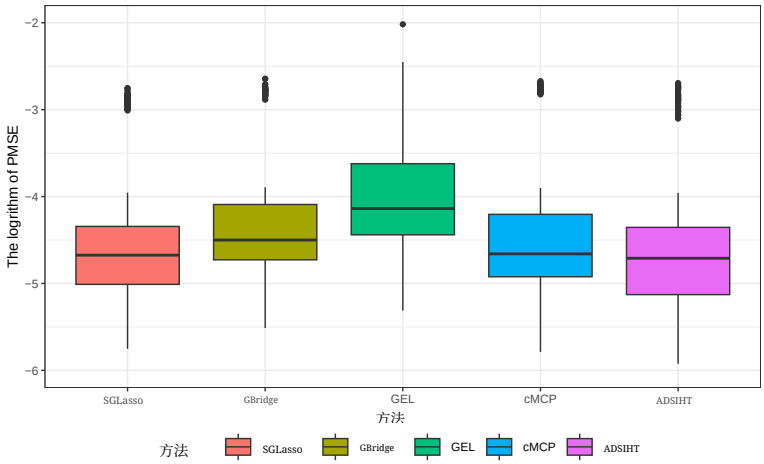

图 8：预测均方误差的箱线图。

为进一步开展研究，我们使用全部 120 个样本来拟合 TRIM32 表达的双稀疏线性模型。图 9 展示了由我们提出的方法及对比方法估计得到的残差 QQ 图。cMCP 和 ADSIHT 的子图中，点大致落在对角线上，表明正态性假设成立。相比之下，图 9 显示，SGLasso、GBridge 和 GEL 的残差分布左侧尾部较长，这意味着使用拟合的线性模型分析该数据集可能缺乏说服力。此外，我们按照表 4 所述，计算了每种方法的 R^2 和调整后的 R^2。表 4 中的计算结果表明，ADSIHT 的拟合表现较好。具体而言，ADSIHT 有效识别出 14 个重要组以及这些组中的 31 个显著变量，共同解释了 TRIM32 表达变异的 79%。尽管 SGLasso 对 TRIM32 表达变异的解释度最高，但它选择的模型规模过大，可能存在过拟合问题。

表 4：各方法的 R^2 和调整后的 R^2。由于 SGLasso 选择的模型规模过大，因此省略调整后的 R^2。

	SGLasso	GBridge	GEL	cMCP	ADSIHT
R^2	88%	44%	54%	75%	79%
调整后的 R^2	×	42%	52%	68%	71%

6 结论

在我们的工作中，我们为高维双重稀疏线性回归提出了一种极小极大最优的 IHT 风格程序。具体而言，我们引入了一种新颖的双重稀疏迭代

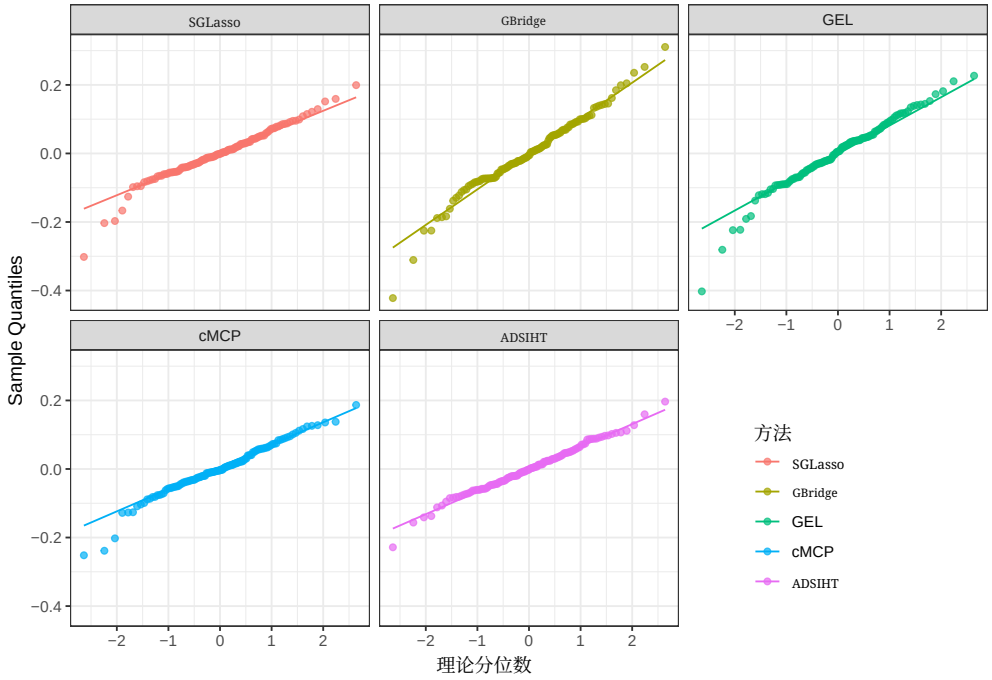

图 9：残差的 QQ 图。

硬阈值（DSIHT）算子。为有效控制假发现，我们迭代降低 DSIHT 算子中的阈值，直至达到最优阈值。在一定条件下，我们证明了 DSIHT 算法能够获得一个极小极大最优估计量。

值得注意的是，对于 (s, s_0)-稀疏结构，我们设计了一种完全自适应最优程序，使我们的算法能够在稀疏度 s, s_0 和方差 σ^2 未知的情况下，得到一个极小极大最优估计量。首先，给定 s_0，我们引入一种自适应程序，利用 Birg´e-Massart 准则的一种变体来确定最优停止时间，该程序独立于 s 和 σ。重要的是，我们强调了稀疏度 s_0 在 IHT 与组 IHT 之间权衡中的作用。在此结果的基础上，我们提出了一种新颖的双重稀疏信息准则来选择最优的 s_0，从而使我们的方法成为完全自适应程序。理论上，我们证明了两步自适应程序能够以快速收敛达到最优统计精度。更重要的是，为了说明我们的算法为何优于稀疏组Lasso，我们证明，在满足 beta-min 条件时，我们的算法能够达到预言机估计速率，而凸估计量无法达到这一速率，并且能够近乎完全恢复真实支撑集。最后，数值实验表明，与其他最先进方法相比，我们的方法具有更准确、更稳健的统计性能。

本文研究了线性回归中的双重稀疏结构，并且可以将类似的方法推广到广义线性模型或单指标模型中。此外，我们的技术结果还可应用于其他具有同时稀疏结构的各种问题，例如稀疏加性模型 [Raskutti 等人，2012年，Yuan 和 Zhou， 2016]以及

高维变点问题 [Liu 等人, 2021]。我们将这些方向视为未来潜在的研究方向。

附录

附录包含所有定理和推论的技术性证明。主要结果的证明见附录 A。附录 B 包含辅助引理的证明。附录 C 给出了次高斯随机设计下 DSRIP 条件的一个示例。为简化附录中的记号，记

$$\Delta := \log \frac{ed}{s_0} + \frac{1}{s_0} \log \frac{em}{s} \quad \text{and} \quad \Delta' := \log \frac{ed}{s_0} + \frac{1}{s_0} \log em.$$

ˆ 给定一个 p 维向量 β，其中 $\|\beta\|_0 = A$ 且 $\|\beta\|_{0,\ 2} = \hat{s}$，记

$$\Omega^*(\beta) := (s + \hat{s}) \log \frac{em}{s + \hat{s}} + (ss_0 + \hat{A}) \log \frac{ed(s + \hat{s})}{ss_0 + \hat{A}}.$$

附录 A：主要结果的证明

引理 1 的证明

根据 Hsu 等人的定理 2.1 [2012], $\forall S \in \mathcal{S}$，我们有

$$P\left(\frac{\|X_S^\top \xi\|_2^2}{\sigma^2} \geq Tr(X_S^\top X_S) + 2\|X_S^\top X_S\|_F \sqrt{t} + 2\lambda_{\max}(X_S^\top X_S)t \right) \leq e^{-t}, \tag{21}$$

其中常数为 $t \geq 0$。由于 $X_S \in \mathbb{R}^{n \times ss_0}$ 和 $\|X_j\|_2 = \sqrt{n}, j \in [p]$，因此我们有

$$Tr(X_S^\top X_S) = \left(\sum_{j=1}^{ss_0} \sum_{i=1}^{n} X_{ij}^2 \right) \leq ss_0 n. \tag{22}$$

一方面，我们有

$$\lambda_{\max}(X_S^\top X_S) \leq n(1 + \delta). \tag{23}$$

另一方面，由 (22) 和 (23) 可得

$$\|X_S^\top X_S\|_F = \sqrt{Tr(X_S^\top X_S X_S^\top X_S)} \leq (1 + \delta)\sqrt{ss_0} n, \tag{24}$$

将（22）—（24）代入（21），得到

$$P\left(\frac{1}{n\sigma^2} \|X_S^\top \xi\|_2^2 \geq 2(1 + \delta) \left[\sqrt{t} + \frac{\sqrt{ss_0}}{2} \right]^2 + \frac{1 - \delta}{2} ss_0 \right) \leq e^{-t}.$$

注意到 $\delta < 1$ 和 $\Delta \gg 1$。对于某个正常数 C，令 $t = (1+C)ss_0\Delta$，则有 $2(1+\delta)\left[\sqrt{t + \frac{\sqrt{ss_0}}{2}}\right]^2 + \frac{1-\delta}{2} ss_0 \leq 4ss_0\Delta$。因此，我们得到

$$P\left(\frac{1}{n}\|X_S^\top \xi\|_2^2 \geq 4\sigma^2 ss_0\Delta\right) \leq e^{-(1+C)ss_0\Delta}. \tag{25}$$

注意到

$$|\mathcal{S}^{m,d}(s,s_0)| \leq \binom{m}{s} \times \binom{sd}{ss_0} \leq (\frac{em}{s})^s \times (\frac{ed}{s_0})^{ss_0} \leq e^{ss_0\Delta}. \tag{26}$$

因此，结合 (25) 和 (26)，我们得到

$$\begin{aligned}
&P\left(\forall S \in \mathcal{S}^{m,d}(s,s_0), \sum_{i\in S} \Xi_i^2 \leq \frac{4\sigma^2 ss_0\Delta}{n}\right) \\
=&1 - P\left(\exists S \in \mathcal{S}^{m,d}(s,s_0), \sum_{i\in S} \Xi_i^2 > \frac{4\sigma^2 ss_0\Delta}{n}\right) \\
\geq&1 - |\mathcal{S}^{m,d}(s,s_0)| P\left(\sum_{i\in S} \Xi_i^2 > \frac{4\sigma^2 ss_0\Delta}{n}\right) \\
\geq&1 - e^{-Css_0\Delta},
\end{aligned}$$

其中，第一个不等式由并集界得出。引理 1 的证明至此完成。

定理 1 的证明

我们在假设事件 $\mathcal{E}$ 成立的条件下证明定理 1。首先，很容易确认，对于 $t = 0$，结论显然成立。接着，假设结论在第 t 步成立，并证明其在第 $t+1$ 步也成立。

我们首先用反证法证明 (5) 和 (6)。假设对于$t+1$，(5) 和 (6) 不成立，即 $S_{G^*} \cap S^{t+1} \cap (S^*)^c \notin \mathcal{S}^{m,\ d}(s,\ s_0)$ 且 $S_{G^*}^c \cap S^{t+1} \notin \mathcal{S}^{m,\ d}(s,\ s_0)$。

步骤 1

对于结果 (5)，注意 $S_{G^*} \cap (S^*)^c$ 至多覆盖 s 个组。根据第 2 节中的~**情形1**，存在一个 (s, s_0)-形子集 $S_{1,t+1}$，它属于 $S_{G^*} \cap (S^*)^c$，且基数为 ss_0，使得

$$ss_0\lambda_{t+1}^2 \leq \sum_{i\in\tilde{S}_{1,t+1}} \left\{\mathcal{T}_{\lambda_{t+1},s_0}(H^{t+1})\right\}_i^2.$$

˜注意，对于 $i \in S_{1,t+1} \subseteq (S^*)^c$，有 $\beta_i^* = 0$。然后，利用方程 (4) 和三角不等式，我们得到

$$\sqrt{ss_0}\lambda_{t+1} \leq \sqrt{\sum_{i\in\tilde{S}_{1,t+1}} \langle \Phi_i^\top, \beta^t - \beta^* \rangle^2} + \sqrt{\sum_{i\in\tilde{S}_{1,t+1}} \Xi_i^2}.$$

回顾一下，β^* 是 $(s,\ s_0)$-稀疏的，并且根据假设，(5) 和 (6) 对 t 均成立。因此，我们有 $\beta^t - \beta^*$ 是 $(2s,\ \frac{3}{2}s_0)$-稀疏的。于是，利用 DSRIP 条件和引理 1，可得

$$\begin{aligned}
\sqrt{ss_0}\lambda_{t+1} \leq & \delta\|\beta^* - \beta^t\|_2 + 2\sigma\sqrt{\frac{ss_0\Delta}{n}} \\
\leq & \frac{3}{2}(1+\sqrt{2})\delta\sqrt{ss_0}\lambda_t + 2\sigma\sqrt{\frac{ss_0\Delta}{n}} \\
\leq & \frac{3}{2}(1+\sqrt{2})\delta^{\frac{9}{10}}\sqrt{ss_0}\lambda_{t+1} + \frac{1}{2}\sqrt{ss_0}\lambda_\infty \\
\leq & \left(\frac{1}{2} + \frac{3}{2}(1+\sqrt{2})\delta^{\frac{9}{10}}\right)\sqrt{ss_0}\lambda_{t+1} \\
< & \sqrt{ss_0}\lambda_{t+1},
\end{aligned}$$

这导致了矛盾。由于我们已假设 (7) 对 t 成立，第二个不等式由此成立，而最后一个不等式则由 $\delta < 0.11 \wedge \kappa^{10}$ 得出。因此，我们有 $S_{G^*} \cap S^{t+1} \cap (S^*)^c \in \mathcal{S}^{m,\ d}(s,\ s_0)$，这表明 (5) 对 $t+1$ 成立。

步骤 2

对于结果 (6)，如果 $S_{G^*}^c \cap S^{t+1}$ 覆盖的组不超过 s 个，则结果 (6) 的分析与**步骤 1**相同。否则，根据第 2 节中的**情形 2**，存在 $S_{G^*}^c$ 的一个 $(s,\ s_0)$-形状的 ˜ 子集 $S_{2,\ t+1}$，使得

$$ss_0\lambda_{t+1}^2 \leq \sum_{i\in\tilde{S}_{2,t+1}} \left\{\mathcal{T}_{\lambda_{t+1},s_0}(H^{t+1})\right\}_i^2.$$

(6) 的其余证明与**步骤 1**类似。因此，对于 $t+1$，(6) 成立。

步骤 3

现在我们转向证明（7）。注意，结果（5）和（6）对 $t+1$ 成立，这意味着 $\beta^{t+1} - \beta^*$ 是 $(2s,\ \frac{3}{2}s_0)$-稀疏的。注意，对任意 $i \in [p]$,

$$\beta_i^{t+1} - \beta_i^* = -H_i^{t+1}\mathrm{I}(i \notin S^{t+1}) + \langle \Phi_i^\top, \beta^* - \beta^t \rangle + \Xi_i. \tag{27}$$

一方面，将（27）两边对集合 $S^{t+1}\cap(S^*)^c$ 求和，可得

$$\begin{aligned}\|\beta^{t+1}_{(S^*)^c}\|_2 \leq&\sqrt{\sum_{i\in S^{t+1}\cap(S^*)^c}\langle\Phi_i^\top,\beta^*-\beta^t\rangle^2}+\sqrt{\sum_{i\in S^{t+1}\cap(S^*)^c}\Xi_i^2}\\ \leq&\delta\|\beta^*-\beta^t\|_2+2\sigma\sqrt{\frac{2ss_0\Delta}{n}},\end{aligned}\tag{28}$$

其中，第二个不等式右侧来自对应于 (5) 和 (6) 的两部分随机误差的累积。另一方面，将 (27) 两边在支撑集 S^* 上求和，可得

$$\begin{aligned}\|(\beta^{t+1}-\beta^*)_{S^*}\|_2 \leq&\sqrt{\sum_{i\in S^*}(H_i^{t+1})^2\mathrm{I}(i\notin S^{t+1})}+\sqrt{\sum_{i\in S^*}\langle\Phi_i^\top,\beta^*-\beta^t\rangle^2}+\sqrt{\sum_{i\in S^*}\Xi_i^2}\\ \leq&\sqrt{2ss_0}\lambda_{t+1}+\delta\|\beta^*-\beta^t\|_2+2\sigma\sqrt{\frac{ss_0\Delta}{n}}.\end{aligned}\tag{29}$$

由于算子 $\mathcal{T}_{\lambda,s_0}(\cdot)$ $\sum_{i\in S^*}(H_i^{t+1}$) 的程序分为两个步骤，(29) 中的项 $\sum_{i\in S^*}(H_i^{t+1}$)^2 上界为 $2ss_0\lambda_{t+1}^2$。结合 (28) 和 (29)，我们得到

$$\begin{aligned}\|\beta^{t+1}-\beta^*\|_2 &\leq \|\beta^{t+1}_{(S^*)^c}\|_2+\|(\beta^{t+1}-\beta^*)_{S^*}\|_2\\ &\leq\sqrt{2ss_0}\lambda_{t+1}+2\delta\|\beta^*-\beta^t\|_2+2(1+\sqrt{2})\sigma\sqrt{\frac{ss_0\Delta}{n}}\\ &\leq\left(\sqrt{2}+3(1+\sqrt{2})\delta^{\frac{9}{10}}+\frac{1+\sqrt{2}}{2}\right)\sqrt{ss_0}\lambda_{t+1}\\ &\leq\frac{3}{2}(1+\sqrt{2})\sqrt{ss_0}\lambda_{t+1},\end{aligned}$$

其中，第三个和最后一个不等式由 $\delta<0.11\wedge\kappa^{10}$ 得出。我们证明，对于 $t+1$，(7) 成立。

最后，我们已经证明，在归纳假设下，定理 1 中的结果对 $t+1$ 成立。定理 1 得证。

定理 2 的证明

考虑 $\frac{ss_0}{4}$-打包集 $\{\beta^1,\ldots,\beta^M\}$，其中 M是打包数 $\widetilde{M}$（$\frac{ss_0}{4}$；$\Theta^{m,d}$（s，s_0）$\|\cdot\|_H$）的简写。我们将 $\beta\in\{\beta^1,\ldots,\beta^M\}$的所有非零元素设为 δ，其中 δ 是一个需要在下文确定的参数。对于任意 $\beta^i\neq\beta^j$，由于每个 β^i 至多有 ss_0 个非零元素，因此我们有

$$\|\beta^i-\beta^j\|_2^2\leq 2ss_0\delta^2,\ \forall i,j\in[M].\tag{30}$$

~另一方面，由于 $\{\beta^1, \ldots, \beta^M\}$ 是 $\Theta^{m,\ d}(s,\ s_0)$ 的 $\frac{ss_0}{4}$-打包集，因此我们有

$$\|\beta^i - \beta^j\|_2^2 \geq \frac{1}{4} s s_0 \delta^2,\ \forall i, j \in [M]. \tag{31}$$

给定设计矩阵 X，记 $y^i = X\beta^i + \xi,\ \forall i \in [M]$。我们考虑不同分布对之间的 Kullback-Leibler 散度：

$$\begin{aligned} KL\left(y^i || y^j\right) &= \frac{1}{2\sigma^2} \|X(\beta^i - \beta^j)\|_2^2 \\ &\leq \frac{n\vartheta_{\max}^2}{2\sigma^2} \|\beta^i - \beta^j\|_2^2, \end{aligned}$$

其中，最后一个不等式由 X 和 $\beta^i - \beta^j \in \widetilde{}\Theta^{m,\ d}(2s,\ 2s_0)$ 的特征值条件得出。记 B 为在打包集上均匀分布的随机向量。注意到：

$$\begin{aligned} I(y; B) \leq & \frac{1}{\binom{M}{2}} \sum_{i \neq j} KL(y^i || y^j) \\ \leq & \frac{1}{\binom{M}{2}} \sum_{i \neq j} \frac{n\vartheta_{\max}^2}{2\sigma^2} \|\beta^i - \beta^j\|_2^2 \\ \leq & \frac{n\vartheta_{\max}^2}{\sigma^2} s s_0 \delta^2, \end{aligned} \tag{32}$$

其中，最后一个不等式使用了（30）。结合广义 Fano 引理 [Cover 和 Thomas, 2006] 以及（32），我们有

$$\begin{aligned} P(B \neq \widetilde{\beta}) &\geq 1 - \frac{I(y; B) + \log 2}{\log M} \\ &\geq 1 - \frac{\frac{n\vartheta_{\max}^2}{\sigma^2} s s_0 \delta^2 + \log 2}{\log M}, \end{aligned}$$

~ ~其中 β取值于打包集。为保证 $P($ $B \neq \beta$) $\geq \frac{1}{2}$，只需选择$\delta = \frac{1}{2}\sqrt{\frac{\sigma^2 \log_{n\vartheta_{\max}^2} M}{ss_0}}$。将其代入方程（31），并根据引理 2，我们有

$$\begin{aligned}
&\inf_{\hat{\beta}} \sup_{\beta^* \in \Theta^{m,d}(s,s_0)} \mathbf{E}_{\hat{\beta}} \|\hat{\beta} - \beta^*\|_2^2 \\
\geq & \frac{1}{16} ss_0 \delta^2 \cdot \inf_B P\left(B \neq \tilde{\beta}\right) \\
\geq & \frac{\sigma^2 \log M}{128 n \vartheta_{\max}^2} \\
\geq & \frac{\sigma^2}{512 n \vartheta_{\max}^2} \left(ss_0 \log \frac{ed}{s_0} + s \log \frac{em}{s}\right),
\end{aligned}$$

由此完成定理 2 的证明。

定理 3 的证明

根据引理 3，以至少 $1 - \exp\{-Css_0\Delta\}$ 的概率，我们有

$$\sigma_0 \geq \frac{19}{20}\sigma - \sqrt{1+\delta}\|\beta^*\|_2. \tag{33}$$

以至少 $1 - \exp\{-Css_0\Delta\}$ 的概率，我们有

$$\begin{aligned}
\|M_{S^*}\|_2 &= \| \left(\beta^* + \Phi\beta^* + \Xi\right)_{S^*} \|_2 \\
&\geq \|\beta^*\|_2 - \|\Phi\beta^*\|_2 - \|\Xi_{S^*}\|_2 \\
&\geq (1-\delta)\|\beta^*\|_2 - 2\sigma\sqrt{\frac{ss_0\Delta}{n}}.
\end{aligned} \tag{34}$$

注意到 $\sqrt{ss_0}\|M\|_\infty \geq \|M_{S^*}\|_2$。结合 (33) 和 (34)，以至少$1 - \exp\{-Css_0\Delta\}$的概率，我们有

$$\begin{aligned}
\sqrt{ss_0}\lambda_0 &\geq \frac{100}{9}\sigma_0\sqrt{\frac{ss_0\Delta'}{n}} \vee \frac{19}{4}\sqrt{ss_0}\|M\|_\infty \\
&\geq \frac{9}{19} \times \frac{100}{9}\sigma_0\sqrt{\frac{ss_0\Delta'}{n}} + \frac{10}{19} \times \frac{19}{4}\|M_{S^*}\|_2 \\
&\geq \frac{100}{19}\left(\frac{19}{20}\sigma - \sqrt{1+\delta}\|\beta^*\|_2\right)\sqrt{\frac{ss_0\Delta'}{n}} + \frac{5}{2}\left((1-\delta)\|\beta^*\|_2 - 2\sigma\sqrt{\frac{ss_0\Delta}{n}}\right) \\
&\geq \left(\frac{5}{2}(1-\delta) - \frac{100}{19}\sqrt{1+\delta}\sqrt{\frac{ss_0\Delta'}{n}}\right)\|\beta^*\|_2 \\
&\geq \|\beta^*\|_2
\end{aligned}$$

其中，第四个不等式利用了 $\Delta' \geq \Delta$，最后一个不等式利用了 $n > 105^2 s s_0 \Delta'$ 和 $\delta < 0.11$。定理 3 的证明至此完成。

定理 4 的证明

注意，由于 $\Delta' \geq \Delta$， $t_0 \leq t_\infty$ 成立。我们首先断言 $t_0 \geq \bar{t}$。对于任意 $t \leq t_0$，根据 t_0 的定义，我们有

$$\sigma \sqrt{\frac{\Delta'}{n}} \leq \frac{1}{12} \lambda_t. \tag{35}$$

由引理 ma 3 和定理 1，以至少 $1 - \exp\{-Css_0\Delta\}$ 的概率，我们有 e

$$\begin{aligned} \sigma_t &\leq \sqrt{1+\delta} \|\beta^* - \beta^t\|_2 + \frac{21}{20}\sigma \\ &\leq \sqrt{1+\delta} \frac{3}{2}(1+\sqrt{2}) \sqrt{ss_0} \lambda_t + \frac{21}{20}\sigma. \end{aligned} \tag{36}$$

由（36）可得三分彩

$$\begin{aligned} 8\sigma_t \sqrt{\frac{\Delta'}{n}} &\leq 12(1+\sqrt{2})\sqrt{1+\delta}\lambda_t \sqrt{\frac{ss_0\Delta'}{n}} + \frac{42}{5}\sigma\sqrt{\frac{\Delta'}{n}} \\ &\leq 12(1+\sqrt{2})\sqrt{1+\delta}\lambda_t \sqrt{\frac{ss_0\Delta'}{n}} + \frac{7}{10}\lambda_t \\ &\leq \lambda_t, \end{aligned} \tag{37}$$

其中第一个不等式使用了 (35)，第二个不等式则由 $\delta < 0.11$ and$n > 105^2 s s_0 \Delta'$ 得出。(37) 表明 $t \leq \bar{t}$，由此可推得 $t_0 \leq \bar{t}$ 以高概率成立。

¯接下来，我们转向 $t \leq t_\infty$的证明。由于 $t_0 \leq t_\infty$，定理 1 表明

$$\|\beta^{t_0} - \beta^*\|_2 \leq 18(1+\sqrt{2})\sigma \sqrt{\frac{ss_0\Delta'}{n}}. \tag{38}$$

由引理 3，对于任意 $t_0 \leq t \leq t_\infty$，以至少 $1 - \exp\{-Css_0\Delta\}$ 的概率有

$$\begin{aligned} |\sigma_t - \sigma| &\leq \sqrt{1+\delta}\|\beta^* - \beta^t\|_2 + \frac{1}{20}\sigma \\ &\leq 18(1+\sqrt{2})\sqrt{1+\delta}\sigma\sqrt{\frac{ss_0\Delta'}{n}} + \frac{1}{20}\sigma \\ &\leq (\frac{9}{20} + \frac{1}{20})\sigma = \frac{1}{2}\sigma, \end{aligned}$$

其中，第二个不等式由（38）得出，最后一个不等式由 $\delta < 0.11$ 和 $n > 105^2 s s_0 \Delta'$ 得出。结合上述不等式，可得

$$\frac{8\sigma_t}{\sqrt{n}}\sqrt{\Delta'} \geq \frac{4\sigma}{\sqrt{n}}\sqrt{\Delta'} \geq \frac{4\sigma}{\sqrt{n}}\sqrt{\Delta}.$$

该结果意味着 $t \leq t_\infty$，从而完成定理 4 的证明。

定理5的证明

根据引理 3 和式 (38)，以至少 $1 - \exp\{-Css_0\Delta\}$ 的概率，我们有

$$\begin{aligned} |\sigma_{\bar{t}} - \sigma| &\leq \sqrt{1+\delta}\|\beta^{\bar{t}} - \beta^*\|_2 + \frac{1}{20}\sigma \\ &\leq \sigma\left(18(1+\sqrt{2})\sqrt{1+\delta}\sqrt{\frac{ss_0\Delta'}{n}} + \frac{1}{20}\right) \\ &\leq \frac{1}{10}\sigma, \end{aligned} \tag{39}$$

其中，最后一个不等式用到了 $\delta < 0.11$ 和 $n > 1000^2 s s_0 \Delta'$。

首先，我们用反证法证明 $\|\beta^t\|_G \leq 4s$。假设 $\|\beta^t\|_G > 4s$。根据 t的定义，我们有

$$\frac{1}{n}\left\|y - X\beta^{\tilde{t}}\right\|_2^2 + \frac{1000\sigma_{\tilde{t}}^2\Omega(\beta^{\tilde{t}})}{n} \leq \frac{1}{n}\left\|y - X\beta^{t_\infty}\right\|_2^2 + \frac{1000\sigma_{\tilde{t}}^2\Omega(\beta^{t_\infty})}{n}. \tag{40}$$

另一方面，我们有

$$\begin{aligned} \left\|y - X\beta^{\tilde{t}}\right\|_2^2 &\geq \|\xi\|_2^2 + \left\|X(\beta^{\tilde{t}} - \beta^*)\right\|_2^2 - 2\left|\left\langle \xi, X(\beta^{\tilde{t}} - \beta^*)\right\rangle\right| \\ &\geq \|\xi\|_2^2 + \left\|X(\beta^{\tilde{t}} - \beta^*)\right\|_2^2 - 2\sigma\sqrt{3\Omega^*(\beta^{\tilde{t}})}\left\|X(\beta^{\tilde{t}} - \beta^*)\right\|_2 \\ &\geq \|\xi\|_2^2 + \left\|X(\beta^{\tilde{t}} - \beta^*)\right\|_2^2 - 2\sigma\sqrt{3 \times \frac{5}{4}\Omega(\beta^{\tilde{t}})}\left\|X(\beta^{\tilde{t}} - \beta^*)\right\|_2 \\ &\geq \|\xi\|^2 - \frac{15}{4}\sigma^2\Omega(\beta^{\tilde{t}}), \end{aligned}$$

其中，第二个不等式由引理 4 得出，第三个不等式利用了这一事实：

$\frac{5}{4}\Omega(\beta^t) \geq \Omega^*(\beta^t)$，当 $\|\beta^t\|_G > 4s$ 时成立。$\Omega^*(\beta)$ 的定义见开头

的附录。通过一些简单代数运算，可得

$$\begin{aligned}\frac{1}{n}\left\|y-X\beta^{\tilde{t}}\right\|_2^2+\frac{1000\sigma_{\tilde{t}}^2\Omega(\beta^{\tilde{t}})}{n}\geq&\frac{\|\xi\|_2^2}{n}-\frac{15\sigma^2\Omega(\beta^{\tilde{t}})}{4n}+\frac{1000\sigma_{\tilde{t}}^2\Omega(\beta^{\tilde{t}})}{n}\\ \geq&\frac{\|\xi\|_2^2}{n}+\frac{950\sigma_{\tilde{t}}^2\Omega(\beta^{\tilde{t}})}{n}.\end{aligned}\tag{41}$$

另一方面，我们有

$$\begin{aligned}\left\|y-X\beta^{t_\infty}\right\|_2^2&\leq\|\xi\|_2^2+\left\|X(\beta^{t_\infty}-\beta^*)\right\|_2^2+2\left|\left\langle\xi,X(\beta^{t_\infty}-\beta^*)\right\rangle\right|\\ &\leq\|\xi\|_2^2+\left\|X(\beta^{t_\infty}-\beta^*)\right\|_2^2+6\sigma\sqrt{\Omega(\beta^*)}\left\|X(\beta^{t_\infty}-\beta^*)\right\|_2\\ &\leq\|\xi\|_2^2+2\left\|X(\beta^{t_\infty}-\beta^*)\right\|_2^2+9\sigma^2\Omega(\beta^*),\end{aligned}$$

其中第二个不等式由引理 4 中的 (87) 得出，因为 $\beta^{t_\infty}-\beta^*$ 是 $(2s,\frac{3}{2}s_0)$-稀疏的。通过一些简单代数运算，可得

$$\begin{aligned}\frac{1}{n}\left\|y-X\beta^{t_\infty}\right\|_2^2+\frac{1000\sigma_{\tilde{t}}^2\Omega(\beta^{t_\infty})}{n}\leq&\frac{\|\xi\|_2^2}{n}+\frac{2}{n}\left\|X(\beta^{t_\infty}-\beta^*)\right\|_2^2+\frac{9\sigma^2\Omega(\beta^*)}{n}\\ &+\frac{1000\sigma_{\tilde{t}}^2\Omega(\beta^{t_\infty})}{n}\\ \leq&\frac{\|\xi\|_2^2}{n}+2(1+\delta)\left(\frac{3}{2}(1+\sqrt{2})\right)^2\times\frac{16\sigma^2ss_0\Delta}{n}\\ &+\frac{9\sigma^2ss_0\Delta}{n}+\frac{1000\sigma_{\tilde{t}}^2}{n}\left(2s\log\frac{em}{2s}+3ss_0\log\frac{ed}{s_0}\right)\\ \leq&\frac{\|\xi\|_2^2}{n}+\frac{480\sigma^2ss_0\Delta}{n}+\frac{3000\sigma_{\tilde{t}}^2ss_0\Delta}{n}\\ \leq&\frac{\|\xi\|_2^2}{n}+\frac{3600\sigma_{\tilde{t}}^2ss_0\Delta}{n},\end{aligned}\tag{42}$$

其中，第三个不等式对 $\delta<\frac{1}{10}$ 成立，最后一个不等式由 (39) 得出。结合 (40)–(42)，根据三角关系，我们有

$$950\Omega(\beta^{\tilde{t}})\leq3600ss_0\Delta.$$

回顾 $\Omega(\beta^t)>4ss_0\Delta$ 在假设 $\|\beta^t\|_G>4s$ 下。于是，我们有

$$3800ss_0\Delta<950\Omega(\beta^{\tilde{t}})\leq3600ss_0\Delta.$$

这与 $\|\beta^t\|_G>4s$ 的假设相矛盾。因此，我们必定有 $\|\beta^t\|_G\leq4s$。

˜接下来，我们给出估计误差 $\|\beta^t - \beta^*\|$ 的上界。由（40）可得

$$\begin{aligned}\left\|y - X\beta^{t\infty}\right\|_2^2 + 1000\sigma_{\tilde{t}}^2\Omega(\beta^{t\infty}) &\geq \left\|y - X\beta^{\tilde{t}}\right\|_2^2 + 1000\sigma_{\tilde{t}}^2\Omega(\beta^{\tilde{t}}) \\ &\geq \left\|y - X\beta^{\tilde{t}}\right\|_2^2 \\ &\geq \|\xi\|_2^2 + \left\|X(\beta^{\tilde{t}} - \beta^*)\right\|_2^2 - 2\sigma\sqrt{3\Omega^*(\beta^{\tilde{t}})}\left\|X(\beta^{\tilde{t}} - \beta^*)\right\|_2.\end{aligned}$$

结合上述不等式和式（42），可得

$$\left\|X(\beta^{\tilde{t}} - \beta^*)\right\|_2^2 - 2\sigma\sqrt{3\Omega^*(\beta^{\tilde{t}})}\left\|X(\beta^{\tilde{t}} - \beta^*)\right\|_2 \leq 3600\sigma_{\tilde{t}}^2 s s_0 \Delta. \tag{43}$$

解二次不等式（43），可得

$$\left\|X(\beta^{\tilde{t}} - \beta^*)\right\|_2 \leq 140\sigma\sqrt{s s_0 \Delta}.$$

˜注意到 β^t 是（$4s$， s_0）-稀疏的。因此，由 DSRIP 条件可得

$$\left\|\beta^{\tilde{t}} - \beta^*\right\|_2 \leq \frac{\left\|X(\beta^{\tilde{t}} - \beta^*)\right\|_2}{\sqrt{n(1-\delta)}} \leq 150\sigma\sqrt{\frac{s s_0 \Delta}{n}},$$

由此完成定理 5 的证明。

推论2的证明

注意到，以至少 $1 - p^{-C}$ 的概率，

$$\begin{aligned}\|M\|_\infty \leq& \|\beta^* + \Phi\beta^*\|_\infty + \|\Xi\|_\infty \\ \leq& \|\beta^* + \Phi\beta^*\|_2 + \|\Xi\|_\infty \\ \leq& (1+\delta)\|\beta^*\|_2 + 2\sigma\sqrt{\frac{\log ep}{n}} \\ \leq& 4\left(\|\beta^*\|_2 \vee \sigma\sqrt{\frac{\log ep}{n}}\right),\end{aligned} \tag{44}$$

其中最后一个 不等式使用了 $\delta \leq 1$。将（44）代入 λ_0 的定义，我们得到

$$\begin{aligned}\lambda_0 =& \frac{100}{9}\sigma_0\sqrt{\frac{\Delta'}{n}} \vee \frac{19}{4}\|M\|_\infty \\ \leq& 19\left(\|\beta^*\|_2 \vee \sigma\sqrt{\frac{\log ep}{n}}\right),\end{aligned} \tag{45}$$

其中，最后一个不等式使用了 $\Delta' \leq \log(ep)$。注意到 $\kappa^T \lambda_0 \leq 4\frac{\sigma_{\bar{t}}}{\sqrt{n}}$。通过一些简单代数运算，以至少 $1 - \exp\{-Css_0\Delta\}$ 的概率，我们有

$$\begin{aligned} T &\leq \log\left(\frac{\lambda_0\sqrt{n}}{4\sigma_{\bar{t}}}\right) / \log(\frac{1}{\kappa}) \\ &\leq \log\left(\frac{5\lambda_0\sqrt{n}}{18\sigma}\right) / \log(\frac{1}{\kappa}) \\ &\leq \log\left(6(\frac{\sqrt{n}\|\beta^*\|_2}{\sigma} \vee \sqrt{\log ep})\right) / \log(\frac{1}{\kappa}), \end{aligned}$$

其中，第二个不等式由 39 得出，最后一个不等式用到了 (45)。

因此，

$$\sup_{S^* \in \mathcal{S}^{m,d}(s,s_0)} P\left(T \geq \log\left(6(\frac{\sqrt{n}\|\beta^*\|_2}{\sigma} \vee \sqrt{\log ep})\right) / \log(\frac{1}{\kappa}) + 1\right) \leq e^{-Css_0\Delta}.$$

定理 6 的证明

我们用于调整 s_0 的技术与 Verzelen[2012] 的方法明显不同，也更加复杂。正如第 3.2 节所讨论的，$\bar{s}_0 > s_0$ 和 $\bar{s}_0 < s_0$ 两种情况下的理论性质存在显著差异。我们可以在逐元素或逐组层面控制稀疏性，但无法同时在两个层面进行控制。此外，如图 4 所示，不同 $\bar{s}_0$ 值对应的极小极大速率呈现“U 形”曲线，而非逐元素稀疏性下观察到的单调递增趋势 [Raskutti 等人，2011年，Verzelen， 2012]。因此，我们必须分别分析 $\bar{s}_0 > s_0$ 和 $\bar{s}_0 < s_0$ 两种情况。

Verzelen 方法的基本不等式

在本部分中，我们给出证明中使用的基本可比不等式。本部分与 Verzelen [2012] 中的定理 5.2 类似。记】【。

$$\mathrm{pen}\left(\bar{s}_0\right) = \frac{K}{n}\left(\hat{A}(\bar{s}_0)\log ed + \hat{s}(\bar{s}_0)\log\frac{em}{\hat{s}(\bar{s}_0)}\right), \quad \mathrm{pen}'\left(\bar{s}_0\right) = -1 + \exp\left(\mathrm{pen}\left(\bar{s}_0\right)\right).$$

根据 $\hat{s}_0$ 的定义，我们有

$$\frac{1}{n}\left\|y - X\hat{\beta}\left(\hat{s}_0\right)\right\|_2^2 \cdot \left(1 + \mathrm{pen}'(\hat{s}_0)\right) \leq \frac{1}{n}\left\|y - X\hat{\beta}\left(s_0\right)\right\|_2^2 \cdot \left(1 + \mathrm{pen}'\left(s_0\right)\right). \tag{46}$$

对于（46）的右侧，采用与（42）类似的策略可得

$$\frac{1}{n}\left\|y - X\hat{\beta}\left(s_0\right)\right\|_2^2 \leq \frac{1}{n}\|\xi\|_2^2 + C_1\frac{\sigma^2 ss_0\Delta}{n}. \tag{47}$$

回顾一下，我们假设 n 足够大，使得

$$\mathrm{pen}(s_0) \leq \frac{4K}{n}\left(ss_0 \log ed + s\log(em/s)\right) < 0.1,$$

这意味着 即 $1+\mathrm{pen}'(s_0) = \exp\{\mathrm{pen}(s_0)\} \leqslant e$。结合（46）和（47），我们有

$$\frac{1}{n}\left\|y - X\hat{\beta}\left(\hat{s}_0\right)\right\|_2^2 \cdot \left(1+\mathrm{pen}'\left(\hat{s}_0\right)\right) \leq \frac{1}{n}\|\xi\|_2^2\left(1+\mathrm{pen}'\left(s_0\right)\right) + C_2\frac{\sigma^2 ss_0\Delta}{n}. \tag{48}$$

对于（46）的左侧，采用与（41）类似的策略可得

$$\begin{aligned}\frac{1}{n}\left\|y - X\hat{\beta}\left(\hat{s}_0\right)\right\|_2^2 \geq \frac{1}{n}\|\xi\|_2^2 + \frac{1}{n}\left\|X\left(\beta^* - \hat{\beta}\left(\hat{s}_0\right)\right)\right\|_2^2 \\ - \frac{2}{n}\left\|X\left(\beta^* - \hat{\beta}\left(\hat{s}_0\right)\right)\right\|_2 \cdot \left|\left\langle \xi, \frac{X\left(\beta^* - \hat{\beta}\left(\hat{s}_0\right)\right)}{\left\|X\left(\beta^* - \hat{\beta}\left(\hat{s}_0\right)\right)\right\|_2}\right\rangle\right|.\end{aligned} \tag{49}$$

然后，根据引理 4，我们也可以将内积上界为

$$\begin{aligned}\left|\left\langle \frac{\xi}{\sigma}, \frac{X\left(\beta^* - \hat{\beta}\left(\hat{s}_0\right)\right)}{\left\|X\left(\beta^* - \hat{\beta}\left(\hat{s}_0\right)\right)\right\|_2}\right\rangle\right|^2 &\precsim \left(ss_0 + \hat{A}(\hat{s}_0)\right)\log\frac{ed(s+\hat{s}(\hat{s}_0))}{ss_0+\hat{A}(\hat{s}_0)} + (s+\hat{s}(\hat{s}_0))\log\frac{em}{s+\hat{s}(\hat{s}_0)} \\ &\leq \left(ss_0 + \hat{A}(\hat{s}_0)\right)\log ed + (s+\hat{s}(\hat{s}_0))\log\frac{em}{s+\hat{s}(\hat{s}_0)}.\end{aligned}$$

记 $\mathcal{L} := \left\|X\left(\beta^* - \hat{\beta}\left(\hat{s}_0\right)\right)\right\|_2$ and $\hat{\Gamma} := \left(ss_0 + \hat{A}(\hat{s}_0)\right)\log ed + (s+\hat{s}(\hat{s}_0))\log\frac{em}{s+\hat{s}(\hat{s}_0)}$。

因此，由（46）–（50），我们得到

$$\mathcal{L}^2 - C_2'\sigma\sqrt{\hat{\Gamma}}\mathcal{L} \leq \mathrm{pen}'(s_0)\|\xi\|_2^2 + C_2\sigma^2 ss_0\Delta. \tag{51}$$

根据 Laurent 和 Massart 的引理 1 [2000], ，我们得出 $0.9\sigma^2 \leq \frac{1}{n}\|\xi\|_2^2 \leq 1.1\sigma^2$以至少 $1-\exp(-C_4 n)$ 的概率成立。此外，由于 $\mathrm{pen}(s_0) < 0.1$，可推得 $\mathrm{pen}'(s_0) = \exp(\mathrm{pen}(s_0)) - 1 \leq 2\mathrm{pen}(s_0)$，因此

$$\mathrm{pen}'(s_0)\|\xi\|_2^2 + C_2\sigma^2 ss_0\Delta \leq C_3\sigma^2\left(ss_0\log d + s\log(em/s)\right) \leq C_3\sigma^2\hat{\Gamma},$$

由此得出

$$\mathcal{L}^2 - C_2'\sigma\sqrt{\hat{\Gamma}}\mathcal{L} \leq C_3\sigma^2\hat{\Gamma}. \tag{52}$$

通过求解不等式（52），我们得到上界 $\mathcal{L}^2 \leq C_4\sigma^2\hat{\Gamma}$。因此，要获得估计误差的最优上界，我们只需证明

$$\hat{\Gamma} \precsim ss_0\log ed + s\log(em/s). \tag{53}$$

ˆ 到目前为止，根据表1，我们知道对于 ˆ$s_0 \le s_0$， A(ˆs_0) $\le 4ss_0$ ，而对于 ˆ$s_0 \ge s_0$, s(ˆs_0) $\le 4s$ 。因此，以高概率，我们得出结论：

$$\hat{\Gamma} \le \begin{cases} 5ss_0 \log ed + (s + \hat{s}(\hat{s}_0)) \log \frac{em}{s+\hat{s}(\hat{s}_0)}, & \hat{s}_0 \le s_0. \\ \left(ss_0 + \hat{A}(\hat{s}_0)\right) \log ed + 5s \log \frac{em}{s}, & \hat{s}_0 > s_0. \end{cases} \tag{54}$$

为方便起见，我们将接下来的证明分为两种情形： $ss_0 \log\, ed \le s \log\, \frac{em}{s}$或 $ss_0 \log\, ed \ge s \log \frac{em}{s}$ 。

假设 A： $ss_0 \log\, ed \ge s \log\, \frac{em}{s}$。 —

情形1： ˆ$s_0 \ge s_0$。

ˆ ˆ 根据 (54)，我们需要证明 A(ˆs_0) $\le 9ss_0$。通过反证法，首先假设 A(ˆs_0) $> 9ss_0$成立，并得到：

$$\begin{aligned} \frac{1 + \mathrm{pen}'(\hat{s}_0)}{1 + \mathrm{pen}'(s_0)} &\ge \exp\left\{ \frac{K}{n} \left(\hat{A}(\hat{s}_0) \log ed + \hat{s}(\hat{s}_0) \log \frac{em}{\hat{s}(\hat{s}_0)} \right) - 4\frac{K}{n} \left(ss_0 \log ed + s \log \frac{em}{s} \right) \right\} \\ &\ge \exp\left\{ \frac{K}{n} \hat{A}(\hat{s}_0) \log ed - 4\frac{K}{n} \left(ss_0 \log ed + s \log \frac{em}{s} \right) \right\} \\ &\ge \exp\left\{ \frac{K}{n} \hat{A}(\hat{s}_0) \log ed - 8\frac{K}{n} ss_0 \log ed \right\} \\ &\ge \exp\left\{ \frac{K}{9n} \hat{A}(\hat{s}_0) \log ed \right\}. \end{aligned} \tag{55}$$

结合 (47)，我们有

$$\frac{1}{n} \left\| y - X\hat{\beta}\left(s_0\right) \right\|_2^2 \le \frac{1}{n} \|\xi\|_2^2 + C_1 \frac{\sigma^2 ss_0 \Delta}{n} \le \frac{1}{n} \|\xi\|_2^2 + C_6 \sigma^2 \frac{\hat{A}(\hat{s}_0) \log ed}{n}. \tag{56}$$

此外，根据 (49)，我们还有

$$\frac{1}{n} \left\| y - X\hat{\beta}\left(\hat{s}_0\right) \right\|_2^2 \ge \frac{1}{n} \|\xi\|_2^2 - \frac{1}{n} C_7 \sigma^2 \left(\hat{A}(\hat{s}_0) \log ed \right), \tag{57}$$

其中我们使用了 $a^2 - 2ab \ge -b^2$，且（49）中的内积项由（50）给出上界。因此，结合（46）和（55）–（57），可得

$$\left(\frac{1}{n} \|\xi\|_2^2 - \frac{C_7 \sigma^2}{n} \hat{A}(\hat{s}_0) \log ed \right) \exp\left\{ \frac{K}{9n} \hat{A}(\hat{s}_0) \log ed \right\} \le \frac{1}{n} \|\xi\|_2^2 + \frac{C_6 \sigma^2}{n} \hat{A}(\hat{s}_0) \log ed. \tag{58}$$

令 $t = \frac{1}{9n}$ˆA(ˆs_0) log ed。注意，根据假设 2， n 足够大，使得 $t \in (0, \frac{1}{K})$。

为导出矛盾，我们需要验证对于 $t \in (0, \frac{1}{K})$，

$$F(t) = \left(1 - \frac{9C_7 t}{\|\xi\|_n^2/\sigma^2}\right)\exp(Kt) - \left(1 + \frac{9C_6 t}{\|\xi\|_n^2/\sigma^2}\right) > 0 \tag{59}$$

始终成立。注意，$F(0) = 0$ 且 $F'(t) = \exp(Kt)\left\{K - \frac{9C_7(Kt+1)}{\|\xi\|_n^2/\sigma^2}\right\} - \frac{9C_6}{\|\xi\|_n^2/\sigma^2}$。由 $0.9\sigma^2 \le \frac{1}{n}\|\xi\|_2^2 \le 1.1\sigma^2$ 和 $Kt \in (0,1)$，我们可以选取充分大的 $K \ge 10C_6 + 20C_7$。因此，我们验证了对于 $\forall t \in (0, \frac{1}{K})$，$F'(t) > 0$，这以高概率导致与 (58) 矛盾。

因此，我们证明 $A(\hat{s}_0) \le 9ss_0$ 以高概率成立。然后，基于 (54)，我们推导出

$$\hat{\Gamma} \le 10\left(ss_0 \log ed + s\log(em/s)\right), \ \forall \hat{s}_0 \ge s_0, \tag{60}$$

这证明 (53) 以高概率成立。

情形 2： $\hat{s}_0 < s_0$.

我们将这种情况分为两个子情况，并分别进行分析。

(a) 如果 $9ss_0 \log ed > \hat{s}(\hat{s}_0) \log \frac{em}{\hat{s}(\hat{s}_0)}$，我们只需将 (50) 中的内积界定为:

$$\begin{aligned}\left|\left\langle \xi, \frac{X\left(\beta^* - \hat{\beta}\left(\hat{s}_0\right)\right)}{\left\|X\left(\beta^* - \hat{\beta}\left(\hat{s}_0\right)\right)\right\|_2}\right\rangle\right|^2 &\le 5ss_0 \log ed + (s + \hat{s}(\hat{s}_0))\log\frac{em}{(s+\hat{s}(\hat{s}_0))} \\ &\le 5ss_0\log ed + s\log(em/s) + 9ss_0\log ed \\ &\le 14\left(ss_0\log ed + s\log(em/s)\right),\end{aligned}$$

其中，第一个不等式利用了对于 $\hat{s}_0 < s_0$，$A(\hat{s}_0) < 4ss_0$；第二个不等式利用了 $\hat{s}(\hat{s}_0)\log\frac{em}{\hat{s}(\hat{s}_0)} < 9ss_0\log ed$。通过求解（51），我们推导出 $\mathcal{L}$ 的一个上界，如下:

$$\mathcal{L}^2 \precsim \frac{\sigma^2}{n}\left(ss_0\log ed + s\log\frac{em}{s}\right), \tag{61}$$

因此，根据 DSRIP 条件，我们证明了（14）。

(b) 若 $9ss_0\log ed \le \hat{s}(\hat{s}_0)\log\frac{em}{\hat{s}(\hat{s}_0)}$。那么，与情形1类似，我们将以高概率得到矛盾。首先，我们得到

$$\begin{aligned}\frac{1+\mathrm{pen}'(\hat{s}_0)}{1+\mathrm{pen}'(s_0)} &\ge \exp\left\{\frac{K}{n}\hat{s}(\hat{s}_0)\log\frac{em}{\hat{s}(\hat{s}_0)} - 8\frac{K}{n}ss_0\log ed\right\} \\ &\ge \exp\left\{\frac{K}{9n}\hat{s}(\hat{s}_0)\log\frac{em}{\hat{s}(\hat{s}_0)}\right\}.\end{aligned}$$

然后，采用与 (56) 和 (57) 中类似的技巧，我们得到以下不等式：

$$\begin{aligned}\frac{1}{n}\left\|y-X\hat{\beta}\left(s_0\right)\right\|_2^2 &\leq \frac{1}{n}\|\xi\|_2^2+C_1\frac{\sigma^2 s s_0\Delta}{n}\\ &\leq \frac{1}{n}\|\xi\|_2^2+\frac{2C_1}{9n}\sigma^2\hat{s}(\hat{s}_0)\log\frac{em}{\hat{s}(\hat{s}_0)},\\ \frac{1}{n}\left\|y-X\hat{\beta}\left(\hat{s}_0\right)\right\|_2^2 &\geq \frac{1}{n}\|\xi\|_2^2-\frac{3\sigma^2}{n}\left(\left(s+\hat{s}(\hat{s}_0)\right)\log\frac{em}{s+\hat{s}(\hat{s}_0)}+4ss_0\log ed\right)\\ &\geq \frac{1}{n}\|\xi\|_2^2-\frac{14\sigma^2}{3n}\hat{s}(\hat{s}_0)\log\frac{em}{\hat{s}(\hat{s}_0)},\end{aligned} \tag{62}$$

因此

$$\begin{aligned}&\left(\frac{1}{n}\|\xi\|_2^2-\frac{14\sigma^2}{3n}\hat{s}(\hat{s}_0)\log\frac{em}{\hat{s}(\hat{s}_0)}\right)\exp\left\{\frac{K}{9n}\hat{s}(\hat{s}_0)\log\frac{em}{\hat{s}(\hat{s}_0)}\right\}\\ \leq&\frac{1}{n}\|\xi\|_2^2+\frac{2C_1}{9n}\sigma^2\hat{s}(\hat{s}_0)\log\frac{em}{\hat{s}(\hat{s}_0)}.\end{aligned} \tag{63}$$

现在令 $t=\frac{1}{9n}\hat{s}(\hat{s}_0)\log\frac{em}{\hat{s}(\hat{s}_0)}$，并采用与 (59) 对应的相同技巧，取足够大的 $K\geq\frac{20C_1+840}{9}$，即可得到矛盾。因此，我们证明以高概率有 $9ss_0\log ed>\hat{s}(\hat{s}_0)\log\frac{em}{\hat{s}(\hat{s}_0)}$ 成立，由此得到 (61)，并完成情形 2 的证明。

至此，我们已完成**假设 A：** $ss_0\log ed\geq s\log\frac{em}{s}$ 时的证明。当 $ss_0\log ed<s\log\frac{em}{s}$ 时，证明策略类似，下面仅给出证明概要。

假设 B： $ss_0\log ed<s\log\frac{em}{s}$。 —

情形 3： $\hat{s}_0\geq s_0$。
与情形 2 类似，我们继续将此情形分为两个子情形：

(a) 若 $A(\hat{s}_0)\log ed\leq 9s\log(em/s)$，则我们得到

$$\begin{aligned}\hat{\Gamma}&=\left(ss_0+\hat{A}(\hat{s}_0)\right)\log ed+(s+\hat{s}(\hat{s}_0))\log\frac{em}{s+\hat{s}(\hat{s}_0)}\\ &\leq ss_0\log ed+9s\log(em/s)+5s\log(em/s)\\ &\leq 14\left(ss_0\log ed+s\log\frac{em}{s}\right).\end{aligned} \tag{64}$$

因此，我们证明了式 (53) 成立。

(b) 如果 $A(\hat{s}_0)\log ed>9s\log(em/s)$，我们证明了

$$\frac{1+\mathrm{pen}'\left(\hat{s}_0\right)}{1+\mathrm{pen}'\left(s_0\right)}\geq\exp\left\{\frac{K}{9n}\hat{A}(\hat{s}_0)\log ed\right\}.$$

ˆ 因此，通过采用与 (62) 和 (63) 类似的策略，我们证明 $A(\hat{s}_0) \log ed > 9s \log(em/s)$ 不可能以高概率成立。因此，在情形 3 中，(64) 以高概率成立，这证明了 (53) 成立。

情况 4： $\hat{s}_0 < s_0$.

在这种情况下，我们只需要控制 $\hat{s}(\hat{s}_0)$。通过使用与情形1类似的矛盾法，首先假设 $\hat{s}(\hat{s}_0) \geq 9s$ 成立，这意味着

$$\begin{aligned}\frac{1+\mathrm{pen}'(\hat{s}_0)}{1+\mathrm{pen}'(s_0)} &\geq \exp\left\{\frac{K}{n}\left(\hat{A}(\hat{s}_0)\log ed + \hat{s}(\hat{s}_0)\log\frac{em}{\hat{s}(\hat{s}_0)}\right) - 4\frac{K}{n}\left(ss_0\log ed + s\log\frac{em}{s}\right)\right\} \\ &\geq \exp\left\{\frac{K}{9n}\hat{s}(\hat{s}_0)\log\frac{em}{\hat{s}(\hat{s}_0)}\right\},\end{aligned} \tag{65}$$

并且通过采用与（56）—（59）类似的策略，我们证明 $\hat{s}(\hat{s}_0) \geq 9s$ 以高概率不成立。因此，我们得到

$$\begin{aligned}\hat{\Gamma} &= \left(ss_0 + \hat{A}(\hat{s}_0)\right)\log ed + (s+\hat{s}(\hat{s}_0))\log\frac{em}{s+\hat{s}(\hat{s}_0)} \\ &\leq 5ss_0\log ed + 10s\log(em/s) \\ &\leq 10\left(ss_0\log ed + s\log\frac{em}{s}\right),\ \forall \hat{s}_0 < s_0,\end{aligned} \tag{66}$$

这就得到（53）。

总体而言，综合这4种情况，我们推导出 $\mathcal{L}$ 的上界。

$$\mathcal{L}^2 \precsim \frac{\sigma^2\left(ss_0\log ed + s\log\frac{em}{s}\right)}{n},$$

并且根据 DSRIP 条件，我们最终完成了定理 6 的证明。

定理 7 的证明

我们采用与定理 1 类似的策略来证明这些结果。为便于展示，我们定义 $\Upsilon(A,\tilde{\beta}^t) := \sum_{(i,j)\in A}\langle\Phi_{ij}^\top, \beta^* - \tilde{\beta}^t\rangle^2$。具体而言，如果 $A \cup \mathrm{supp}(\beta^* - \tilde{\beta}^t) \in \mathcal{S}^{m,d}(3s, \frac{5s_0}{3})$，根据 DSRIP$(3s, \frac{5}{3}s_0, \delta)$ 条件，我们有 $\Upsilon(A,\tilde{\beta}^t) \leq \delta^2\|\beta^* - \tilde{\beta}^t\|_2^2$。在定理 7 和 8 的证明中（以及引理 5–7 中），我们使用双重指标 (i,j) 来表示第 j 组的第 i 个条目（变量） G_j。首先，我们给出本文中经常使用的概率不等式

证明：

$$
\begin{aligned}
P\left\{\forall S \in \mathcal{S}(s', s_0),\ \|\Xi_S\|_2^2 > \frac{6\sigma^2 s' s_0}{n}\Delta(s', s_0)\right\} &= o(1), \quad \text{where } s' = \frac{s}{8\Delta^2};\\
P\left\{\sum_{(i,j)\in S_{G^*}} \Xi_{ij}^2 \mathrm{I}\Big\{|\Xi_{ij}| \geq \tilde{\lambda}_1\Big\} \geq \frac{\sigma^2 s s_0}{n\Delta}\right\} &= o(1);\\
P\left(\sum_{(i,j)\in S^*} \tilde{\lambda}_2^2 \cdot \mathrm{I}\Big(|\Xi_{ij}| > \frac{\epsilon}{2}\tilde{\lambda}_2\Big) \geq \frac{\sigma^2 s s_0}{n\Delta}\right) &= o(1);\\
P\left\{\sum_{j\in G^*} s_0\tilde{\lambda}_2^2 \cdot \mathrm{I}\left(\sum_{k\in S^*\cap S_{G_j}} \Xi_{kj}^2 > \frac{\epsilon^2}{4}(s_j \vee s_0)\tilde{\lambda}_2^2\right) \geq \frac{\sigma^2 s s_0}{n\Delta}\right\} &= o(1);\\
P\left(\|\Xi_{S^*}\|_2^2 \geq \frac{2\sigma^2 s s_0}{n}\right) &= o(1),
\end{aligned}
\tag{67}
$$

作为 $\min\{\Delta,\ ss_0/\Delta\} \to \infty$。定义 $\Delta(s',\ s_0) := \frac{1}{s_0}\log\binom{em}{s'} + \log\binom{ed}{s_0}$. 我们在附录 B 中给出上述不等式的证明。

这里我们通过数学归纳法证明定理 7。根据假设，初始估计量 $\tilde{\beta}^0 = \hat{\beta}$ 是 $(2s,\ \frac{3}{2}s_0)$-稀疏且极小极大最优的。不难验证，定理 7 中的三个结论对 $t = 0$ 均成立。现在，对于 $\forall t \geq 0$，假设定理 7 中的结论对第 t 次迭代成立，我们将通过反证法和归纳法证明这些结论对第 $(t+1)$ 次迭代也成立。

步骤 1（控制误发现的组）。

假设在第 $(t+1)$ 次迭代中误发现的组超过 s 组。那么，我们总能任意选取 s 个误发现的组，并构造一个 $(s,\ s_0)$-稀疏集$S'_{OG} \in \tilde{S}^{t+1} \cap S^c_{G^*}$。选择过程的具体细节如下：

对于任意选定的组 $j \notin G^*$，如果其中误发现条目的数量超过 s_0，则从这些误发现条目中任意选取 s_0 个非零条目放入 S'_{OG}；如果数量少于 s_0，则将所有这些误发现条目放入 S'_{OG}。对于任意 s 个误发现的组，我们重复此操作 s 次，从而得到一个 $(s,\ s_0)$-稀疏集S'_{OG}。

然后，根据 DSIHT 算子 $\mathcal{T}_{s_0,\tilde{\lambda}_2}(\cdot)$ 的定义，对于任意错误发现的

选入集合 S'_{OG} 的组 j，我们有 $\|\tilde{\beta}^{t+1}_{G_j\cap S'_{OG}}\|_2^2 \geq s_0\tilde{\lambda}_2^2$，由此可得

$$\begin{aligned}\sqrt{ss_0}\tilde{\lambda}_2 \leq & \sqrt{\Upsilon\left(S'_{OG},\tilde{\beta}^t\right)} + \sqrt{\sum_{(i,j)\in S'_{OG}} \Xi_{ij}^2 \mathrm{I}\left\{\mathcal{T}_{\lambda_2,s_0}\left(\tilde{H}^{t+1}\right)_{ij}\neq 0\right\}} \\ \overset{(i)}{\leq} & \frac{7}{2}\delta\|\tilde{\beta}^t-\beta^*\|_2 + \sqrt{\frac{\sigma^2 ss_0}{n\Delta}},\end{aligned} \tag{68}$$

其中不等式 (i) 由引理 5 得出。根据数学归纳法的假设，由于 (17) 对 t 次迭代成立，因此我们有 $\left\|\tilde{\beta}^t-\beta^*\right\|_2 < 16\sigma\sqrt{\frac{ss_0\Delta}{n}} + 16\sqrt{\frac{\sigma^2 ss_0}{n}}$。结合 (68)，得到

$$\sqrt{ss_0}\tilde{\lambda}_2 = \sqrt{\frac{32\sigma^2 ss_0\Delta}{n}} < 2.8\sqrt{\frac{\sigma^2 ss_0\Delta}{n}} + 3.8\sqrt{\frac{\sigma^2 ss_0}{n}}, \tag{69}$$

当 $\Delta > 2.5$ 时，这不可能成立。因此得出矛盾。

我们已经证明，在第 $(t+1)$ 次迭代中，错误发现的组不超过 s 个。接下来，我们将证明，在真实组 G^* 之外，错误发现的条目不超过 ss_0 个。否则，我们可以构造一个 $(s,\ s_0)$-稀疏集 $S''_{OG} \in S^t \cap S^c_{G^*}$。于是得到

$$\begin{aligned}\sqrt{ss_0}\tilde{\lambda}_2 \leq & \sqrt{\Upsilon\left(S''_{OG},\tilde{\beta}^t\right)} + \sqrt{\sum_{(i,j)\in S''_{OG}} \Xi_{ij}^2 \mathrm{I}\left\{\mathcal{T}_{\lambda_2,s_0}\left(\tilde{H}^{t+1}\right)_{ij}\neq 0\right\}} \\ \overset{(i)}{\leq} & \frac{7}{2}\delta\|\tilde{\beta}^t-\beta^*\|_2 + \sqrt{\frac{\sigma^2 ss_0}{n\Delta}},\end{aligned} \tag{70}$$

其中不等式 (i) 由引理 5 得出。这与 (69) 矛盾。

步骤 2（控制 S_{G^*} 中的误发现条目）。

假设真实组内有超过 ss_0 个误发现条目 G^*。那么，我们可以构造一个 $(s,\ s_0)$-稀疏集 $S_{IG} \in S_{G^*} \cap S^t \cap (S^*)^c$，使得对于每个

S_{IG}、 $|\tilde{\beta}_{ij}^{t+1}| = |\Xi_{ij} + \langle \Phi_{ij}^{\top}, \ \beta^* - \tilde{\beta}^t \rangle| \geq \tilde{\lambda}_2$ 中的条目始终成立，这意味着

$$\begin{aligned}
\sqrt{ss_0}\tilde{\lambda}_2 \leq & \sqrt{\Upsilon\left(S_{IG}, \tilde{\beta}^t\right)} + \sqrt{\sum_{(i,j)\in S_{IG}} \Xi_{ij}^2 \mathrm{I}\Big\{|\Xi_{ij} + \langle \Phi_{ij}^{\top}, \beta^* - \tilde{\beta}^t \rangle| \geq \tilde{\lambda}_2\Big\}} \\
\leq & \delta \|\tilde{\beta}^t - \beta^*\|_2 + \sqrt{\sum_{(i,j)\in S_{G^*}} \Xi_{ij}^2 \mathrm{I}\Big\{|\Xi_{ij}| \geq \tilde{\lambda}_1\Big\}} \\
& + \sqrt{\sum_{(i,j)\in S_{IG}} \Xi_{ij}^2 \mathrm{I}\Big\{|\Xi_{ij}| < \tilde{\lambda}_1 < |\langle \Phi_{ij}^{\top}, \beta^* - \tilde{\beta}^t \rangle|\Big\}} \\
\overset{(i)}{<} & 2\delta \|\tilde{\beta}^t - \beta^*\|_2 + \sqrt{\frac{\sigma^2 ss_0}{n\Delta}},
\end{aligned} \tag{71}$$

其中，不等式 (i) 由引理 6 得出。由于式 (17) 对第 t 次迭代成立，因此再次得到与式 (69) 一样的矛盾。

步骤 3（ℓ_2 $\tilde{\beta}^{t+1}$ 的估计误差）。

现在我们已经证明了定理 7 中的前两个结论在第$(t+1)$次迭代中仍然成立，接下来我们将证明第三个结论在第$(t+1)$次迭代中也成立。注意，

$$\tilde{\beta}_{ij}^{t+1} - \beta_{ij}^* = -\tilde{H}_i^{t+1} \cdot \mathrm{I}\left((i,j) \notin \tilde{S}^{t+1}\right) + \langle \Phi_{kj}^{\top}, \beta^* - \tilde{\beta}^t \rangle + \Xi_{ij} \tag{72}$$

现在我们分别关注 S^* 和 $\tilde{S}^{t+1} \cap (S^*)^c$ 上的估计误差。在 S^*上，我们有

$$\begin{aligned}
\|\tilde{\beta}_{S^*}^{t+1} - \beta_{S^*}^*\|_2 \leq & \sqrt{\sum_{(i,j)\in S^*} \left(\tilde{H}_{ij}^{t+1}\right)^2 \mathrm{I}\left((i,j) \notin \tilde{S}^{t+1}\right)} + \sqrt{\Upsilon\left(S^*, \tilde{\beta}^t\right)} + \sqrt{\sum_{(i,j)\in S^*} \Xi_{ij}^2} \\
\overset{(i)}{\leq} & \frac{4}{\epsilon}\delta \left\|\tilde{\beta}^t - \beta^*\right\|_2 + 2\sqrt{\frac{\sigma^2 ss_0}{n\Delta}} + \sqrt{\Upsilon\left(S^*, \tilde{\beta}^t\right)} + \sqrt{\sum_{(i,j)\in S^*} \Xi_{ij}^2} \\
\overset{(ii)}{\leq} & \left(\frac{4}{\epsilon} + 1\right)\delta \left\|\tilde{\beta}^t - \beta^*\right\|_2 + 2\sqrt{\frac{\sigma^2 ss_0}{n\Delta}} + \sqrt{\frac{2\sigma^2 ss_0}{n}},
\end{aligned} \tag{73}$$

其中，不等式 (i) 使用了引理 7 的结果。不等式 (ii) 使用了 Hsu 等人的引理 1 和定理 2.1。 [2012], 也就是说，对于一个固定集合 $S^* \in \mathcal{S}^{m,\ d}(s,\ s_0)$ 以及任意 $t > 0$，我们有

$$P\left(\frac{n}{\sigma^2}\|\Xi_{S^*}\|_2^2 \geq ss_0 + 2(1+\delta)\sqrt{ss_0 t} + 2(1+\delta)t\right) \leq e^{-t}. \tag{74}$$

令 $t = \frac{ss_0}{10}$。根据 $\delta < \frac{1}{5}$，我们得到 $P\left(\|\Xi_{S^*}\|_2^2 \geq \frac{2\sigma^2 ss_0}{n}\right) \leq \exp\left(-\frac{ss_0}{10}\right) = o(1)$，当 $ss_0 \to \infty$ 时。

在 $\tilde{S}^{t+1}\cap(S^*)^c$ 上，当 X 满足 DSRIP$(3s,\ \frac{5}{3}s_0,\ \delta)$ 时，我们有

$$
\begin{aligned}
&\left\|\tilde{\beta}^{t+1}_{\tilde{S}^{t+1}\cap(S^*)^c}-\beta^*_{\tilde{S}^{t+1}\cap(S^*)^c}\right\|_2=\left\|\tilde{\beta}^{t+1}_{\tilde{S}^{t+1}\cap(S^*)^c}\right\|_2\\
\leq&\sqrt{\Upsilon\left(\tilde{S}^{t+1}\cap(S^*)^c,\tilde{\beta}^t\right)}+\sqrt{\sum_{(i,j)\in\tilde{S}^{t+1}\cap(S^*)^c}\Xi_{ij}^2\mathrm{I}\left(T_{\lambda_2,s_0}\left(\tilde{H}^{t+1}\right)_{ij}\neq 0\right)}\\
\leq&\delta\left\|\tilde{\beta}^t-\beta^*\right\|_2+\sqrt{\sum_{(i,j)\in S_{OG}}\Xi_{ij}^2\mathrm{I}\left(T_{\lambda_2,s_0}\left(\tilde{H}^{t+1}\right)_{ij}\neq 0\right)}+\sqrt{\sum_{(i,j)\in S_{IG}}\Xi_{ij}^2\mathrm{I}\left(T_{\lambda_2,s_0}\left(\tilde{H}^{t+1}\right)_{ij}\neq 0\right)}\\
\overset{(i)}{<}&\frac{9}{2}\delta\left\|\tilde{\beta}^t-\beta^*\right\|_2+2\sqrt{\frac{\sigma^2ss_0}{n\Delta}},
\end{aligned}
\tag{75}
$$

其中，不等式 (i) 由引理 5、(71) 式和引理 6 得出。

最后，根据（73）和（75），我们有

$$
\begin{aligned}
\left\|\tilde{\beta}^{t+1}-\beta^*\right\|_2\leq&\left\|\tilde{\beta}^{t+1}_{S^*}-\beta^*_{S^*}\right\|_2+\left\|\tilde{\beta}^{t+1}_{\tilde{S}^{t+1}\cap(S^*)^c}-\beta^*_{\tilde{S}^{t+1}\cap(S^*)^c}\right\|_2\\
\leq&\left(\frac{4}{\epsilon}+\frac{11}{2}\right)\delta\left\|\tilde{\beta}^t-\beta^*\right\|_2+4\sqrt{\frac{\sigma^2ss_0}{n\Delta}}+\sqrt{\frac{2\sigma^2ss_0}{n}}\\
\overset{(i)}{<}&\left(\frac{4}{\epsilon}+\frac{11}{2}\right)\delta\left\|\tilde{\beta}^t-\beta^*\right\|_2+4\sqrt{\frac{\sigma^2ss_0}{n}},
\end{aligned}
\tag{76}
$$

其中，在不等式 (i) 中，我们假设 $\Delta>2.5$，由此得到 $\frac{4}{\epsilon}+\sqrt{2}<4$。

然后，根据已初始化的不等式 $\left\|\tilde{\beta}^0-\beta^*\right\|_2\leq 16\sigma\sqrt{\frac{ss_0\Delta}{n}}$ 和 $\delta\leq\epsilon^4\wedge 0.05$，我们有 $\delta\left(\frac{4}{\epsilon}+\frac{11}{2}\right)\leq\frac{3}{4}$，由此得到

$$
\left\|\tilde{\beta}^{t+1}-\beta^*\right\|_2<\ 16\left(\frac{3}{4}\right)^{t+1}\sigma\sqrt{\frac{ss_0\Delta}{n}}+16\sqrt{\frac{\sigma^2ss_0}{n}}.
\tag{77}
$$

因此，我们证明了定理 7 中的结论对第 $(t+1)$ 次迭代成立，从而完成证明。

定理8的证明

在定理 7 的条件下，注意到 (67) 中的概率不等式仍然成立。

步骤 1（尖锐上界）。

令 $t > 2\log(256\Delta)$，则有

$$16\left(\frac{3}{4}\right)^t \sigma\sqrt{\frac{ss_0\Delta}{n}} < \sqrt{\frac{\sigma^2 ss_0}{n}}. \tag{78}$$

由（17）可知，

$$\left\|\tilde{\beta}^t - \beta^*\right\|_2 < \sqrt{\frac{\sigma^2 ss_0}{n}} + 16\sqrt{\frac{\sigma^2 ss_0}{n}} = 17\sqrt{\frac{\sigma^2 ss_0}{n}}. \tag{79}$$

步骤 2（逐组近乎完全恢复）。

注意，根据定理 7 的第一条结论，在第（$t+1$）次迭代中，错误发现的组不超过 s 个。记 G_{FD}^{t+1} 为第（$t+1$）次迭代中错误发现的组索引集，其满足 $\left|G_{FD}^{t+1}\right| < s$。于是，我们有

$$\begin{aligned}
\|\tilde{\eta}_G^{t+1} - \eta_G^*\|_0 &= \sum_{j=1}^{m} |(\tilde{\eta}_G^{t+1})_j - (\eta_G^*)_j| \\
&= \sum_{j\in G^*} |(\tilde{\eta}_G^{t+1})_j - 1| + \sum_{j\in G_{FD}^{t+1}} |(\tilde{\eta}_G^{t+1})_j - 0| \\
&= \sum_{j\in G^*} \mathrm{I}\left(\mathcal{T}_{\tilde{\lambda}_2,s_0}\left(\tilde{H}_{G_j}^{t+1}\right) = \mathbf{0}\right) + \sum_{j\in G_{FD}^{t+1}} \mathrm{I}\left(\mathcal{T}_{\tilde{\lambda}_2,s_0}\left(\tilde{H}_{G_j}^{t+1}\right) \neq \mathbf{0}\right).
\end{aligned} \tag{80}$$

对于式 (80) 中的第一项，根据引理 7，我们有

$$
\begin{aligned}
&\sum_{j\in G^*} \mathrm{I}\left(\mathcal{T}_{\tilde{\lambda}_2,s_0}(\tilde{H}_{G_j}^{t+1}) = \mathbf{0}\right) \\
\leq &\sum_{j\in G^*} \mathrm{I}\left(\sum_{k\in S_{G_j}\cap S^*} \left(\tilde{H}_{kj}^{t+1}\right)^2 \mathrm{I}\left(|\tilde{H}_{kj}^{t+1}| \geq \tilde{\lambda}_2\right) < s_0\tilde{\lambda}_2^2\right) \\
\leq &\sum_{j\in G^*} \mathrm{I}\left(\sum_{k\in S_{G_j}\cap S^*} \left(\tilde{H}_{kj}^{t+1}\right)^2 < (s_0+s_j)\tilde{\lambda}_2^2\right) \\
\overset{(i)}{\leq} &\sum_{j\in G^*} \mathrm{I}\left(\Upsilon\left(S_{G_j}\cap S^*, \tilde{\beta}^t\right) > \frac{\epsilon^2}{4}(s_j\vee s_0)\tilde{\lambda}_2^2\right) + \sum_{j\in G^*} \mathrm{I}\left(\sum_{k\in S_{G_j}\cap S^*} \Xi_{kj}^2 > \frac{\epsilon^2}{4}(s_j\vee s_0)\tilde{\lambda}_2^2\right) \\
\overset{(ii)}{\leq} &\frac{4\delta^2\left\|\tilde{\beta}^t-\beta^*\right\|_2^2}{\epsilon^2 s_0\tilde{\lambda}_2^2} + \frac{\sigma^2 s}{n\tilde{\lambda}_2^2\Delta} \\
\precsim &\frac{s}{\Delta} + \frac{s}{\Delta^2} \\
= &O\left(\frac{s}{\Delta}\right), \quad \text{as } \Delta\to\infty,
\end{aligned}
\tag{81}
$$

w这里，不等式 (i) 由 (108) 得出，而不等式 (ii) 由 (112) 得出。).

对于（80）中的第二项，根据引理 5，我们有

$$
\begin{aligned}
&\sum_{j\in G_{FD}^{t+1}} \mathrm{I}\left(\mathcal{T}_{\tilde{\lambda}_2,s_0}(\tilde{H}_{G_j}^{t+1}) \neq \mathbf{0}\right) \\
\leq &\sum_{j\in G_{FD}^{t+1}} \mathrm{I}\left(\mathcal{T}_{\tilde{\lambda}_1,s_0}(\Xi_{\tilde{S}^{t+1}\cap S_{G_j}}) \neq \mathbf{0}\right) + \sum_{j\in G_{FD}^{t+1}} \mathrm{I}\left(\mathcal{T}_{\tilde{\lambda}_1,s_0}(\Xi_{\tilde{S}^{t+1}\cap S_{G_j}}) = \mathbf{0},\ \mathcal{T}_{\tilde{\lambda}_2,s_0}(\tilde{H}_{G_j}^{t+1}) \neq \mathbf{0}\right) \\
\overset{(i)}{\leq} &\frac{s}{8\Delta^2} + \sum_{j\in G_{FD}^{t+1}} \mathrm{I}\left(\sum_{k\in\tilde{S}^{t+1}\cap S_{G_j}} \Xi_{kj}^2\mathrm{I}(|\Xi_{kj}|\geq\tilde{\lambda}_1) < s_0\tilde{\lambda}_1^2,\ \mathcal{T}_{\tilde{\lambda}_2,s_0}(\tilde{H}_{G_j}^{t+1}) \neq \mathbf{0}\right) \\
\overset{(ii)}{\leq} &\frac{s}{8\Delta^2} + \sum_{j\in G_{FD}^{t+1}} \mathrm{I}\left(s_0\tilde{\lambda}_1^2 \leq 2\Upsilon\left(\tilde{S}^{t+1}\cap S_{G_j}, \tilde{\beta}^t\right)\right) \\
\precsim &\frac{s}{\Delta^2} + \frac{s}{\Delta} = O\left(\frac{s}{\Delta}\right), \quad \text{as } \Delta\to\infty,
\end{aligned}
\tag{82}
$$

其中，不等式 (i) 由引理 5 中第一项证明里的类似矛盾得出，而不等式 (ii) 则由式 (97) 的结果得出。

综合 (80)、(81) 和 (82)，我们证明了 $\|\tilde{\eta}_G^{t+1} - \eta_G^*\|_0 = O\left(\frac{s}{\Delta}\right)$。 –

步骤 3（逐元素几乎完全恢复）。

根据定理 7 的前两个结论，我们有

$$\begin{aligned}\|\tilde{\eta}^{t+1}-\eta^*\|_0 =& \sum_{j=1}^{m}\sum_{i=1}^{d}|\tilde{\eta}_{ij}^{t+1}-\eta_{ij}^*| \\ =& \sum_{(i,j)\in S^*}|\tilde{\eta}_{ij}^{t+1}-1|+\sum_{(i,j)\in S_{G^*}\cap(S^*)^c\cap\tilde{S}^{t+1}}|\tilde{\eta}_{ij}^{t+1}-0|+\sum_{(i,j)\in S_{G^*}^c\cap\tilde{S}^{t+1}}|\tilde{\eta}_{ij}^{t+1}-0|.\end{aligned} \tag{83}$$

We 只需分别分析这三项。对于第一项，注意到 t

$$\begin{aligned}\sum_{(i,j)\in S^*}|\tilde{\eta}_{ij}^{t+1}-1| =& \sum_{(i,j)\in S^*}\mathrm{I}\left((i,j)\notin\tilde{S}^{t+1}\right) \\ \overset{(i)}{\leq}& \sum_{(i,j)\in S^*}\mathrm{I}\left(|\tilde{H}_{ij}^{t+1}|<\tilde{\lambda}_2\right) \\ &+\frac{1}{\tilde{\lambda}_2^2}\sum_{(i,j)\in S^*}\left(\tilde{H}_{ij}^{t+1}\right)^2\mathrm{I}\left(|\tilde{H}_{ij}^{t+1}|\geq\tilde{\lambda}_2,\ \sum_{k\in S_{G_j}\cap S^*}\left(\tilde{H}_{kj}^{t+1}\right)^2<(s_0+s_j)\tilde{\lambda}_2^2\right) \\ \overset{(ii)}{\leq}& \frac{16\delta^2\left\|\tilde{\beta}^t-\beta^*\right\|_2^2}{\epsilon^2\tilde{\lambda}_2^2}+\frac{4\sigma^2 ss_0}{n\tilde{\lambda}_2^2\Delta}\precsim\ \frac{ss_0}{\Delta}+\frac{ss_0}{\Delta^2} \\ =& O\left(\frac{ss_0}{\Delta}\right),\ \text{ as } \Delta\to\infty,\end{aligned} \tag{84}$$

其中，不等式 (i) 由引理 7 中 (104) 的第一个不等式得到，不等式 (ii) 则由 (104) 的最后一个不等式、(107) 和 (113) 得到。

对于第二项，我们得到

$$\begin{aligned}\sum_{(i,j)\in S_{G^*}\cap(S^*)^c\cap\tilde{S}^{t+1}}|\tilde{\eta}_{ij}^{t+1}-0| \leq& \sum_{(i,j)\in S_{G^*}\cap(S^*)^c\cap\tilde{S}^{t+1}}\mathrm{I}\left(\left|\tilde{H}_{ij}^{t+1}\right|\geq\tilde{\lambda}_2\right) \\ \overset{(i)}{\leq}& \sum_{(i,j)\in S_{G^*}}\mathrm{I}\left(|\Xi_{ij}|\geq\tilde{\lambda}_1\right) \\ &+\sum_{(i,j)\in S_{G^*}\cap(S^*)^c\cap\tilde{S}^{t+1}}\mathrm{I}\big(|\Xi_{ij}|<\tilde{\lambda}_1<|\langle\Phi_{ij}^\top,\beta^*-\tilde{\beta}^t\rangle|\big) \\ \overset{(ii)}{\leq}& \frac{\sigma^2 ss_0}{n\tilde{\lambda}_1^2\Delta}+\frac{\delta^2\left\|\tilde{\beta}^t-\beta^*\right\|_2^2}{\tilde{\lambda}_1^2}\precsim\ \frac{ss_0}{\Delta^2(s,s_0)}+\frac{ss_0}{\Delta} \\ =& O\left(\frac{ss_0}{\Delta}\right),\ \text{ as } \Delta\to\infty,\end{aligned} \tag{85}$$

其中不等式 (i) 根据定理 7 中的 (71) 得出，不等式 (ii) 根据概率-

引理 6 中的不等式（102）和 $\sum_{(i,j)\in S_{G^*}} \tilde{\lambda}_1^2 \mathrm{I}\Big\{|\Xi_{ij}| \geq \tilde{\lambda}_1\Big\} \leq \sum_{(i,j)\in S_{G^*}} \Xi_{ij}^2 \mathrm{I}\Big\{|\Xi_{ij}| \geq \tilde{\lambda}_1\Big\}$。

对于第三项，我们得到

$$
\begin{aligned}
&\sum_{(i,j)\in S_{G^*}^c \cap \tilde{S}^{t+1}} |\tilde{\eta}_{ij}^{t+1} - 0| = \sum_{(i,j)\in S_{G^*}^c \cap \tilde{S}^{t+1}} \mathrm{I}\left(\mathcal{T}_{\tilde{\lambda}_2, s_0}\big(\tilde{H}^{t+1}\big)_{ij} \neq 0\right) \\
\overset{(i)}{\leq} &\sum_{(i,j)\in S_{G^*}^c \cap \tilde{S}^{t+1}} \mathrm{I}\Big\{\mathcal{T}_{\tilde{\lambda}_1, s_0}\big(\Xi_{S_{OG}}\big)_{ij} \neq 0\Big\} + \sum_{(i,j)\in S_{G^*}^c \cap \tilde{S}^{t+1}} \mathrm{I}\Big\{|\Xi_{ij}| < \tilde{\lambda}_1,\ |\Xi_{ij} + \langle \Phi_{ij}^\top, \beta^* - \tilde{\beta}^t\rangle| \geq \tilde{\lambda}_2\Big\} \\
&+ \frac{1}{\tilde{\lambda}_1^2} \sum_{(i,j)\in S_{G^*}^c \cap \tilde{S}^{t+1}} \Xi_{ij}^2 \mathrm{I}\left\{|\Xi_{ij}| \geq \tilde{\lambda}_1,\ \sum_{k\in S_{G_j}\cap \tilde{S}^{t+1}} \Xi_{kj}^2 \mathrm{I}(|\Xi_{kj}| \geq \tilde{\lambda}_1) < s_0 \tilde{\lambda}_1^2,\ \mathcal{T}_{\tilde{\lambda}_2, s_0}\big(\tilde{H}^{t+1}\big)_{ij} \neq 0\right\} \\
\overset{(ii)}{<} &s' s_0 + \frac{\delta^2 \left\|\tilde{\beta}^t - \beta^*\right\|_2^2}{\tilde{\lambda}_1^2} + \frac{2\delta^2 \left\|\tilde{\beta}^t - \beta^*\right\|_2^2}{\tilde{\lambda}_1^2} \\
\precsim &\frac{s s_0}{\Delta^2} + \frac{s s_0}{\Delta} = O\left(\frac{s s_0}{\Delta}\right), \quad \text{as } \Delta \to \infty,
\end{aligned}
\tag{86}
$$

其中不等式 (i) 由引理 5 中的 (92) 得出，而不等式 (ii) 由引理 5 中第一项的框架、(97) 和 (98) 得出，并回顾 $s' = \frac{s}{8\Delta^2}$。

结合（83）、（84）、（85）和（86），我们证明了 $\|\tilde{\eta}^{t+1} - \eta^*\|_0 = O\left(\frac{s s_0}{\Delta}\right)$。

附录 B：辅助引理

引理 3 假设 X 满足 *DSRIP*$(2s, \frac{3}{2}s_0, \frac{\delta}{2})$。则，以至少 $1 - \exp\{-Css_0\Delta\}$ 的概率，有

$$
|\sigma_t - \sigma| \leq \sqrt{1+\delta}\|\beta^* - \beta^t\|_2 + \frac{1}{20}\sigma.
$$

证明 1 记事件为 $\mathcal{A} = \{|\frac{\|\xi\|_2}{\sigma} - \sqrt{n}| \leq \frac{1}{20}\sqrt{n}\}$。由 *Hanson-Wright* 不等式[*Rudelson* 和 *Vershynin*, *2013*], 可得 $P(\mathcal{A}) \geq 1 - e^{-Cn} \geq 1 - e^{-Css_0\Delta}$。因此，

$$
\begin{aligned}
|\sigma_t - \sigma| &\leq |\sigma_t - \frac{\|\xi\|_2}{\sqrt{n}}| + |\frac{\|\xi\|_2}{\sqrt{n}} - \sigma| \\
&\leq \frac{1}{\sqrt{n}}\left|\|X(\beta^* - \beta^t) + \xi\|_2 - \|\xi\|_2\right| + \frac{1}{20}\sigma \\
&\leq \sqrt{1+\delta}\|\beta^* - \beta^t\|_2 + \frac{1}{20}\sigma,
\end{aligned}
$$

其中，第二个不等式由事件 $\mathcal{A}$ 得出，最后一个不等式则由 *DSRIP*

条件得出。

为了控制 ξ 和 $X\left(\hat{\beta}-\beta^{*}\right)$ 之间的内积，我们给出一个有用的引理。

引理 4 给定整数 v_1、$v_2>0$，并假设 β 是一个$(v_1, v_2/v_1)$-稀疏向量。我们有

$$P\left(\sup_{\beta\in\Theta^{m,d}(v_1,\frac{v_2}{v_1})}\left|\left\langle\xi,\frac{X\beta}{\|X\beta\|_2}\right\rangle\right|^2\geq 3\sigma^2(v_1\log\frac{em}{v_1}+v_2\log\frac{edv_1}{v_2})\right)\leq e^{-C\left(v_1\log\frac{em}{v_1}+v_2\log\frac{edv_1}{v_2}\right)}. \tag{87}$$

具体而言，如果 β^{*} 是一个（s，s_0）-稀疏向量，并且 $\hat{}\,\beta\in\Theta^{m,\,d}$（$\hat{}s$，$\hat{}_A\hat{s}$），即 $\|\hat{}\beta\|_0\leq \hat{}A$ 和 $\|\hat{}\beta\|_{0,\,2}\leq\hat{}s$，则有

$$P\left(\sup_{\hat{\beta}\in\Theta^{m,d}(\hat{s},\frac{\hat{A}}{\hat{s}})}\left|\left\langle\xi,\frac{X(\hat{\beta}-\beta^{*})}{\|X(\hat{\beta}-\beta^{*})\|_2}\right\rangle\right|^2\geq 3\sigma^2\Omega^{*}(\hat{\beta})\right)\leq e^{-C\Omega^{*}(\hat{\beta})}, \tag{88}$$

$\hat{}$其中 $\Omega^{*}(\beta)$ 定义于附录开头。

证明 2 对于满足 $supp(\beta)\subseteq S$ 的固定集合 S，将 X_S 记为由 X 中索引为 S 的列所张成的空间，因此我们有 $\langle\xi、X\beta\rangle=\langle\xi、X_S\beta_S\rangle$。记 $\pi_S=X_S(X_S^{\top}X_S)^{-1}X_S^{\top}\in\mathbb{R}^{n\times n}$，它是一个秩不超过 $|S|$ 的正交矩阵。因此，对于 $\forall\beta\in\Theta^{m,d}(v_1,\frac{v_2}{v_1})$，由柯西-施瓦茨不等式可得：

$$\left|\left\langle\xi,\frac{X\beta}{\|X\beta\|_2}\right\rangle\right|=\left|\left\langle\pi_S\xi,\frac{X_S\beta_S}{\|X_S\beta_S\|_2}\right\rangle\right|\leq\|\pi_S\xi\|_2\leq\sup_{S\in\mathcal{S}^{m,d}\left(v_1,\frac{v_2}{v_1}\right)}\|\pi_S\xi\|_2\,, \tag{89}$$

注意到，对于 $\forall S\in\mathcal{S}^{m,\,d}\left(v_1,\frac{v_2}{v_1}\right)$，我们有 $rank(\pi_S)\leq v_2$，因此 $Tr(\pi_S)\leq rank(\pi_S)\cdot\|\pi_S\|_2\leq v_2$。于是，根据 *Hsu* 等人的定理 *2.1*，*[2012]*, 对于 $\forall t>0$，我们有

$$P\left(\frac{\|\pi_S\xi\|_2^2}{\sigma^2}\geq\frac{5}{2}t\right)\leq P\left(\frac{\|\pi_S\xi\|_2^2}{\sigma^2}\geq v_2+2\sqrt{v_2t}+2t\right)\leq e^{-t}, \tag{90}$$

其中，当 $t\gg v_2$ 时，第一个不等式成立。

与（26）类似，我们有 $\left|S^{m,\,d}\left(v_1,\frac{v_2}{v_1}\right)\right|\leq\left(\frac{em}{v_1}\right)^{v_1}\times\left(\frac{edv_1}{v_2}\right)^{v_2}$，因此由（90）可得并集界：

$$P\left(\sup_{S\in\mathcal{S}^{m,d}\left(v_1,\frac{v_2}{v_1}\right)}\|\pi_S\xi\|_2^2\geq 3\sigma^2\left(v_1\log\frac{em}{v_1}+v_2\log\frac{edv_1}{v_2}\right)\right)\leq e^{-C\left(v_1\log\frac{em}{v_1}+v_2\log\frac{edv_1}{v_2}\right)}.$$

令 $t=(1+C)(v_1\log\frac{em}{v_2}+v_1+\log\frac{edv_1}{v_2})$ 为某个常数 $0<C<\frac{1}{5}$，且满足$t\gg v_2$。我们完成了(87)。

对于(88)，任取$\hat{}\ \beta \in \Theta^{m,d}(\hat{}s\text{、}-\hat{}_A\,\hat{s}\,)$，结合 $\beta^* \in \Theta^{m,d}(s, s_0)$，因此有：

$$\hat{\beta} - \beta^* \in \Theta^{m,d}\left(\hat{s} + s, \frac{ss_0 + \hat{A}}{\hat{s} + s}\right).$$

$\hat{}$令 $v_1 = \hat{}s + s$ 和 $v_2 = ss_0 + A$，由此可直接通过 (87) 得到 (88)。

为便于展示，在接下来的三个引理中，我们用双重指标（i, j）表示第 j 组 G_j 中第 i 个条目（变量）。此外，回顾缩写 $\Upsilon(A, \beta^t) = \sum_{(i,j)\in A}\langle \Phi_{ij}^\top, \beta^* - \tilde{\beta}^t\rangle^2$, $\Delta = \frac{1}{s_0}\log(em/s) + \log(ed/s_0)$ 和 $\tilde{\lambda}_a = a\sqrt{\frac{8\sigma^2}{n}\left(\log\frac{ed}{s_0} + \frac{1}{s_0}\log\frac{em}{s}\right)}$。记 $S_{OG} := \tilde{S}^{t+1} \cap S_{G^*}^c$。

首先，为了界定 S_{OG} 中 Ξ 所选条目的 ℓ_2 范数，我们给出以下引理。

引理 5 假设定理 7 中的所有条件均成立。对于 $\forall t \geq 0$，当 $\Delta, \frac{ss_0}{\Delta} \to \infty$ 时，我们有

$$P\left(\sqrt{\sum_{(i,j)\in S_{OG}} \Xi_i^2 \mathrm{I}\left\{\mathcal{T}_{\tilde{\lambda}_2, s_0}\left(\tilde{H}^{t+1}\right)_{ij} \neq 0\right\}} < \sqrt{\frac{\sigma^2 ss_0}{n\Delta}} + \frac{5}{2}\delta\|\tilde{\beta}^t - \beta^*\|_2\right) \to 1. \tag{91}$$

证明 3 注意到

$$
\begin{aligned}
&\sqrt{\sum_{(i,j)\in S_{OG}} \Xi_{ij}^2 \mathrm{I}\Big\{\mathcal{T}_{\tilde{\lambda}_2,s_0}\big(\tilde{H}^{t+1}\big)_{ij}\neq 0\Big\}}\\
\leq&\sqrt{\sum_{(i,j)\in S_{OG}} \Xi_{ij}^2 \mathrm{I}\Big\{\mathcal{T}_{\tilde{\lambda}_1,s_0}\big(\Xi_{S_{OG}}\big)_{ij}\neq 0,\ \mathcal{T}_{\tilde{\lambda}_2,s_0}\big(\tilde{H}^{t+1}\big)_{ij}\neq 0\Big\}}\\
&+\sqrt{\sum_{(i,j)\in S_{OG}} \Xi_{ij}^2 \mathrm{I}\Big\{\mathcal{T}_{\tilde{\lambda}_1,s_0}\big(\Xi_{S_{OG}}\big)_{ij}= 0,\ \mathcal{T}_{\tilde{\lambda}_2,s_0}\big(\tilde{H}^{t+1}\big)_{ij}\neq 0\Big\}}\\
\leq&\sqrt{\sum_{(i,j)\in S_{OG}} \Xi_{ij}^2 \mathrm{I}\Big\{\mathcal{T}_{\tilde{\lambda}_1,s_0}\big(\Xi_{S_{OG}}\big)_{ij}\neq 0,\ \mathcal{T}_{\tilde{\lambda}_2,s_0}\big(\tilde{H}^{t+1}\big)_{ij}\neq 0\Big\}}\\
&+\sqrt{\sum_{(i,j)\in S_{OG}} \Xi_{ij}^2 \mathrm{I}\Big\{|\Xi_{ij}|<\tilde{\lambda}_1,\ \mathcal{T}_{\tilde{\lambda}_2,s_0}\big(\tilde{H}^{t+1}\big)_{ij}\neq 0\Big\}}\\
&+\sqrt{\sum_{(i,j)\in S_{OG}} \Xi_{ij}^2 \mathrm{I}\left\{|\Xi_{ij}|\geq\tilde{\lambda}_1,\ \sum_{k\in S_{G_j}\cap S_{OG}} \Xi_{kj}^2\mathrm{I}(|\Xi_{kj}|\geq\tilde{\lambda}_1)<s_0\tilde{\lambda}_1^2,\ \mathcal{T}_{\tilde{\lambda}_2,s_0}\big(\tilde{H}^{t+1}\big)_{ij}\neq 0\right\}}\\
\leq&\sqrt{\sum_{(i,j)\in S_{OG}} \Xi_{ij}^2 \mathrm{I}\Big\{\mathcal{T}_{\tilde{\lambda}_1,s_0}\big(\Xi_{S_{OG}}\big)_{ij}\neq 0\Big\}}+\sqrt{\sum_{(i,j)\in S_{OG}} \Xi_{ij}^2 \mathrm{I}\Big\{|\Xi_{ij}|<\tilde{\lambda}_1,\ |\Xi_{ij}+\langle\Phi_{ij}^\top,\beta^*-\tilde{\beta}^t\rangle|\geq\tilde{\lambda}_2\Big\}}\\
&+\sqrt{\sum_{j\in\tilde{G}^{t+1}\cap(G^*)^c} s_0\tilde{\lambda}_1^2\cdot \mathrm{I}\left\{\sum_{k\in S_{G_j}\cap S_{OG}} \Xi_{kj}^2\mathrm{I}(|\Xi_{kj}|\geq\tilde{\lambda}_1)<s_0\tilde{\lambda}_1^2,\ \mathcal{T}_{\tilde{\lambda}_2,s_0}\big(\tilde{H}^{t+1}\big)_{G_j}\neq 0\right\}}.
\end{aligned}
\tag{92}
$$

接下来，我们分别对最后一个不等式中的三项进行界定。

第一项。令 $s'=\frac{s}{8\Delta^2}$，~~s'_0~~ $=s_0$。接下来，我们证明，在引理 *1* 中的事件 $\mathcal{E}(s',s_0)$ 发生时，$\Xi_{S_{OG}}$ 中被 $\mathcal{T}_{\tilde{\lambda}_1,s_0}$发现的组少于 s' 个。否则，任取 s' 个已发现的组，并构造一个 $S'\subset S_{OG}$ 和 $S'\in\mathcal{S}^{m,d}(s',s_0)$，使其满足

$$
\sum_{(i,j)\in S'}\Xi_{ij}^2\geq\sum_{(i,j)\in S'}\Xi_{ij}^2\mathrm{I}\Big\{\mathcal{T}_{\tilde{\lambda}_1,s_0}\big(\Xi_{S_{OG}}\big)_{ij}\neq 0\Big\}\geq\ s's_0\tilde{\lambda}_1^2\geq\ \frac{s}{8\Delta^2}\cdot s_0\cdot\frac{8\sigma^2}{n}\Delta=\frac{ss_0\sigma^2}{n\Delta}.
$$

When 当 Δ 足够大时，我们可以证明 $\log(8\Delta^2)\ <\ \frac{s_0}{3}\Delta$，由此可得

$$
\Delta<\Delta(s',s_0):=\frac{1}{s_0}\log\frac{em}{s'}+\log\frac{ed}{s_0}<\frac{4}{3}\Delta.
$$

因此我们有

$$
\sum_{(i,j)\in S'}\Xi_{ij}^2\geq\frac{ss_0\sigma^2}{n\Delta}=\frac{8s's_0\sigma^2\Delta}{n}>\frac{8s's_0\sigma^2}{n}\cdot\frac{3}{4}\Delta(s',s_0)=\frac{6s's_0\sigma^2\Delta(s',s_0)}{n},
$$

这与引理 *1* 中的事件 $\mathcal{E}(s', s_0)$ 以高概率成立相矛盾。因此，我们证明在 $\Xi_{S_{OG}}$ 中被发现的组少于 s' 个。同理，我们可以证明在 $\Xi_{S_{OG}}$ 中被发现的条目少于 $s's_0$ 个。若非如此，取 $S_2 \in \mathcal{S}^{m,d}(s', s_0)$ 和 $S_2 \subset S_{OG}$，它们的条目在 S_{OG}中均被错误地发现，这将导致

$$\sum_{(i,j)\in S_2} \Xi_{ij}^2 \geq s's_0\tilde{\lambda}_1^2 \geq \frac{ss_0\sigma^2}{n\Delta} \geq \frac{6s's_0\sigma^2\Delta(s', s_0)}{n}. \tag{93}$$

在引理 *1* 中的事件 $\mathcal{E}(s', s_0)$ 下，(93) 再次导致矛盾。因此，我们可以将(92)中的第一项上界为

$$\sum_{(i,j)\in S_{OG}} \Xi_{ij}^2 \mathrm{I}\Big\{\mathcal{T}_{\tilde{\lambda}_1,s_0}\big(\Xi_{S_{OG}}\big)_{ij} \neq 0\Big\} \leq \sup_{S_2\in\mathcal{S}(s',s_0)} \sum_{(i,j)\in S_2} \Xi_{ij}^2 \leq \frac{6\sigma^2 s's_0\Delta(s', s_0)}{n} \leq \frac{\sigma^2 ss_0}{n\Delta}. \tag{94}$$

第二项。注意

$$\begin{aligned}
&\sum_{(i,j)\in S_{OG}} \Xi_{ij}^2 \mathrm{I}\Big\{|\Xi_{ij}| < \tilde{\lambda}_1,\ |\Xi_{ij} + \langle\Phi_{ij}^\top, \beta^* - \tilde{\beta}^t\rangle| \geq \tilde{\lambda}_2\Big\} \\
\leq &\sum_{(i,j)\in S_{OG}} \Xi_{ij}^2 \mathrm{I}\Big\{|\Xi_{ij}| < \tilde{\lambda}_1,\ |\Xi_{ij}| + |\langle\Phi_{ij}^\top, \beta^* - \tilde{\beta}^t\rangle| \geq 2\tilde{\lambda}_1\Big\} \\
\leq &\sum_{(i,j)\in S_{OG}} \Xi_{ij}^2 \mathrm{I}\Big\{|\Xi_{ij}| < \tilde{\lambda}_1 \leq |\langle\Phi_{ij}^\top, \beta^* - \tilde{\beta}^t\rangle|\Big\} \\
\leq &\sum_{(i,j)\in S_{OG}} \langle\Phi_{(i,j)}^\top, \beta^* - \tilde{\beta}^t\rangle^2 = \left\|\Phi_{S_{OG}}\left(\beta^* - \tilde{\beta}^t\right)\right\|_2^2.
\end{aligned} \tag{95}$$

因此，我们可以将(92) 中的第二项界定为

$$\sqrt{\sum_{(i,j)\in S_{OG}} \Xi_{ij}^2 \mathrm{I}\Big\{|\Xi_{ij}| < \tilde{\lambda}_1,\ |\Xi_{ij} + \langle\Phi_{ij}^\top, \beta^* - \tilde{\beta}^t\rangle| \geq \tilde{\lambda}_2\Big\}} \leq \left\|\Phi_{S_{OG}}\left(\beta^* - \tilde{\beta}^t\right)\right\|_2 \leq \delta\left\|\tilde{\beta}^t - \beta^*\right\|_2. \tag{96}$$

第三项。对于任意组 $j \notin G^*$，满足 $\sum_{k\in S_{G_j}\cap S_{OG}} \Xi_{kj}^2 \mathrm{I}(|\Xi_{kj}| \geq \tilde{\lambda}_1) < s_0\tilde{\lambda}_1^2$ 且

$\mathcal{T}_{\tilde{\lambda}_2,\ s_0}\big(\tilde{H}^{t+1}\big)_{G_j} \neq \mathbf{0}$ （其中索引遍历 S_{OG}），我们有

$$
\begin{aligned}
s_0\tilde{\lambda}_2^2 \leq & \sum_{k\in S_{G_j}\cap S_{OG}} \Big(\underbrace{\Xi_{kj}+\langle\Phi_{kj}^\top,\beta^*-\tilde{\beta}^t\rangle}_{\tilde{H}_{kj}^{t+1}}\Big)^2 \mathrm{I}\Big(|\Xi_{kj}+\langle\Phi_{kj}^\top,\beta^*-\tilde{\beta}^t\rangle|\geq\tilde{\lambda}_2\Big) \\
\leq & \sum_{k\in S_{G_j}\cap S_{OG}} 2\Xi_{kj}^2\mathrm{I}\Big(|\tilde{H}_{kj}^{t+1}|\geq\tilde{\lambda}_2\Big)+\sum_{k\in S_{G_j}\cap S_{OG}} 2\langle\Phi_{kj}^\top,\beta^*-\tilde{\beta}^t\rangle^2\mathrm{I}\Big(|\tilde{H}_{kj}^{t+1}|\geq\tilde{\lambda}_2\Big) \\
\leq & \sum_{k\in S_{G_j}\cap S_{OG}} 2\Xi_{kj}^2\mathrm{I}\Big(|\Xi_{kj}|\geq\tilde{\lambda}_1\Big)+\sum_{k\in S_{G_j}\cap S_{OG}} 2\Xi_{kj}^2\mathrm{I}\Big(|\Xi_{kj}|<\tilde{\lambda}_1\leq|\langle\Phi_{kj}^\top,\beta^*-\tilde{\beta}^t\rangle|\Big) \\
& +2\sum_{k\in S_{G_j}\cap S_{OG}}\langle\Phi_{kj}^\top,\beta^*-\tilde{\beta}^t\rangle^2 \\
\leq & 2s_0\tilde{\lambda}_1^2+4\Upsilon\left(S_{G_j}\cap S_{OG},\tilde{\beta}^t\right),
\end{aligned}
\tag{97}
$$

由此得到 $s_0\tilde{\lambda}_1^2 \leq 2\Upsilon\left(S_{G_j}\cap S_{OG},\tilde{\beta}^t\right)$。因此，我们可以将(92) 中的第三项界定为

$$
\begin{aligned}
& \sqrt{\sum_{(i,j)\in S_{OG}}\Xi_{ij}^2\mathrm{I}\Big\{|\Xi_{ij}|\geq\tilde{\lambda}_1,\ \sum_{k\in S_{G_j}\cap S_{OG}}\Xi_{kj}^2\mathrm{I}(|\Xi_{kj}|\geq\tilde{\lambda}_1)<s_0\tilde{\lambda}_1^2,\ \mathcal{T}_{\tilde{\lambda}_2,s_0}\big(\tilde{H}^{t+1}\big)_{ij}\neq 0\Big\}} \\
\leq & \sqrt{\sum_{j\in\tilde{G}^{t+1}\cap(G^*)^c} s_0\tilde{\lambda}_1^2\mathrm{I}\Big\{s_0\tilde{\lambda}_1^2\leq 2\Upsilon\left(S_{G_j}\cap S_{OG},\tilde{\beta}^t\right)\Big\}} \\
\leq & \sqrt{2\Upsilon\left(S_{OG},\tilde{\beta}^t\right)} \\
< & \frac{3}{2}\delta\left\|\tilde{\beta}^t-\beta^*\right\|_2.
\end{aligned}
\tag{98}
$$

*Co*将这三项(94)、(96)和(98)合并起来，最终得到

$$
P\left(\sqrt{\sum_{(i,j)\in S_{OG}}\Xi_i^2\mathrm{I}\Big\{\mathcal{T}_{\tilde{\lambda}_2,s_0}\big(\tilde{H}^{t+1}\big)_{ij}\neq 0\Big\}}<\sqrt{\frac{\sigma^2 s s_0}{n\Delta}}+\frac{5}{2}\delta\|\tilde{\beta}^t-\beta^*\|_2\right)\to 1, \tag{99}
$$

如 Δ，$\frac{ss_0}{\Delta}\to\infty$。

类似地，我们可以界定真实组G^*中 Ξ 所选条目的 ℓ_2-范数，这可表述为以下引理。

引理 6 假设定理 *7* 中的所有条件均成立。当 $\Delta\to\infty$ 时，我们有

$$
P\left(\sqrt{\sum_{(i,j)\in S_{G^*}}\Xi_{ij}^2\mathrm{I}\Big\{|\Xi_{ij}|\geq\tilde{\lambda}_1\Big\}}<\sqrt{\frac{\sigma^2 s s_0}{n\Delta}}\right)\to 1. \tag{100}
$$

证明 4 由于对于任意 $(i,j) \in S_{G^*}$, Ξ_{ij} 都是参数为 $\frac{\sigma^2}{n}$ 的次高斯变量，因此我们得出结论：

$$\begin{aligned}
\mathbf{E}\Big(\Xi_{ij}^2 \mathrm{I}\left\{|\Xi_{ij}| \geq \tilde{\lambda}_1\right\}\Big) &= \int_0^\infty P\Big(\Xi_{ij}^2 \mathrm{I}\left\{|\Xi_{ij}| \geq \tilde{\lambda}_1\right\} > u\Big) \mathrm{d}u \\
&= \int_0^{\tilde{\lambda}_1^2} P\Big(|\Xi_{ij}| \geq \tilde{\lambda}_1\Big) \mathrm{d}u + \int_{\tilde{\lambda}_1^2}^\infty P\Big(|\Xi_{ij}| \geq \sqrt{u}\Big) \mathrm{d}u \\
&\leq 2\tilde{\lambda}_1^2 \exp\left(-\frac{n\tilde{\lambda}_1^2}{2\sigma^2}\right) + \int_{\tilde{\lambda}_1^2}^\infty 2\exp\left(-\frac{nu}{2\sigma^2}\right) \mathrm{d}u \\
&= \left(2\tilde{\lambda}_1^2 + \frac{4\sigma^2}{n}\right) \exp\left(-\frac{n\tilde{\lambda}_1^2}{2\sigma^2}\right) \\
&\leq 3\tilde{\lambda}_1^2 \exp\left(-\frac{n\tilde{\lambda}_1^2}{2\sigma^2}\right),
\end{aligned} \tag{101}$$

其中，最后一个不等式由 $\tilde{\lambda}_1^2 = \frac{8\sigma^2}{n}\Delta \geq \frac{4\sigma^2}{n}$ 得出。因此，根据马尔可夫不等式，我们有

$$\begin{aligned}
P\left(\sum_{(i,j)\in S_{G^*}} \Xi_{ij}^2 \mathrm{I}\Big\{|\Xi_{ij}| \geq \tilde{\lambda}_1\Big\} \geq \frac{\sigma^2 s s_0}{n\Delta}\right) &\leq \frac{n\Delta}{\sigma^2 s s_0} \cdot \mathbf{E}\left(\sum_{(i,j)\in S_{G^*}} \Xi_{ij}^2 \mathrm{I}\Big\{|\Xi_{ij}| \geq \tilde{\lambda}_1\Big\}\right) \\
&\leq 3\Delta \cdot \frac{\tilde{\lambda}_1^2 n}{\sigma^2} \cdot \frac{d}{s_0} \cdot \exp\left(-\frac{n\tilde{\lambda}_1^2}{2\sigma^2}\right) \\
&\leq \frac{3}{4}\left(\frac{n\tilde{\lambda}_1^2}{\sigma^2}\right)^2 \exp\left(-\frac{n\tilde{\lambda}_1^2}{4\sigma^2}\right) \\
&= o(1), \quad as\ \Delta \to \infty,
\end{aligned} \tag{102}$$

其中，最后一个不等式利用了 $\tilde{\lambda}_1^2 \geq \frac{4\sigma^2\Delta}{n}$ 和 $\frac{d}{s_0} < \exp(\Delta) \leq \exp\left(\frac{n\tilde{\lambda}_1^2}{4\sigma^2}\right)$。

现在我们来分析 S^* 上的估计误差项。在满足适当的 beta-min 条件时，我们可以通过以下引理对 $\sum_{(i,j)\in S^*} \left(\tilde{H}_{ij}^{t+1}\right)^2 \mathrm{I}\left((i,j) \notin \tilde{S}^{t+1}\right)$ 进行界定。

引理 7 假设定理 7 中的所有条件均成立。那么，对于任意 $\epsilon > 0$，我们有

$$P\left(\sqrt{\sum_{(i,j)\in S^*} \left(\tilde{H}_{ij}^{t+1}\right)^2 \mathrm{I}\left((i,j) \notin \tilde{S}^{t+1}\right)} < \frac{4}{\epsilon}\delta \left\|\tilde{\beta}^t - \beta^*\right\|_2 + 2\sqrt{\frac{\sigma^2 s s_0}{n\Delta}}\right) \to 1, \tag{103}$$

作为 $\Delta \to \infty$。

证明 5 注意

$$
\begin{aligned}
&\sqrt{\sum_{(i,j)\in S^*}\left(\tilde{H}_{ij}^{t+1}\right)^2 \mathrm{I}\left((i,j)\notin \tilde{S}^{t+1}\right)}\\
\leq&\sqrt{\sum_{(i,j)\in S^*}\left(\tilde{H}_{ij}^{t+1}\right)^2 \mathrm{I}\left(|\tilde{H}_{ij}^{t+1}|<\tilde{\lambda}_2\right)}\\
&+\sqrt{\sum_{(i,j)\in S^*}\left(\tilde{H}_{ij}^{t+1}\right)^2 \mathrm{I}\left(|\tilde{H}_{ij}^{t+1}|\geq\tilde{\lambda}_2,\ \sum_{k\in S^*\cap S_{G_j}}\left(\tilde{H}_{kj}^{t+1}\right)^2 \mathrm{I}(|\tilde{H}_{kj}^{t+1}|\geq\tilde{\lambda}_2)<s_0\tilde{\lambda}_2^2\right)}\\
<&\sqrt{\sum_{(i,j)\in S^*}\tilde{\lambda}_2^2\cdot \mathrm{I}\left(|\tilde{H}_{ij}^{t+1}|<\tilde{\lambda}_2\right)}\\
&+\sqrt{\sum_{j\in G^*}s_0\tilde{\lambda}_2^2\cdot \mathrm{I}\left(\sum_{k\in S^*\cap S_{G_j}}\left(\tilde{H}_{kj}^{t+1}\right)^2 \mathrm{I}(|\tilde{H}_{kj}^{t+1}|\geq\tilde{\lambda}_2)<s_0\tilde{\lambda}_2^2\right)}\\
\leq&\sqrt{\sum_{(i,j)\in S^*}\tilde{\lambda}_2^2\cdot \mathrm{I}\left(|\tilde{H}_{ij}^{t+1}|<\tilde{\lambda}_2\right)}+\sqrt{\sum_{j\in G^*}s_0\tilde{\lambda}_2^2\cdot \mathrm{I}\left(\sum_{k\in S^*\cap S_{G_j}}\left(\tilde{H}_{kj}^{t+1}\right)^2<(s_j+s_0)\tilde{\lambda}_2^2\right)}.
\end{aligned}
\tag{104}
$$

接下来，我们分别分析这两项。

第一项。回顾可知，对于每个支撑集元素， $\tilde{H}_{ij}^{t+1}=\beta_{ij}^*+\langle\Phi_{ij}^\top、\ \beta^*-\tilde{\beta}^t\rangle+\Xi_{ij}$ 和 $|\beta_{ij}^*|\geq(1+\epsilon)\tilde{\lambda}_2$ 均成立。因此，我们有

$$
\begin{aligned}
\sqrt{\sum_{(i,j)\in S^*}\tilde{\lambda}_2^2\cdot \mathrm{I}\left(|\tilde{H}_{ij}^{t+1}|<\tilde{\lambda}_2\right)}\leq&\sqrt{\sum_{(i,j)\in S^*}\tilde{\lambda}_2^2\cdot \mathrm{I}\Big(|\beta_{ij}|-|\langle\Phi_{ij}^\top,\beta^*-\tilde{\beta}^t\rangle+\Xi_{ij}|<\tilde{\lambda}_2\Big)}\\
\leq&\sqrt{\sum_{(i,j)\in S^*}\tilde{\lambda}_2^2\cdot \mathrm{I}\Big(\epsilon\tilde{\lambda}_2<|\langle\Phi_{ij}^\top,\beta^*-\tilde{\beta}^t\rangle|+|\Xi_{ij}|\Big)}\\
\leq&\sqrt{\sum_{(i,j)\in S^*}\tilde{\lambda}_2^2\cdot \mathrm{I}\Big(|\langle\Phi_{ij}^\top,\beta^*-\tilde{\beta}^t\rangle|>\frac{\epsilon}{2}\tilde{\lambda}_2\Big)}\\
&+\sqrt{\sum_{(i,j)\in S^*}\tilde{\lambda}_2^2\cdot \mathrm{I}\Big(|\Xi_{ij}|>\frac{\epsilon}{2}\tilde{\lambda}_2\Big)}\\
<&\frac{2}{\epsilon}\sqrt{\Upsilon\left(S^*,\tilde{\beta}^t\right)}+\sqrt{\sum_{(i,j)\in S^*}\tilde{\lambda}_2^2\cdot \mathrm{I}\Big(|\Xi_{ij}|>\frac{\epsilon}{2}\tilde{\lambda}_2\Big)}.
\end{aligned}
\tag{105}
$$

在固定的 S^* 下，根据马尔可夫不等式，我们有

$$\begin{aligned} P\left(\sum_{(i,j)\in S^*}\tilde{\lambda}_2^2\cdot \mathrm{I}\Big(|\Xi_{ij}|>\frac{\epsilon}{2}\tilde{\lambda}_2\Big)\geq\frac{\sigma^2 s s_0}{n\Delta}\right) \leq & \frac{n\Delta}{\sigma^2 s s_0}\sum_{(i,j)\in S^*}\tilde{\lambda}_2^2\cdot P\Big(|\Xi_{ij}|>\frac{\epsilon}{2}\tilde{\lambda}_2\Big) \\ \leq & \frac{1}{16}(\frac{n\tilde{\lambda}_2^2}{\sigma^2})^2\exp(-\frac{\epsilon^2}{8}\cdot\frac{n\tilde{\lambda}_2^2}{\sigma^2}) \\ = & o(1),\quad as\ \Delta\to\infty, \end{aligned} \tag{106}$$

其中回顾可知 $\tilde{\lambda}_2=2\sqrt{\frac{8\sigma^2}{n}\Delta}$。因此，(104) 中的第一项可界定为

$$\sqrt{\sum_{(i,j)\in S^*}\tilde{\lambda}_2^2\cdot \mathrm{I}\left(|\tilde{H}_{ij}^{t+1}|<\tilde{\lambda}_2\right)}<\frac{2}{\epsilon}\delta\left\|\tilde{\beta}^t-\beta^*\right\|_2+\sqrt{\frac{\sigma^2 s s_0}{n\Delta}}. \tag{107}$$

第二项。令 $s_j=\|\beta^*_{G_j}\|_0$，其中 $j\in G^*$。对于 $\forall j\in G^*$，根据逐元素 ***beta-min*** 条件 $\min_{(i,j)\in S^*}|\beta^*_{ij}|\geq(\sqrt{2}+\epsilon)\tilde{\lambda}_2$ 和组级 ***beta-min*** 条件 $\min_{j\in G^*}\|\beta^*_{G_j}\|_2\geq(\sqrt{2}+\epsilon)\sqrt{s_0}\tilde{\lambda}_2$，我们得出

$$\|\beta^*_{S_{G_j}}\|_2\geq(\sqrt{2}+\epsilon)\sqrt{s_0\vee s_j}\tilde{\lambda}_2\geq\sqrt{s_j+s_0}\tilde{\lambda}_2+\epsilon\sqrt{s_j\vee s_0}\tilde{\lambda}_2.$$

因此，我们有

$$\begin{aligned} & \mathrm{I}\left(\sum_{k\in S^*\cap S_{G_j}}\left(\tilde{H}_{kj}^{t+1}\right)^2<(s_j+s_0)\tilde{\lambda}_2^2\right) \\ \leq & \mathrm{I}\left(\sqrt{\sum_{k\in S^*\cap S_{G_j}}\left(\beta^*_{kj}\right)^2}-\sqrt{\sum_{k\in S^*\cap S_{G_j}}\left(\langle\Phi_{ij}^\top,\beta^*-\tilde{\beta}^t\rangle+\Xi_{kj}\right)^2}<\sqrt{s_j+s_0}\tilde{\lambda}_2\right) \\ \leq & \mathrm{I}\left(\sum_{k\in S^*\cap S_{G_j}}\left(\langle\Phi_{ij}^\top,\beta^*-\tilde{\beta}^t\rangle+\Xi_{kj}\right)^2>\epsilon^2(s_j\vee s_0)\tilde{\lambda}_2^2\right) \\ \leq & \mathrm{I}\left(\Upsilon\left(S^*\cap S_{G_j},\tilde{\beta}^t\right)+\sum_{k\in S^*\cap S_{G_j}}\Xi_{kj}^2>\frac{\epsilon^2}{2}(s_j\vee s_0)\tilde{\lambda}_2^2\right) \\ \leq & \mathrm{I}\left(\Upsilon\left(S^*\cap S_{G_j},\tilde{\beta}^t\right)>\frac{\epsilon^2}{4}(s_j\vee s_0)\tilde{\lambda}_2^2\right)+\mathrm{I}\left(\sum_{k\in S^*\cap S_{G_j}}\Xi_{kj}^2>\frac{\epsilon^2}{4}(s_j\vee s_0)\tilde{\lambda}_2^2\right), \end{aligned} \tag{108}$$

由此可得，

$$
\begin{aligned}
&\sqrt{\sum_{j\in G^*} s_0\tilde{\lambda}_2^2 \cdot \mathrm{I}\left(\sum_{k\in S^*\cap S_{G_j}} \left(\tilde{H}_{kj}^{t+1}\right)^2 < (s_j+s_0)\tilde{\lambda}_2^2\right)} \\
\leq &\sqrt{\sum_{j\in G^*} s_0\tilde{\lambda}_2^2 \cdot \mathrm{I}\left(\Upsilon\left(S^*\cap S_{G_j}, \tilde{\beta}^t\right) > \frac{\epsilon^2}{4}(s_j\vee s_0)\tilde{\lambda}_2^2\right)} \\
&+\sqrt{\sum_{j\in G^*} s_0\tilde{\lambda}_2^2 \cdot \mathrm{I}\left(\sum_{k\in S^*\cap S_{G_j}} \Xi_{kj}^2 > \frac{\epsilon^2}{4}(s_j\vee s_0)\tilde{\lambda}_2^2\right)} \qquad (109)\\
\leq &\frac{2}{\epsilon}\sqrt{\Upsilon\left(S^*, \tilde{\beta}^t\right)} + \sqrt{\sum_{j\in G^*} s_0\tilde{\lambda}_2^2 \cdot \mathrm{I}\left(\sum_{k\in S^*\cap S_{G_j}} \Xi_{kj}^2 > \frac{\epsilon^2}{4}(s_j\vee s_0)\tilde{\lambda}_2^2\right)} \\
\leq &\frac{2}{\epsilon}\delta\left\|\tilde{\beta}^t - \beta^*\right\|_2 + \sqrt{\sum_{j\in G^*} s_0\tilde{\lambda}_2^2 \cdot \mathrm{I}\left(\sum_{k\in S^*\cap S_{G_j}} \Xi_{kj}^2 > \frac{\epsilon^2}{4}(s_j\vee s_0)\tilde{\lambda}_2^2\right)}.
\end{aligned}
$$

现在，根据 *Hsu* 等人中的引理 *1* 和定理 *2.1*， *[2012],* 对于每个 $t>0$ 以及每个支撑集组 G_j，我们得到：

$$
P\left(\frac{n}{\sigma^2}\|\Xi_{S^*\cap S_{G_j}}\|_2^2 \geq s_j + 2(1+\delta)\sqrt{s_j t} + 2(1+\delta)t\right) \leq e^{-t}. \qquad (110)
$$

L且 $t=\frac{n\epsilon^2}{24\sigma^2}(s_j\vee s_0)\tilde{\lambda}_2^2)$。我们可以证明 $t>s_j$，因为 $\Delta\to\infty$。由 $\delta\leqq\frac{1}{4}$ 可得

$$
s_j + 2(1+\delta)\sqrt{s_j t} + 2(1+\delta)t \;\leq\; 6t = \frac{n\epsilon^2(s_j\vee s_0)}{4\sigma^2}\tilde{\lambda}_2^2, \qquad (111)
$$

这意味着 $P\left(\sum_{k\in S^*\cap S_{G_j}}\Xi_{kj}^2 > \frac{\epsilon^2}{4}(s_j\vee s_0)\tilde{\lambda}_2^2\right) \leq \exp\left(-\frac{n\epsilon^2}{24\sigma^2}(s_j\vee s_0\ \tilde{\lambda}_2^2)\right)$。因此，

由马尔可夫不等式，我们有

$$
\begin{aligned}
&P\left\{\sum_{j\in G^*} s_0\tilde{\lambda}_2^2\cdot \mathrm{I}\left(\sum_{k\in S^*\cap S_{G_j}}\Xi_{kj}^2>\frac{\epsilon^2}{4}(s_j\vee s_0)\tilde{\lambda}_2^2\right)\geq\frac{\sigma^2 s s_0}{n\Delta}\right\}\\
\leq&\frac{n\Delta}{\sigma^2 s s_0}\sum_{j\in G^*} s_0\tilde{\lambda}_2^2\cdot P\left(\sum_{k\in S^*\cap S_{G_j}}\Xi_{kj}^2>\frac{\epsilon^2}{4}(s_j\vee s_0)\tilde{\lambda}_2^2\right)\\
\leq&\frac{n\Delta}{\sigma^2 s s_0}\sum_{j\in G^*} s_0\tilde{\lambda}_2^2\cdot \exp\left(-\frac{n\epsilon^2(s_j\vee s_0)}{24\sigma^2}\tilde{\lambda}_2^2\right) \qquad (112)\\
\leq&\frac{n\Delta}{\sigma^2}\tilde{\lambda}_2^2\cdot\exp(-\frac{n\epsilon^2 s_0}{24\sigma^2}\tilde{\lambda}_2^2)\\
=&\frac{1}{32}(\frac{n\tilde{\lambda}_2^2}{\sigma^2})^2\exp(-\frac{\epsilon^2 s_0}{24}\cdot\frac{n\tilde{\lambda}_2^2}{\sigma^2})\\
=&o(1),\quad as\ \Delta\to\infty.
\end{aligned}
$$

结合(109) 和(112)，我们将第二项界定为

$$
\sqrt{\sum_{j\in G^*} s_0\tilde{\lambda}_2^2\cdot \mathrm{I}\left(\sum_{k\in S^*\cap S_{G_j}}\left(\tilde{H}_{kj}^{t+1}\right)^2<(s_j+s_0)\tilde{\lambda}_2^2\right)}\leq\frac{2}{\epsilon}\delta\left\|\tilde{\beta}^t-\beta^*\right\|_2+\sqrt{\frac{\sigma^2 s s_0}{n\Delta}}. \qquad (113)
$$

最后，根据(107)和(113)，可得

$$
P\left(\sqrt{\sum_{(i,j)\in S^*}\left(\tilde{H}_{ij}^{t+1}\right)^2\mathrm{I}\left((i,j)\notin\tilde{S}^{t+1}\right)}<\frac{4}{\epsilon}\delta\left\|\tilde{\beta}^t-\beta^*\right\|_2+2\sqrt{\frac{\sigma^2 s s_0}{n\Delta}}\right)\to 1, \qquad (114)
$$

如 $\Delta\to\infty$。

附录 C：次高斯随机设计示例

假设 ζ_1，ζ_2，$\cdots$，ζ_n 是独立同分布的 p 维各向同性次高斯随机向量，构成随机矩阵 $Z\in\mathbb{R}^{n\times p}$，其第 i 行 Z_i 记为 ζ_i。本文考虑如下生成的随机设计矩阵 X:

$$
X=Z\Sigma^{\frac{1}{2}}, \qquad (115)
$$

其中 Σ 是协方差矩阵。

根据 Zhou [2009] 和 Mendelson 等人 [2008], 给出的向量空间 $\mathcal{V}\in\mathbb{R}^p$的理论框架, 关键在于构造受限等距性质

Xv与 $\Sigma^{\frac{1}{2}}v$之间针对 $v \in \mathcal{V}$的性质。经验过程技术发挥着重要作用，我们首先定义高斯复杂度:

定义 3（高斯复杂度）给定一个子集 $\mathcal{V} \subseteq \mathbb{R}^p$，我们将 $\mathcal{V}$ 的高斯复杂度定义如下:

$$\ell^*(\mathcal{V}) \coloneqq \mathbf{E}_g \sup_{\theta \in \mathcal{V}} \left| \sum_{i=1}^{p} g_i \theta_i \right|,$$

其中 θ_i 是向量 θ 的各个分量，而 g_1、g_2、$\cdots$、g_p 是从 $\mathcal{N}(0,1)$ 分布中独立抽取的。特别地，给定一个半正定矩阵 Σ，我们定义

$$\tilde{\ell}^*(\mathcal{V}) \coloneqq \ell^*(\Sigma^{\frac{1}{2}}\mathcal{V}) = \mathbf{E}_g \sup_{v \in \mathcal{V}} \left| \langle \Sigma^{\frac{1}{2}} v, g \rangle \right| = \mathbf{E}_g \sup_{v \in \mathcal{V}} \left| \langle v, \Sigma^{\frac{1}{2}} g \rangle \right|.$$

根据范数的齐次性，我们只需考虑单位球面的子集 S^{p-1}，其定义如下:

$$S^{p-1} \coloneqq \{v \in \mathbb{R}^p : \|v\|_2 = 1\}.$$

我们使用的主要技术是以下经验过程结果:

引理8（Mendelson 等人的定理 2.1 [2008]）设 $1 \le n \le p$ 和 $0 < \delta < 1$。设 $\zeta \in \mathbb{R}^p$ 为一个参数为 α的各向同性次高斯随机向量。设 $\zeta_1, \ldots, \zeta_n$为 ζ的独立副本。将 X 定义为(115)中的随机矩阵，并令 $\mathcal{V}$ 满足：对于所有 $v \in \mathcal{V}$，均有$\Sigma^{\frac{1}{2}}v \in S^{p-1}$。如果样本量 n满足 $\bar{n} > c'\alpha^4\delta^2\tilde{\ell}^*(\mathcal{V})^2,$，那么以至少 $1 - \exp(-\bar{c}\delta^2 n/\alpha^4)$的概率，对于所有 $v \in \mathcal{V}$，均有:

$$1 - \delta \le \|Xv\|_2/\sqrt{n} \le 1 + \delta,$$

其中， c'、$\bar{c} > 0$ 为某些绝对常数。

记参数空间为 $\mathcal{V} \coloneqq \Theta^{m,\,d}(s,\ s_0) \cap \{v :\ \Sigma^{\frac{1}{2}}v \in S^{p-1}\}$. 然后，给定任意 $v \in \mathcal{V}$，我们假设

$$\rho_{\min} \le \frac{\left\|\Sigma^{1/2}v\right\|_2}{\|v\|_2} \le \rho_{\max}.$$

接下来，我们推导 $\tilde{}$双重稀疏结构的高斯复杂度 $\ell^*(\mathcal{V})$。我们记

$$U \coloneqq \Theta^{m,d}(s, s_0) \cap \{v : \|\Sigma^{\frac{1}{2}}v\|_2 \le 1\}.$$

回顾 $m \times d = p$。于是，我们有

$$\begin{aligned}\tilde{\ell}^{*}(\mathcal{V}) \leq \widetilde{\ell}^{*}(U) &= \mathbf{E}_{g} \sup_{t \in U} \left|\left\langle t, \Sigma^{1/2} g \right\rangle\right| \\ &\leq 3\sqrt{\log\left|\Theta^{m,d}(s, s_0)\right|} \sup_{t \in U} \sqrt{\mathbf{E}_{g}\left|\left\langle t, \Sigma^{1/2} g \right\rangle\right|^{2}} \\ &\leq C\sqrt{s s_0 \log(ed/s_0) + s\log(em/s)} \sup_{t \in U} \left\|\Sigma^{1/2} t\right\|_{2} \\ &\leq C\sqrt{s s_0 \log(ed/s_0) + s\log(em/s)},\end{aligned}$$

其中，第一个不等式由 Ledoux 和 Talagrand [1991] 的第3章得出。请注意，

$$\frac{\|Xv\|_2}{\sqrt{n}\|v\|_2} = \frac{\|Xv\|_2}{\sqrt{n}\|\Sigma_S^{\frac{1}{2}} v\|_2} \cdot \frac{\|\Sigma_S^{\frac{1}{2}} v\|_2}{\|v\|_2}.$$

Theref ore，根据引理8，对于 $n > C'\alpha^4\delta^2 \cdot (s s_0 \log(ed/s_0) + s\log(em/s))$，我们有

$$(1-\delta)\rho_{\min} \leq \frac{\|Xv\|_2}{\sqrt{n}\|v\|_2} \leq (1+\delta)\rho_{\max}.$$

这证明了 s 次高斯随机设计下满足 DSRIP 条件 .

参考文献

Balas Kausik Natarajan. 线性系统的稀疏近似解。*SIAM* 计算杂志，24(2):227–234, 1995。

Dimitris Bertsimas、Angela King 和 Rahul Mazumder。以现代优化视角进行最佳子集选择。统计年刊，44(2):813–852，2016年04月。doi:10.1214/15- AOS1388。URL https://doi.org/10.1214/15-AOS1388。

Xiao-Tong Yuan、Ping Li 和 Tong Zhang。梯度硬阈值追踪。机器学习研究杂志，18(166):1–43，2018年。URL http://jmlr.org/papers/v18/ 14-415.html。

Jian Huang、Yuling Jiao、Yanyan Liu 和 Xiliang Lu。一种针对 l_0惩罚回归的构造性方法。机器学习研究杂志，19(10):1–37，2018年。URL http://jmlr.org/papers/v19/17-194.html。

Junxian Zhu、Canhong Wen、Jin Zhu、Heping Zhang 和 Xueqin Wang。最佳子集选择问题的多项式时间算法。美国国家科学院院刊，117(52)：33117–33123，2020年。ISSN 0027-8424。doi:10.1073/pnas.2014241117。URL https://www.pnas.org/content/117/52/33117。

Yonina C. Eldar、Patrick Kuppinger 和 Helmut Bolcskei。块稀疏信号：不确定性关系与高效恢复。*IEEE*信号处理汇刊，58(6)：3042–3054，2010年。doi:10.1109/TSP.2010.2044837。

Junzhou Huang、Tong Zhang 和 Dimitris Metaxas。利用结构化稀疏性进行学习。机器学习研究杂志，12(103):3371–3412，2011年。URL http://jmlr.org/papers/v12/huang11b.html。

Hussein Hazimeh、Rahul Mazumder 和 Peter Radchenko。利用离散优化进行分组变量选择：计算与统计视角。统计年刊，51(1):1 – 32，2023。doi:10.1214/21-AOS2155。URL https://doi.org/10. 1214/21-AOS2155。

Yanhang Zhang、Junxian Zhu、Jin Zhu 和 Xueqin Wang。采用拼接方法进行最佳分组子集选择。*INFORMS*计算期刊，35(1):104–119，2023。doi:10.1287/ijoc.2022.1241。URL https://doi.org/10.1287/ijoc.2022.1241。

Jerome Friedman、Trevor Hastie 和 Robert Tibshirani。关于分组套索和稀疏组Lasso的一则说明。*arXiv* 预印本 *arXiv:1001.0736*，2010年。

Noah Simon、Jerome Friedman、Trevor Hastie 和 Robert Tibshirani。稀疏分组套索。计算与图形统计学期刊，22(2):231–245，2013年。

Robert Tibshirani。通过 Lasso 进行回归收缩与变量选择。皇家统计学会会刊：*B*辑（方法论），58(1):267–288，1996年。doi:https://doi.org/10.1111/j.2517-6161.1996.tb02080.x。网址 https://rss. onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1996.tb02080.x。

Ming Yuan 和 Yi Lin。含分组变量的回归模型中的模型选择与估计。皇家统计学会会刊：*B*辑（统计方法论），68(1):49–67，2006年。doi:https://doi.org/10.1111/j.1467-9868.2005.00532.x。网址 https://rss. onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2005.00532.x。

Yasutoshi Ida、Yasuhiro Fujiwara 和 Hisashi Kashima。快速稀疏分组套索。神经信息处理系统进展，32，2019年。

Yangjing Zhang、Ning Zhang、Defeng Sun 和 Kim-Chuan Toh。用于求解大规模稀疏组套索问题的高效基于海森矩阵的算法。*Mathematical Programming*，179(1):223–263，2020年。

T. Tony Cai、Anru R. Zhang 和 Yuchen Zhou。稀疏组Lasso：最优样本复杂度、收敛速度与统计推断。*IEEE*信息论汇刊，68(9)：5975–6002，2022年。doi:10.1109/TIT.2022.3175455。

Zhifan Li、Yanhang Zhang 和 Jianxin Yin。估计$\ell_u(\ell_q)$-球上的双重稀疏结构：极小极大速率与相变。*IEEE*信息论汇刊，70(10)：7066–7088，2024年。doi:10.1109/TIT.2024.3451512。Pierre C Bellec。Lasso及其他凸估计量的噪声屏障与大信号偏差。*arXiv*预印本 *arXiv:1804.01230*，2018年。Thomas Blumensath 和 Mike E Davies。压缩感知的迭代硬阈值方法。应用与计算调和分析，27(3)：265–274，2009年。Thomas Blumensath 和 Mike E Davies。归一化迭代硬阈值：保证的稳定性与性能。*IEEE*信号处理精选主题期刊，4 (2)：298–309，2010年。Prateek Jain、Ambuj Tewari 和 Purushottam Kar。关于高维m估计的迭代硬阈值方法。神经信息处理系统进展，27，2014年。

Xiao-Tong Yuan、Bo Liu、Lezi Wang、Qingshan Liu 和 Dimitris N. Metaxas。双重迭代硬阈值法。机器学习研究杂志，21(152):1–50，2020年。网址：http://jmlr.org/papers/v21/18-487.html。

Meiling Hao、Lianqiang Qu、Dehan Kong、Liuquan Sun 和 Hongtu Zhu。大规模矩阵线性回归模型的最优极小极大变量选择。机器学习研究杂志，22(147):1–39，2021。网址：http://jmlr.org/papers/v22/19-969. html。

Christophe Giraud。《高维统计学导论》。CRC出版社，2021年。Mohamed Ndaoud。高维线性回归中的尺度化极小极大最优性：一种非凸算法正则化方法。*arXiv*预印本 *arXiv:2008.12236*，2020年。Nicolas Verzelen。稀疏回归的极小极大风险：超高维现象。《电子统计学杂志》，6:38–90，2012年。Cristina Butucea、Mohamed Ndaoud、Natalia A. Stepanova 和 Alexandre B. Tsybakov。采用汉明损失的变量选择。统计年刊，46(5):1837 – 1875，2018年。doi:10.1214/17-AOS1572。网址：https://doi.org/10.1214/17-AOS1572。E.J. Candes 和 T. Tao。通过线性规划进行解码。*IEEE*信息论汇刊，51(12):4203–4215，2005年。doi:10.1109/TIT.2005.858979。Garvesh Raskutti、Martin J. Wainwright 和 Bin Yu。 ℓ_q -球上高维线性回归估计的极小极大速率。*IEEE*信息论汇刊，57(10):6976–6994，2011。doi:10.1109/TIT.2011.2165799。

Pierre C. Bellec、Guillaume Lecué 和 Alexandre B. Tsybakov。Slope 与 Lasso 相遇：改进的预言界与最优性。统计年刊，46(6B):3603 – 3642，2018年。doi:10.1214/17-AOS1670。网址：https://doi.org/10.1214/17-AOS1670。Junzhou Huang 和 Tong Zhang。组稀疏性的益处。统计年刊，38(4):1978 – 2004，2010年。doi:10.1214/09-AOS778。网址：https://doi.org/ 09-AOS778。Karim Lounici、Massimiliano Pontil、Sara van de Geer 和 Alexandre B. Tsybakov。组稀疏性下的预言不等式与最优推断。统计年刊，39(4):2164 – 2204，2011年。doi:10.1214/11-AOS896。网址：https://doi.org/ 11-AOS896。Edgar N Gilbert。信号字母表的比较。《贝尔系统技术期刊》，31(3):504–522，1952年。Lucien Birgé 和 Pascal Massart。高斯模型选择。《欧洲数学学会杂志》，3(3):203–268，2001。Christophe Giraud、Sylvie Huet 和 Nicolas Verzelen。未知方差下的高维回归。《统计科学》，27(4):500 – 518，2012年。doi:10.1214/12-STS398。网址：https://doi.org/10.1214/12-STS398。Yonina C. Eldar 和 Moshe Mishali。从结构化子空间并集中稳健恢复信号。*IEEE*信息论汇刊，55(11):5302–5316，2009年。doi:10.1109/TIT.2009.2030471。Xiaoxuan Liang、Aaron Cohen、Anibal Sólon Heinsfeld、Franco Pestilli 和 Daniel J. McDonald。sparsegl：用于估计稀疏组Lasso的 R软件包。《统计软件杂志》，110(6):1–23，2024年。doi:10.18637/jss.v110.i06。网址：https://www.jstatsoft.org/index.php/jss/article/view/v110i06。Jian Huang、Shuange Ma、Huiliang Xie 和 Cun-Hui Zhang。用于变量选择的组桥方法。《生物计量学》，96(2):339–355，2009年。Patrick Breheny。用于双层变量选择的组指数Lasso。《生物统计学》，71 (3):731–740，2015年。Patrick Breheny 和 Jian Huang。用于双层变量选择的惩罚方法。《统计学及其应用》，2(3):369，2009年。Jiahua Chen 和 Zehua Chen。针对大型模型空间模型选择的扩展贝叶斯信息准则。《生物计量学，95(3):759–771，2008年。

Todd E. Scheetz、Kwang-Youn A. Kim、Ruth E. Swiderski、Alisdair R. Philp、Terry A. Braun、Kevin L. Knudtson、Anne M. Dorrance、Gerald F. DiBona、Jian Huang、Thomas L. Casavant、Val C. Sheffield 和 Edwin M. Stone。哺乳动物眼睛中基因表达的调控及其与眼病的相关性。美国国家科学院院刊，103(39):14429–14434，2006年。ISSN 0027-8424。doi:10.1073/pnas.0602562103。URL https://www.pnas.org/content/103/39/14429。

Jian Huang、Joel L. Horowitz 和 Fengrong Wei。非参数加性模型中的变量选择。统计年刊，38(4):2282 – 2313，2010年。doi:10.1214/09- AOS781。URL https://doi.org/10.1214/09-AOS781。

Jianqing Fan、Yang Feng 和 Rui Song。稀疏超高维加性模型中的非参数独立性筛选。美国统计学会会刊，106(494)：544–557，2011年。doi:10.1198/jasa.2011.tm09779。URL https://doi.org/10. 1198/jasa.2011.tm09779。

Annie P Chiang、John S Beck、Hsan-Jan Yen、Marwan K Tayeh、Todd E Scheetz、Ruth E Swiderski、Darryl Y Nishimura、Terry ABraun、Kwang-Youn A Kim、Jian Huang 等。利用SNP芯片进行纯合性作图，鉴定出TRIM32（一种E3泛素连接酶）为巴德-毕德综合征基因（BBS11）。美国国家科学院院刊，103(16)：6287–6292，2006年。

Wenliang Pan、Xueqin Wang、Weinan Xiao 和 Hongtu Zhu。一种通用的确定性独立筛选程序。美国统计学会会刊，114(526)：928–937，2019年。doi:10.1080/01621459.2018.1462709。URL https://doi.org/10.1080/01621459. 2018.1462709。PMID: 31692981。

Patrick Breheny 和 Jian Huang。用于含分组预测变量的非凸惩罚线性回归和逻辑回归模型的组下降算法。统计与计算，25(2): 173–187，2015年。

Garvesh Raskutti、Martin J. Wainwright 和 Bin Yu。通过凸规划求解核类上的稀疏加性模型的极小极大最优速率。机器学习研究期刊，13(13):389–427，2012年。网址：http://jmlr.org/papers/v13/raskutti12a.html。

Ming Yuan 和 Ding-Xuan Zhou。高维加性模型中的极小极大最优估计率。统计年刊，44(6):2564 – 2593，2016。doi:10.1214/15-AOS1422。网址：https://doi.org/10.1214/15-AOS1422。

Haoyang Liu、Chao Gao 和 Richard J. Samworth。《稀疏高维变点检测的极小极大速率》。《统计年刊》，49(2):1081–1112，2021。doi:10.1214/20-AOS1994。URL https://doi.org/10.1214/20-AOS1994。

Daniel Hsu、Sham Kakade 和 Tong Zhang。次高斯随机向量二次型的尾不等式。《概率电子通信》，17:1–6，2012。Thomas M. Cover 和 Joy A Thomas。《信息论基础》。Wiley-Interscience，2006年。B. Laurent 和 P. Massart。通过模型选择自适应估计二次泛函。《统计年刊》，28(5):1302 – 1338，2000。doi:10.1214/aos/1015957395。URL https://doi.org/10.1214/aos/1015957395。Mark Rudelson 和 Roman Vershynin。Hanson-Wright 不等式与次高斯集中。《概率电子通信》，18(none):1 – 9，2013。doi:10.1214/ECP.v18-2865。URL https://doi.org/10.1214/ECP.v18-2865。Shuheng Zhou。次高斯随机矩阵上的受限特征值条件。《arXiv 预印本 arXiv:0912.4045》，2009。Shahar Mendelson、Alain Pajor 和 Nicole Tomczak-Jaegermann。伯努利与次高斯系综的一致不确定性原理。《构造性逼近》，28: 277–289，2008。Michel Ledoux 和 Michel Talagrand。《巴拿赫空间中的概率：等周问题与随机过程》，第23卷。Springer Science & Business Media，1991。