arXiv:2303.02118v2 [stat.ML] 6 Jul 2023

混合稀疏线性回归中的统计—计算权衡

Gabriel Arpino* Ramji Venkataramanan†

2023年7月7日

摘要

我们研究含两个分量的混合稀疏线性回归问题：需要从 n 个无标签含噪线性测量中恢复两个k-稀疏信号 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathbb{R}^p$ 。稀疏度允许相对于维数呈次线性增长（$k = o(p)$），并假设加性噪声为相互独立的高斯噪声，方差为 σ^2。先前工作表明，该问题存在$\frac{k}{\mathrm{SNR}^2}$到-$\frac{k^2}{\mathrm{SNR}^2}$ 的统计-计算差距，与其他具有计算挑战性的高维推断问题相似，例如稀疏主成分分析和鲁棒稀疏均值估计 [Brennan 和 Bresler, 2020b]；此处，$\mathrm{SNR} := \|\boldsymbol{\beta}_1\|_2/\sigma^2 = \|\boldsymbol{\beta}_2\|_2/\sigma^2$ 表示信噪比。我们通过低次多项式方法证明，该问题存在更广泛的$\frac{k}{\mathrm{SNR}^2}$-到-k^2（$\mathrm{SNR}+1^{\,2}$）计算障碍，但同时表明，只有在一个非常狭窄的对称参数区域内，该问题才具有计算困难性。对于此~~困难区域中~~的任何随机化算法，我们确定了样本复杂度 n 与运行时间 $\exp(\tilde{\Theta}(k^2(\mathrm{SNR}+1)^2/(n\mathrm{SNR}^2))$ 之间平滑的信息-计算权衡。通过一个简单的归约，这为稀疏相位恢复中精确支撑恢复存在计算障碍提供了新的严格证据，其样本复杂度为 $n = \tilde{o}(k^2)$。我们的第二项贡献是分析一种简单的阈值算法：在问题不困难的狭窄区域之外，该算法能在$O(np)$时间内解决相应的混合回归检测问题，并达到（非混合）稀疏线性回归所需的样本复杂度 $k^{(\mathrm{SNR}+1)\,\mathrm{SNR}} \log p$；由此，随后可以利用稠密情形中的先进技术解决恢复问题。作为我们结果的一个特例，我们表明，在一大类算法中，该简单算法在解决稀疏线性回归的精确有符号支撑恢复问题时具有阶最优性。据我们所知，这是首次全面研~~究混合对~~称性、信号稀疏性及其共同影响如何决定混合稀疏线性回归的计算困难性。

*剑桥大学，ga442@cam.ac.uk。
†剑桥大学，rv285@cam.ac.uk。

目录

1 引言 **1**1.1 我们的贡献 .31.2 与先前工作的联系 .41.3 低次方法.6

2 主要结果 **7**
2.1 MSLR 的下界.72.2 MSLR 的算法.82.3 稀疏线性回归（SLR）的下界 .92.4 SLR 的算法.10

3 证明思路 **10**4 讨论 **12**A 关于低次方法的补充背景 **18**B 低次下界的证明 **19**

B.1 MSLR的低次分析：一般混合 .21B.2 特例： SLR − D .27B.3 特例： SB-MSLR − D .28

C 多项式时间归约的证明 多项式时间归约 **3** **0**
C.1 从 SB-MSLR − D 归约到 SB-MSLR .33C.2 从 SB-MSLR归约到 PSB-MSLR .35C.3 从 SB-MSLR − D 到 SPR .37

D 高效算法的证明 **39**
D.1 CORR 用于 MSLR .39D.2 CORR 用于 MSLR − D .50MSLR 的恢复算法 .50D.4 CORR 用于 SLR 中带符号支撑恢复 SLR .56

1 引言

本文研究双分量混合稀疏线性回归（MSLR）问题，目标是根据 n 无标签含噪线性测量，估计两个 k-稀疏信号 $\boldsymbol{\beta}_1$、$\boldsymbol{\beta}_2 \in \mathbb{R}^p$。该模型定义如下。

Defi定义 1.1（MSLR）。对于 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$、$\boldsymbol{w} \in \mathbb{R}^n$ 和 $\boldsymbol{z} \in \mathbb{R}^n$，考虑以下模型：

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_1 \odot \boldsymbol{z} + \boldsymbol{X}\boldsymbol{\beta}_2 \odot (1-\boldsymbol{z}) + \boldsymbol{w},$$

其中，$\odot$ 表示向量之间的逐元素乘积，$X_{i,\ j}$ 独立同分布$\sim$ $\mathcal{N}(0,1)$，w_i 独立同分布$\sim$ $\mathcal{N}(0,\sigma^2)$，z_i 独立同分布$\sim$伯努利分布(ϕ)，且 $\boldsymbol{\beta}_1$、$\boldsymbol{\beta}_2 \in \mathbb{R}^p$ 均为 k-稀疏。给定 $(\boldsymbol{X},\boldsymbol{y})$，目标是估计 $\boldsymbol{\beta}_1$、$\boldsymbol{\beta}_2$。

该模型由 Quandt 和 Ramsey [1978] 提出，此后在机器学习和统计学领域得到了广泛研究；例如参见 St¨adler 等人 [2010], Chen 等人 [2014],Yi 等人 [2014], Fan 等人 [2018], Javanmard 等人 [2022]以及其中的参考文献。

如果潜变量（z_i）$_{i\in[n]}$ 是已观测到的，那么问题就简化为求解两个独立的线性回归。然而，在许多应用中，潜变量可能是未知的，因为数据可能来自不同的无标签子群体。MSLR 模型刻画了这一现象，并已应用于多种场景，包括市场细分 [Wedel 和 Kamakura， 2000], 音乐感知 [Viele 和 Tong， 2002], 医疗保健 [Deb 和 Holmes，2000，Luo 等人，2022，Im 等人，2022], 以及其他各种领域 [Li 等人，2022，Kazor 和 Hering， 2019]。被称为分层专家混合模型的混合回归模型变体早已在机器学习领域得到研究 [Jordan 和 Jacobs， 1994], ，并被用于集成学习以及门控循环单元和注意力网络中 [Makkuva 等人， 2019]。

最大似然估计量是估计信号 $\boldsymbol{\beta}_1$、$\boldsymbol{\beta}_2$ 的自然选择。然而，由此产生的优化问题是非凸且NP难的 [Yi 等人， 2014]。因此，该问题在统计和计算方面都颇具挑战性，研究人员也提出了各种高效的估计量。这些方法包括谱方法 [Chaganty and Liang, 2013, Yi 等，2014, Zhang et al., 2022], 期望最大化（EM） [Khalili and Chen, 2007, Faria and Soromenho, 2010, St¨adler et al.， 2010], 交替最小化 [Yi 等，2014, Shen and Sanghavi, 2019, Ghosh and Kannan, 2020], 凸松弛 [Chen 等人， 2014], 矩下降方法 [Li and Liang, 2018, Chen 等人， 2020], 以及使用可处理的非凸目标函数 [Zhong et al., 2016, Barik and Honorio

，

2022]。

尽管近期已有研究探讨混合线性回归在统计和计算上的可行性（包括但不限于 Azizyan 等人 [2013], Pal 等人 [2022, 2021]），但对于高维稀疏区域中的这一问题，人们仍知之甚少。在该区域，样本量 n和稀疏度 k 都可能随维度 p次线性增长，且涵盖所有 信噪比 缩放范围，其中信噪比：$= \|\boldsymbol{\beta}_1\|_2^2/\sigma^2 = \|\boldsymbol{\beta}_2\|_2^2/\sigma^2$ 表示信号与噪声之比。这一区域受到近年来各类统计应用的推动，应用领域从生物学到通信（相关参考文献可见专著Hastie 等人 [2015], Giraud [2021]，其中包含多篇参考文献）。高维统计文献中常将定义 1.1 中的独立同分布高斯数据行 $\boldsymbol{x}_i$ 和加性高斯噪声 w_i 假设广泛视为一种理想化假设（例如，Wainwright[2009a、b], Arias-Castro 等人 [2011], Janson 等人 [2017]）。

这一表述中开始显现的一个特点是，在对称参数区域内，MSLR 问题很难，也就是说，在信息论最优样本复杂度下，无法通过多项式时间算法求解 $n_{\mathtt{IT}} = \tilde{\Theta}(k/\text{信噪比}^2)$ [Fan 等人， 2018]。穷举搜索通常能为信号支撑集得到统计上近乎最优的估计量，但其运行时间随 k呈指数增长。近期，Brennan 和 Bresler [2020b], Fan 等人

[2018] 以不同方式量化了这一现象，揭示了算法面临的根本性障碍：在一个非常狭窄且对称的参数范围内，对于所有样本复杂度 $n = \tilde{o}(k^2/\text{信噪比}^2)$ 和稀疏度$k = o(\sqrt{p})$，算法都受到这一障碍的限制。我们将该参数范围称为对称平衡稀疏线性回归混合（SB-MSLR），定义如下：

$$\texttt{SB-MSLR}: \ \phi = 1/2 \text{ and } \boldsymbol{\beta}_1 = -\boldsymbol{\beta}_2. \tag{1}$$

（这里回顾一下，ϕ 是定义 1.1 中的混合参数，因此 $\phi = \frac{1}{2}$ 意味着每个y_i 来自 $\boldsymbol{\beta}_1$ 或 $\boldsymbol{\beta}_2$的概率相同。）这一现象被称为 $\frac{k}{\text{SNR}^2}$至$\frac{k^2}{\text{SNR}^2}$ 的统计-计算差距：该问题只需数量级为 k/SNR^2 的样本即可求解，但高效算法至少需要数量级为 k^2/SNR^2 的样本。（本文中，高效算法是指运行时间为 $O(p^\eta)$ 的算法，其中 $\eta > 0$为某个常数。）这一计算阈值的数量级与许多统计估计问题的阈值相似，这些问题涵盖了植入团的各种变体，例如稀疏主成分分析和稳健均值估计 [Brennan 和 Bresler, 2020b]。值得注意的是，SB-MSLR 与一种重要的稀疏相位恢复问题的表述相近$\boldsymbol{y} = |\boldsymbol{X}\boldsymbol{\beta}| + \boldsymbol{w}$ [Brennan 和 Bresler, 2020b, Fan 等人，2018], ；该问题已得到广泛研究，并被认为存在 k至k^2 的统计-计算差距 [Liu 等人, 2021, Wu and Rebeschini,2021]。

过去几十年来，稀疏线性回归（SLR）的特殊情形——只有一个信号的情形（即定义 1.1 中的 $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$） —— 已得到广泛研究 [Candes 和 Tao，2005，Donoho，2006，Wainwright，2009b]。对于 SLR，统计-计算差距要小得多，但仍然存在。事实上，在 $k = o(p)$ 的情形下，SLR 的信息论最优样本复杂度为 $\frac{k}{\log(p/k)} \log(1+\text{SNR})$ 量级 [Wang 等人，2010，Reeves 等人，2019]；相比之下，Bandeira 等人 [2022], Gamarnik 和 Zadik [2022] 近期的研究通过对重叠间隙性质和低次多项式的分析，确立了在 SNR $\to \infty$ 这一情形下的下界，并证明，高效求解 SLR 所需的样本复杂度至少为 $k \log p$ 量级。此外，使用多种算法可以得到同一量级的上界[Wainwright，2009b，Bandeira 等人，2022，Gamarnik 和 Zadik， 2022]。

本文针对 MSLR以及其特殊情形 SLR，通过分析一种简单的阈值算法，给出了新的算法下界和上界。我们将这种阈值算法称为 CORR，Bandeira 等人 [2022] 曾用它为 $o(k)$ 个误差以内的近似支撑恢复推导出 SLR的上界，其设定为二值信号且信噪比 SNR $\to \infty$ 随 k增长。在我们的所有结果中，我们都明确给出了对 SNR 的依赖，因此结果适用于所有 SNR 区间，包括 SNR $= \Theta(1)$，以及 SNR $= o(1)$。在总结结果之前，我们先定义所考虑的信号先验分布类别 $\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2$。

信号先验 我们考虑关于 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ 的联合先验，其在 k-稀疏向量 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathbb{R}^p$ 上边缘均匀，且范数相等 $\|\boldsymbol{\beta}\|_2$。两个信号范数相等的情形更具挑战性，因为无论观测 $\boldsymbol{y}$ 中的每个条目对应哪个信号，其方差都相同。我们将此类先验记为 $\mathcal{P}_{\|\boldsymbol{\beta}\|_2}(\mathcal{D})$，其中每个向量的非零条目取值于 $\mathcal{D} \subseteq \mathbb{R}$。我们假设 $\boldsymbol{\beta}_{\min} := \min\{|\beta| \mid \beta \in \mathcal{D}\} > 0$。

记号 我们用粗体表示向量和矩阵，用正体表示标量（例如$\boldsymbol{a}$ 和 a，分别表示向量和矩阵）。对于 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$， $\boldsymbol{x}_i$ 表示该矩阵的第 i行， $\boldsymbol{X}_j$ 表示该矩阵的第 j列。全文采用标准渐近记号$O(\cdot)$、$\Omega(\cdot)$ $o(\cdot)$ $\omega(\cdot)$ 和 $\Theta(\cdot)$。我们用 $O(\cdot)$及类似形式表示相差至多多对数因子的关系。用 $\lesssim$、 $\simeq$分别表示相差至多常数因子的不等式和等式。令 $[n] := \{1, 2 \cdots n\}$。对于定义 1.1 中的 MSLR 设置以及参数范围 SB-MSLR,

(1)，我们将 MSLR \ SB-MSLR 称为 MSLR 问题，其相关参数不在SB-MSLR参数范围内。

1.1 我们的贡献

下文中的计算下界结果对有界幅度信号在尺度范围 $p \to \infty$, $n \to \infty$ 和 $k = o(\sqrt{p})$ 内完全普遍成立。我们的算法可实现性结果则以高概率适用于一般信号，在次线性稀疏度区间$p \to \infty$, $n \to \infty$, $k = o(p)$，以及 $n = \omega(k)$。

MSLR 的计算下界 我们通过研究低次多项式，提供了新的严格证据 [Kunisky 等人，2022年，Schramm 和 Wein，2022年，Hopkins， 2018]，表明在所有样本复杂度 $n = o(k^{2}(\mathrm{SNR}+1)^{2}/\mathrm{SNR}^{2}\cdot\log p)$ 和稀疏度 $k = o(\sqrt{p})$ 下，求解 SB-MSLR 的检测（假设检验）变体都存在根本性的算法障碍。此外，我们表明，这一计算障碍意味着样本复杂度与时间复杂度之间存在平滑的权衡，从而阻止运行时间小于 $\exp(\tilde{\Theta}(\frac{k^2}{n}\cdot(\text{信噪比}+1)^2/\text{信噪比}^2))$ 的算法取得成功。这些结果扩展了 Brennan 和 Bresler [2020b], Fan 等人 [2018] 的研究：它们表明 SB-MSLR 在所有信噪比范围内（包括无噪声和信噪比 $= \omega(1)$范围）均存在显著的统计-计算差距，并在困难范围内确定了样本量与运行时间之间的平滑权衡。

随后，我们针对信号取非零值于 $\{1, -1\}$ 的情形，给出 SB-MSLR 的检测与恢复两种变体之间的多项式时间归约，从而将我们的困难性结果转化为证据，表明随着信噪比增大，精确支撑恢复同样困难。我们还通过将 SB-MSLR 的精确恢复问题归约到更一般的部分对称平衡 MSLR（即 PSB-MSLR）情形中的精确恢复问题，证明任何将 SB-MSLR 包含为子问题的 MSLR 情形都必然是困难的，其中

$$\text{PSB-MSLR}: \phi = \frac{1}{2}, \text{ and } \boldsymbol{\beta}_{1,j} = -\boldsymbol{\beta}_{2,j} \text{ for } j \in J \subseteq \mathrm{supp}(\boldsymbol{\beta}_1) \cap \mathrm{supp}(\boldsymbol{\beta}_2), \text{ with } |J| = \Theta(k). \tag{2}$$

我们针对 SB-MSLR 无噪声版本所得到的计算下界，在稀疏相位恢复中为精确支撑集恢复带来了等价的下界，其中 $\boldsymbol{y} = |\boldsymbol{X}\boldsymbol{\beta}| + \boldsymbol{w}$。这为计算障碍以及在 $n = \tilde{o}(k^2)$ 个样本下求解稀疏相位恢复中的精确支撑集恢复问题时平滑的信息-计算权衡，提供了新颖而严谨的证据，并回应了关于这一问题难度的一个重要开放问题 [Liu 等人，2021年，Brennan 和 Bresler，2020b，Wu andRebeschini， 2021]。

用于 MSLR 然而，或许令人惊讶的是，我们证明在 SB-MSLR之外，上述算法障碍便不复存在。我们证明，一种名为 CORR 的简单阈值算法能够解决 MSLR 在 SB-MSLR 之外的检测变体，其运行时间为 $O(np)$，样本复杂度为 n，数量级为 $^{k}(\mathrm{SNR}+1)\ \mathrm{SNR}\log p$，与高效求解稀疏线性回归所需的复杂度相匹配。我们注意到，SB-MSLR 是一个非常狭窄的参数范围。事实上，对于相对于勒贝格测度绝对连续的信号先验（针对非零值），约束（1）几乎必然不成立，因此，CORR在一个测度为一的集合上均能成功。

就原始恢复问题而言，CORR已被证明能够精确恢复两个信号的联合支撑，其适用范围略宽于 PSB-MSLR（精确表述见定理 2.6 ）。恢复联合支撑后，问题便归结为稠密情形或成比例稀疏情形（$k/p = n/p = \Theta(1)$），此时现有算法能够精确推断 $\boldsymbol{\beta}_1$ 和 $\boldsymbol{\beta}_2$。这拓展了Mazumdar 和 Pal [2022] 近期的工作，该工作提出了一种精确的联合支撑恢复算法

该算法适用于二元信号（取自 $\{0,1\}^p$），样本复杂度为$k(\mathrm{SNR}+1)\ \mathrm{SNR}\ \log^3 p$。我们强调，假设二元信号的所有非零条目均等于 1 是有限制的，因为它未涵盖问题较难的 SB-MSLR 和 PSB-MSLR 这两个重要参数范围。我们可以用集合记号概括算法上较难的参数范围：

$$\underset{\text{SB-MSLR}}{\text{“Low-degree hard detection”}} \subset \underset{\text{PSB-MSLR}}{\text{“Exact support recovery is hard by reduction”}} \subset \text{MSLR}.$$

稀疏线性回归的下界与算法 我们的结果也阐明了稀疏线性回归（SLR）这一特殊情形中出现的计算障碍，其中 $\beta_1 = \beta_2$。正如引言中所述，先前的研究者已经证明，高效算法所需的样本复杂度至少为$k \log p$ [Bandeira 等人，2022，Gamarnik 和 Zadik， 2022]，；而在 SNR $\to \infty$ [的情形下，已有相应的算法上界结果与之匹配 Wainwright，2009b，Bandeira 等人，2022，Gamarnik 和 Zadik， 2022]。我们扩展了这些发现，并提供了严格的低次数证据，表明对于 SLR 的检测变体（在 $\|\boldsymbol{\beta}\|_2^2$ 为 k量级的情形下），多项式时间算法所需的样本复杂度至少为$n_{\mathrm{alg}}^{\mathrm{SLR}} := {}^{k}(\mathrm{SNR}+1)\ \mathrm{SNR}\ \log p$ 。我们的证明技术采用了针对 SLR 的朴素低次数计算；这不同于 Bandeira 等人 [2022], 的方法，他们建立了低次方法与 Franz-Parisi 判据之间的联系，从而得到了 SLR 的计算下界。我们的直接证明技术使我们能够明确量化 SNR 在该问题中的作用。

此外，我们证明，对于所有信噪比缩放方式和一般稀疏信号先验，CORR 使用$9n_{\mathrm{alg}}^{\mathrm{SLR}}$ 个样本即可解决 SLR 中的检测和带符号支撑恢复问题。此外，其运行时间为 $O(np)$，根据所采用的收敛准则，这可能显著优于 Lasso 等替代方案 [Wainwright，2009b]。这进而证明了：对于输入的次数至多为 $O(\log p)$ 的解析多项式这一类算法（包括运行 $O(\log\ p)$ 次迭代的谱方法），CORR 在 SLR 中实现精确有符号支撑恢复具有阶最优性。我们注意到，除非信噪比 $= \omega(1)$，否则 SLR 中的统计-计算差距仅在乘法常数意义下介于 ${}^{k}_{\log(p/k)\ \log(1+\mathrm{SNR})}$与 $n_{\mathrm{alg}}^{\mathrm{SLR}}$ 之间。

下表 1 总结了我们的贡献及已有结果，其中信号取值于 $\{-1,0,1\}$。在表 1中， n_{IT} 表示检测的信息论阈值（通过归约也适用于恢复），而 n_{alg} 表示高效算法的样本阈值。重要的是，我们表明，MSLR在狭窄的 SLR 之外，其表现与 SB-MSLR 区域相似；并且我们统一了文献中已有的结果：这些结果证明，在二元情形下，可达到的样本复杂度为 k [Mazumdar 和 Pal， 2022] 而在一般情形下为 k^2 [St¨adler 等人， 2010]。这些结果使我们认为，$\frac{k}{\mathrm{SNR}^2}$-至-$k^2(\mathrm{SNR}+1\ ^2)$差距源于信号中的脆弱对称性，也表明 SB-MSLR 与 SLR 在计算上是截然不同的问题；前者在高维情形下只能低效求解。

1.2 与先前工作的联系

最早严格证明统计到计算的差距的研究之一是Barak 等人[2016] ，他们利用平方和（SOS）层次结构证明了植入团（PC）问题的紧计算下界。基于平方和方法，Hopkins [2018] 随后提出了一个猜想（即下一小节所述猜想1.2 的一个版本），认为低次多项式对于假设检验具有最优性。这一方法为稀疏主成分分析等高维推断问题中的计算障碍提供了证据[Hopkins 和 Steurer，2017年；Bandeira 等人， 2020]。证明计算障碍的其他方法还包括

	信息论下界 n_{IT}	算法下界 n_{alg}	算法
MSLR（先前）	$\tilde{\Theta}(k/\text{信噪比}^2)$ [Fan 等人，2018]	$\tilde{\Theta}(k^2/\text{信噪比}^2)$ [Fan 等人，2018年, Brennan 和 Bresler, 2020b]	ℓ_1-惩罚 $(n = \Omega(k^2)$, 信噪比 $\to \infty)$; 多项式恒等式-对于取值为 0-1 的信号，(复杂度 $n = \Omega(\frac{k(\text{SNR}+1)}{\text{SNR}} \log^3 p)$) [St¨adler 等人，2010年, Mazumdar 和 Pal，2022]
MSLR（本文工作） SB-MSLR、PSB-MSLR MSLR \ SB-MSLR	$\Theta\left(\frac{2k \text{ 对数 } (p/k)}{\text{对数}_2(1+\text{SNR})}\right)$	$\Theta\left(\frac{k^2(\text{SNR}+1^{\,2}\quad 1)}{\text{SNR}^2} \frac{}{\text{对数 } p}\right)$	CORR $(n = \Omega(\frac{k(\text{SNR}+1)}{\text{SNR}} \text{对数 } p))$
SLR（前文）	[Wang 等人，2010年, Gamarnik 和 Zadik，2022，Reeves 等人，2019]	$\Theta(k \text{ 对数 } p)$ [Wainwright，2009b, Gamarnik 和 Zadik, 2022年, Bandeira 等人, 2022年，Arpino，2021]	Lasso, CORR，搜索, OMP $(n = \Omega(k \log p)$, SNR $\to \infty)$ [Wainwright, 2009b, Bandeira 等人, 2022年, Gamarnik 和 Zadik, 2022年, Wainwright，2009a, Cai和Wang，2011]
SLR（本文工作）		$\Theta\left(\frac{k(\text{SNR}+1)}{\text{SNR}} \text{对数 } p\right)$	CORR $(n \geq \frac{8k\ (\text{SNR}+1)}{\beta_{\min}^2 \text{SNR}} \text{对数 } 2p)$

表1：对取值于 $\{-1, 0, 1\}$ 的信号所作贡献的总结。

统计查询等算法类别的失效 [Diakonikolas 等人，2019], 局部 [Linial，1992，Gamarnik 和 Sudan，2017] 以及消息传递算法 [Zdeborova 和 Krzakala，2016年，Krzakala 等人，2007], 以及从植入团等典型“困难”问题的变体进行归约[Berthet 和 Rigollet，2013，Brennan 和 Bresler，2020b]。

值得注意的是，高维 MSLR 问题引起了广泛关注，因为已有研究表明，SB-MSLR 这一特例存在 k-至-k^2 统计-计算差距，我们将其更精确地定义为 $\frac{k}{\text{SNR}^2}$至-$\frac{k^2}{\text{SNR}^2}$ 差距。这一发现源于对植入团的平均情形归约 [Brennan 和 Bresler，2020b] 以及统计查询模型 [Fan 等人，2018] 的研究。Fan 等人注意到，已知的 SB-MSLR 多项式时间算法均无法在样本复杂度低于 $\tilde{\Theta}(k^2/\text{信噪比}^2)$ 时取得成功；据此，他们 [2018] 推导出了相关检测问题的信息论极限和计算极限的下界。具体而言，他们证明，信息论最小样本复杂度为 $n = \tilde{\Theta}(k/\text{信噪比}^2)$，而统计查询算法（以及推测中的多项式时间算法）已被证明在所有低于更大阈值 $n = \tilde{o}\left(k^2/\text{信噪比}^2\right)$ 的样本复杂度下都会失效。这与文献中许多现有算法的失效阈值在数量级上相符，不过尚未严格证明计算下界是紧的。

类似地，Brennan 和 Bresler [2020b] 证明，我们在本文中考察的相关检测问题（SB-MSLR − D）在一种包含于样本复杂度 $n = o\left(k^2/\text{SNR}^2\right)$ 的情形下，可归约为一种 PC 检测问题变体，称为“Secret Leakage PC”。植入团的检测版本可以表述为：判断一个大小为 k 的团是否被人为“植入”到一个大小为 n 的 Erd¨os-R´enyi 图中。对于$k = \Omega(\log n)$，该问题可通过穷举搜索解决。植入团猜想认为，当 $k = o(\sqrt{n})$ 时，不存在能在多项式时间内解决 PC 的算法。关于 PC 猜想，有多种证据来源，参见 Feldman 等人。

[2013], Barak 等人 [2016], Brennan 和 Bresler [2020b] 以及其中的参考文献。

上述结果为 $\frac{k}{\mathrm{SNR}^2}$至$\frac{k^2}{\mathrm{SNR}^2}$ 统计-计算差距提供了证据，这一差距存在于 SBSB-MSLR的信息论极限与计算极限之间。更广泛地说，Brennan 和 Bresler [2020的] 工作迈出了一步，有助于理解高维统计中 k-至-k^2 差距的普遍性。该工作表明，学习具有 k-稀疏均值的混合分布时，高效算法至少需要 $\tilde{\Omega}(k^2)$的样本复杂度。在定理2.2中，我们强化了现有的 SB-MSLR 计算下界，揭示了更广泛的 $\frac{k}{\mathrm{SNR}^2}$-至-$\frac{k^2}{\mathrm{SNR}^2}$差距；与早期的下界不同，这表明即使在无噪声区域（信噪比 $= \infty$），也存在显著的计算障碍。

1.3 低次方法

低次方法是一种用于获得假设检验问题复杂度下界的框架，源自对平方和层级的研究 [Barak 等人, 2016年, Hopkins 等人, 2017年, Hopkins 和 Steurer, 2017年, Hopkins, 2018]。低次方法归结为严格排除利用输入的低次多项式函数解决给定假设检验问题的可能性。考虑简单二元假设检验的情形：我们希望区分定义在 $\mathbb{R}^N$上的两个分布 $\mathbb{P}_N$ 和 $\mathbb{Q}_N$，其中 N 是（可能不断增长的）问题规模。给定一个从 $\mathbb{P}_N$ 或 $\mathbb{Q}_N$中抽取的样本 $\boldsymbol{x}$，目标是通过假设检验判断 $\boldsymbol{x}$ 源自前者还是后者。在我们的 MSLR 设定中，可以将 $N = np + n$ 视为数据的总维度（$\boldsymbol{X}$, $\boldsymbol{y}$），并注意到 $\log N = O(\log p)$。我们考虑检验成功的两种定义：

- 强检测/区分：当 $p \to \infty$ 时，检验以概率 $1 - o(1)$ 成功。
- 弱检测/区分：检验以某个常数 $\epsilon > 0$ 的概率 $\frac{1}{2} + \epsilon$ 成功。

D 次多项式算法表示一列（可能是随机的）多元多项式 $g_N : \mathbb{R}^N \to \mathbb{R}$，其次数为 D；而 $f_{\leq D}$ we 表示将函数 f 正交投影到次数为 D 的多项式空间上。在过去十年间，研究已证实，对于一大类高维检验问题（包括稀疏主成分分析、植入团、社群检测等），次数为 $O(\log p)$ 的多项式算法类，其能力与已知最优多项式时间算法相当 [Bandeira 等人，2020年，Ding 等人，2023, Hopkins，2018年，Hopkins 和 Steurer，2017年，Hopkins 等人，2017年，Kunisky 等人，2022]。这可以用以下猜想形式化。

猜想 1.2（低次猜想 Coja-Oghlan 等人 [2022], Hopkins [2018]）。定义 $\mathbb{P}_N$ 与 $\mathbb{Q}_N$ 之间的卡方散度为 $\chi^2(\mathbb{P}_N \| \mathbb{Q}_N) := \mathbb{E}_{\boldsymbol{x} \sim \mathbb{Q}_N} \left(\frac{d\mathbb{P}_N}{d\mathbb{Q}_N}(\boldsymbol{x}) \right)^2 - 1$，并令 $\chi^2_{\leq D}(\mathbb{P}_N \| \mathbb{Q}_N)$ 表示~~其在次~~数为 D 的多项式空间上的投影。

- 如果对于某个 $D = \omega(\log N)$，$\chi^2_{\leq D}(\mathbb{P}_N \| \mathbb{Q}_N) = O(1)$，则强检测不存在多项式时间$\tilde{}$算法，并且还需要 $\exp(\Omega(D))$ 的运行时间。
- 如果对于某个 $D = \omega(\log N)$，$\chi^2_{\leq D}(\mathbb{P}_N \| \mathbb{Q}_N) = o(1)$，则弱检测不存在多项式时间$\tilde{}$算法，并且还需要 $\exp(\Omega(D))$ 的运行时间。

许多最先进的算法都可以用低次多项式近似，因此可以通过上述形式的低次下界严格排除，其中包括重要的谱方法类别（见Kunisky 等人 [2022]的定理 4.4），以及所有统计查询算法 [Brennan 等人， 2021]。近期研究还证明了低次

多项式算法与统计物理学衍生出的成熟算法解法，在某些问题类别中是等价的 [Bandeira 等人, 2022年, Montanari 和 Wein, 2022]。尽管尚未证明次数为$O(\log p)$ 的多项式涵盖所有多项式时间算法，但这类多项式在假设检验中的成功，往往预示着一般多项式时间算法也能成功。鉴于此，我们旨在通过证明某个相关检测问题的低次下界，为 MSLR 和SLR 中的计算困难性提供具体证据。该检测问题具有如下形式 $\chi^2_{\leq D}(\mathbb{P}_N \| \mathbb{Q}_N) = O(1)$，之后可将其归约为恢复问题。有关低次方法的更多背景信息，请参见附录A。

2 主要结果

2.1 MSLR 的下界

我们首先定义 MSLR 的一个检测变体：给定 $(\boldsymbol{X}, \boldsymbol{y})$，目标是在两个假设之间进行区分：一个假设认为数据符合 MSLR 模型，另一个假设认为 $\boldsymbol{X}$ 和 $\boldsymbol{y}$ 相互独立。

定义 2.1（检测变体 MSLR − D）。对于 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$、 $\sigma > 0$ 和 $\boldsymbol{w} \in \mathbb{R}^n$，考虑以下假设检验问题：

$$\mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y}) : \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X} \\ \sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} + 1} \cdot \boldsymbol{w} \end{bmatrix}$$
$$\mathbb{P}(\boldsymbol{X}, \boldsymbol{y}) : \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X} \\ \frac{1}{\sigma}\boldsymbol{X}\boldsymbol{\beta}_1 \odot \boldsymbol{z} + \frac{1}{\sigma}\boldsymbol{X}\boldsymbol{\beta}_2 \odot (1-\boldsymbol{z}) + \boldsymbol{w} \end{bmatrix}$$

其中 $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \sim \mathcal{P}_{\|\boldsymbol{\beta}\|_2}(\mathcal{D})$，且 $X_{i,\ j}$ 独立同分布$\sim \mathcal{N}(0,1)$， w_i 独立同分布$\sim \mathcal{N}(0,1)$， z_i 独立同分布$\sim$ 伯努利分布(ϕ)。任务是构造一个函数 f，使其能够有力地区分 $\mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})$ 与 $\mathbb{P}(\boldsymbol{X}, \boldsymbol{y})$。

注意， $\mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})$ 和 $\mathbb{P}(\boldsymbol{X}, \boldsymbol{y})$ 的边缘分布相同，因此可以排除仅对 $\boldsymbol{y}$ 的矩进行阈值处理而忽略 $\boldsymbol{X}$ 的解。SB-MSLR 的相应检测变体记为 SB-MSLR − D，其定义与参数范围为 (1) 中所给 SB-MSLR 的 MSLR − D 类似。根据这一表述，我们得到关于 SB-MSLR − D 的以下困难性结果。证明见附录 B.3。

定理 2.2（SB-MSLR − D 的低次下界）。考虑 SB-MSLR − D 的设定，其中$\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \sim \mathcal{P}_{\|\boldsymbol{\beta}\|_2}(\mathcal{D})$，且信号具有有界幅度（$\boldsymbol{\beta}_{min} = \Theta(\|\boldsymbol{\beta}\|_\infty)$）。对于样本量 n，其中$n = \omega(\max\{k, \log\ p\})$ 且 $n = o\left(k^2 \cdot \frac{(SNR+1)^2}{SNR^2} \cdot 1)\log p\right)$，猜想~~1.2意味着，任~~何随机化算法要在$k = o(\sqrt{p})$的条件下求解 SB-MSLR − D，都需要 $\exp\left(\tilde{\Omega}\left(\min\left\{k^2 \cdot \frac{(SNR+1)^2}{nSNR^2}\text{的运行时间}, n\right\}\right)\right)$。

定理 2.2 是我们关于低次困难性的主要结果。我们关心三种情形，下面用 $n_{\text{alg}}^{\text{SB-MSLR}} := k^2 \cdot \frac{(\text{SNR}+1)^2}{\text{SNR}^2}$ 来描述。首先，若 $n = \Omega\left(n_{\text{alg}}^{\text{SB-MSLR}} \log\ p\right)$，定理 2.2 中运行时间的下界等于 $e^{O(\log p)}$，因此无法排除多项式时间解。否则，定理 2.2（通过猜想 1.2）意味着样本量 n 与超多项式（但次指数）运行时间 $\exp\left(\tilde{\Omega}\left(n_{\text{alg}}^{\text{SB-MSLR}}/n\right)\right)$之间存在平滑的权衡关系，其中$n = \omega((n_{\text{alg}}^{\text{SB-MSLR}})^{\frac{1}{2}})$；这令人联想到稀疏主成分分析中类似的权衡关系 [Ding 等人, 2023]。在第三种情形下，即 $n = o((n_{\text{alg}}^{\text{SB-MSLR}})^{\frac{1}{2}})$，定理 2.2 表明需要 $e^{\tilde{\Omega}(n)}$ 运行时间。因此

存在三种不同的计算机制，具体取决于样本复杂度 n：第一种允许多项式时间解，第二种形成信息与计算之间平滑的反向权衡关系，最后一种则意味着运行时间会随着样本量增加而呈指数级增长。这扩展了 [Brennan 和 Bresler、2020b、Fan 等人的结果 2018] ，这些结果指出， $n = \tilde{o}(k^2/\text{信噪比}^2)$样本区间对SB-MSLR $-$ D 中具有 $\{-1,0,1\}^p$信号的情形带来了与统计查询和隐 planted clique 相关的算法障碍；请注意，在无噪声设定下，以及在信噪比 $=\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} = \Theta(k)$的自然设定下，阶为信噪比^2 的下界都是空泛的。

Fan 等人 [2018] 证明，SB-MSLR $-$ D 的信息论最小样本复杂度为 $n = \tilde{\Theta}(k/\text{信噪比}^2)$，当信噪比 $= \omega(\sqrt{k})$ 时，这一结果毫无意义。然而，对于一大类信噪比，相关的稀疏相位恢复（SPR）检测问题的信息论最小样本复杂度已知为 $k\log p$ 量级（例如，参见 Cai 等人 [2016] 的定理 3.2，以及 Lecu´e 和 Mendelson [2015]的第 6.1 节）。通过从 SLR 到 SB-MSLR，再到 SPR 的直接归约，可以证明，SB-MSLR 中检测的信息论样本复杂度介于 $^k\log(p/k)\log(1+\text{SNR})$ 和 $k\log p$ 之间。由此，定理 2.2 证实，在广泛的信噪比范围内，SB-MSLR $-$ D 存在统计-计算差距，其量级至少为 k。

我们强调，定理 2.2 严格排除了输入的次数至多为 $O(\log p)$ 的解析多项式成功的可能性，其中包括谱方法。当信号为 $k = o(\sqrt{p}$ 稀疏时， k-sparse（参见 [Brennan 和 Bresler、2020b、Fan 等人、2018年、Ding 等人，以及 2023] 其中的参考文献）时，这一假设通常是检测下界中的标准假设；有时也可以通过对某个坏事件 [Bandeira 等人、 2022].

注释 2.3。 为便于解释，我们在定理 2.2 中加入了有界幅度假设。对 $\|\boldsymbol{\beta}\|_\infty$ 的依赖关系可以通过将定理 k^2 中的 2.2 替换为$\|\boldsymbol{\beta}\|_2^4/\|\boldsymbol{\beta}\|_\infty^4$来明确写出。我们认为对 $\|\boldsymbol{\beta}\|_\infty$ 的依赖是证明技术带来的假象；参见附录B.3。

通过附录 C 中的定理 C.3，我们给出了从 SB-MSLR $-$ D 到 PSB-MSLR 中精确支撑集恢复的多项式时间归约，适用于 $\{-1,0,1\}^p$ 且信噪比 $= \omega(1)$ 的信号，从而将定理 2.2 的困难性转移到这一情形。在附录 C.3 中，我们给出了从 SB-MSLR $-$ D 到稀疏相位恢复（SPR）中精确支撑集恢复和检测的多项式时间归约，适用于非零条目属于 $\{-1,0,1\}^p$ 的信号，从而将定理 2.2 的困难性结果转移到 SPR。这为如下猜想提供了新的严格证据：对于样本量 $n = \tilde{o}(k^2)$，SPR 在计算上不可行 [Wu and Rebeschini, 2021, Li 等人, 2022, Brennan 和 Bresler, 2020b]。

2.2 MSLR的算法

我们分别将 $\boldsymbol{\beta}_1$、$\boldsymbol{\beta}_2$ 的支撑集记为 $\mathcal{S}_1$、$\mathcal{S}_2$。注意， $|\mathcal{S}_1| = |\mathcal{S}_2| = k$。定义以下量：

$$\langle\boldsymbol{\beta}\rangle_{\text{min}}^2 := \min_{j\in\mathcal{S}_1\cup\mathcal{S}_2}(\phi\boldsymbol{\beta}_{1,j} + (1-\phi)\boldsymbol{\beta}_{2,j})^2,$$
$$\langle\boldsymbol{\beta}\rangle_{>0}^2 := \min_{\substack{j\in\mathcal{S}_1\cup\mathcal{S}_2 \\ (\phi\boldsymbol{\beta}_{1,j}+(1-\phi)\boldsymbol{\beta}_{2,j})>0}}(\phi\boldsymbol{\beta}_{1,j} + (1-\phi)\boldsymbol{\beta}_{2,j})^2.$$

注意，在 SB-MSLR 区域之外，对于（$\boldsymbol{\beta}_1$， $\boldsymbol{\beta}_2$）， $\langle\boldsymbol{\beta}\rangle_{>0}^2 > 0 \sim \mathcal{P}_{\|\boldsymbol{\beta}\|_2}(\mathcal{D})$。另外，回顾$\boldsymbol{\beta}_{\text{min}} = \min\{|\beta| \mid \beta\in\mathcal{D}\} > 0$。

定义 2.4（CORR）。令 CORR 为一种算法，根据 $\widehat{\mathcal{S}_1\cup\mathcal{S}_2} = \left\{j\in[p]: \left|\langle\boldsymbol{X}_j, \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|_2}\rangle\right| \geq \tau\right\}$，输出 $\mathcal{S}_1\cup\mathcal{S}_2$ $\boldsymbol{\beta}_1$、 $\boldsymbol{\beta}_2$ 的联合支撑集的估计，其中 $\tau = \sqrt{2(1+\frac{\epsilon}{2})\log 2p}$，且 $\epsilon\in$ 属于 (0, 1)。

定理 2.5（CORR 在 SB-MSLR 之外的 MSLR – D 上的成功性）。考虑 MSLR – D \ SB-MSLR 的一般情形，其中 $(\beta_1, \beta_2) \sim \mathcal{P}_{\|\beta\|_2}(\mathcal{D})$。设 $\epsilon \in (0, 1)$ 为 CORR 中使用的参数。那么，只要

$$n \geq \frac{32(1+\epsilon)}{\min\{\phi^2\beta_{min}^2, (1-\phi)^2\beta_{min}^2, \langle\beta\rangle_{>0}^2\}} \frac{\|\beta\|_2^2 (SNR+1)}{SNR} \log 2p,$$

CORR 算法在 MSLR – D \ SB-MSLR 中实现强检测。

定理 2.5 的证明见附录 D.2。在 $\|\beta\|_2^2$ is of order$k$ 的自然设定下，该定理意味着，除 SB-MSLR 区域外，CORR 算法能够解决 MSLR – D 问题，所需样本数是定理 2.2 中低次下界所隐含样本数的平方根（忽略对数因子）。事实上，定理 2.5 中的样本复杂度与更简单的 SLR – D 问题的最优样本复杂度相符；见下文定理 2.8。该定理有效地量化了以 SLR 的样本复杂度解决 MSLR 的能力。定理 2.5 的证明也适用于更一般的情形，即 $\langle\beta\rangle_{>0}^2 > 0$ 和信号范数 $\|\beta_1\|_2, \|\beta_2\|_2$ 不受必须相等的约束。对于非零项的信号先验若相对于勒贝格测度绝对连续，则事件 $\{\langle\beta\rangle_{>0}^2 > 0\}$ 的测度为一，因为 $\phi\beta_1 + (1-\phi)\beta_2 \neq 0$ 几乎必然成立。

定理 2.6（$\langle\beta\rangle_{\min}^2 > 0$下 MSLR 中 CORR 恢复成功）。考虑 MSLR 的一般设定，其中为 $\sigma = 0, \phi \neq 1/2$（无噪声），或 $\phi = 1/2, SNR = \Omega(k)$（平衡情形）。令 $(\beta_1, \beta_2) \sim \mathcal{P}_{\|\beta\|_2}(\mathcal{D})$。令 $\epsilon \in (0, 1)$ 为 CORR 中使用的参数，并且

$$n \geq \frac{32(1+\epsilon)}{\min\{\phi^2\beta_{min}^2, (1-\phi)^2\beta_{min}^2, \langle\beta\rangle_{min}^2\}} \frac{\|\beta\|_2^2 (SNR+1)}{SNR} \log 2p.$$

那么，存在一种算法，与 CORR 结合后，以至少 $1 - c_1(\frac{k}{p} + ke^{-c_2 n} + \frac{k}{n} + \frac{1}{p^{c_2}})$ 的概率精确恢复 β_1 和 β_2（允许重新标记），其中 $c_1, c_2 > 0$ 为常数。

定理 2.6的证明见附录 D.3，首先使用 CORR 进行支撑集恢复，随后采用已有的恢复算法处理稠密（$k/p = \Theta(1)$）混合线性回归的无噪声和平衡情形 [Yi 等人， 2014年， Chen 等人， 2014]。在条件 $\langle\beta\rangle_{\min}^2 > 0$下，该条件比 SB-MSLR略为严格，定理2.6 给出的样本复杂度与 SLR的样本复杂度处于同一数量级。我们注意到，对于非零项的信号先验若相对于勒贝格测度绝对连续，则事件 $\{\langle\beta\rangle_{\min}^2 > 0\}$ 几乎处处成立。我们强调，无噪声情形可表述为一种采用独立高斯设计的压缩感知混合变体 [Yu 和 Sapiro， 2011]。

注 2.7。 定理 2.6 仅讨论无噪声和平衡情形，这是因为现有算法在稠密情形下所能提供的保证所致；实验表明，这些算法在远超这些范围的情形下也能成功 [Yi 等人，2014年，Chen 等人， 2014]。

2.3 稀疏线性回归（SLR）的下界

我们按照定义 SLR 的检测变体，称为 SLR – D，定义为定义 2.1 并附加约束$\beta_1 = \beta_2$。附录 SLR – D 证明了以下下界，见附录 B.2。

定理 2.8（SLR – D 的低次下界）。考虑 SLR – D 的设定（ $\beta_1 = \beta_2$ 下的定义 2.1），其中 $\beta \sim \mathcal{P}_{\|\beta\|_2}(\mathcal{D})$。对于 $n = \omega(\log p)$ 且 $n \leq (1-\epsilon)(1-2\theta)\frac{\|\beta\|_2^2}{\|\beta\|_\infty^2}(SNR+1)\, SNR \log p$，对于任意 $\epsilon \in (0, 1)$, ~~猜想 1~~.2 意味着，任何随机化算法都需要 $e^{\Omega(n)}$的运行时间，才能在 $k = O(p^\theta) \leq \sqrt{p}$ 且 $\theta \in (0, 1/2]$ 的情况下求解 SLR – D。

在 $\|\boldsymbol{\beta}\|_2^2$ 为 k 的自然设定中，且各条目具有有界幅度时， n 的低次下界为 $k\,(\mathrm{SNR}+1)\,\mathrm{SNR}\,\log p$。这与现有 SLR 下界的数量级相匹配, ~~包括~~ Bandeira 等人、2022年、Gamarnik 和 Zadik、2022], 的结果，但其优势在于对所有 信噪比 设定以及稀疏信号上的一般先验 $\boldsymbol{\beta}$均有效。我们认为，下界对 $\|\boldsymbol{\beta}\|_\infty$ 的依赖是证明技术造成的假象；详见附录 B.2。

从 SLR $-$ D 到 SLR 的归约与从 SB-MSLR $-$ D 到 SB-MSLR的归约类似，后者见附录C。

2.4 SLR 的算法 SLR

定理 2.9。考虑参数为 $\boldsymbol{\beta} \sim \mathcal{P}_{\|\boldsymbol{\beta}\|_2}(\mathcal{D})$ 的 SLR 设置。设 $\epsilon \in (0, 1)$ 为 CORR 中使用的参数。那么对于 $n \geq \frac{8(1+\epsilon}{\beta_{\min}^2}\|\boldsymbol{\beta}\|_2^2\frac{(SNR+1)}{SNR})\log ~~2p,~~$ CORR 能够解决 SLR $-$ D 中的强检测问题。

ˆ 接下来，我们考虑 CORR 的一个轻微变体，该变体恢复 $\boldsymbol{\beta}$ 的带符号支撑集。它产生 $\boldsymbol{\beta}$，其条目由下式给出

$$\hat{\boldsymbol{\beta}}_j = \mathbb{I}\left\{\left|\frac{\langle \boldsymbol{X}_j, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \geq \sqrt{2(1+\epsilon/2)\log 2p}\right\}\operatorname{sign}\left(\frac{\langle \boldsymbol{X}_j, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right), \quad \text{for } j \in [p], \tag{3}$$

其中，当 $x > 0$ 时，sign(x) 等于 1；当 $x < 0$ 时，等于 -1；当 $x = 0$ 时，等于 0。

定理 2.10。 考虑参数为 SLR 的设置，其中 $\boldsymbol{\beta} \sim \mathcal{P}_{\|\boldsymbol{\beta}\|_2}(\mathcal{D})$。设 $\epsilon \in (0, 1)$ 为上述 CORR 变体中使用的参数。那么对于 $n \geq \frac{8(1+\epsilon}{\beta_{\min}^2}\|\boldsymbol{\beta}\|_2^2\frac{(SNR+1)}{SNR})\log 2p$, ~~式ˆ $\boldsymbol{\beta}$~~ 中的向量 (3) 以至少 $\boldsymbol{\beta}$ 的概率等于 $1 - (\frac{k}{p} + 2ke^{-c_2 n} + \frac{1}{p^{c_2}})$ 的带符号支撑集，其中 $c_2 > 0$为某个常数。

定理 2.9 和定理 2.10 的证明见附录 D.4。CORR 成功所需的样本复杂度与定理 2.8 中的低次下界仅相差常数，这严格证明了 CORR 在低次多项式算法中的阶最优性，包括运行 O 次迭代的谱方法，其中迭代次数为 $(\log p)$，且适用于所有信噪比范围。这些可达到的样本复杂度也与先前工作的结果相符 [Wainwright，2009b、Bandeira 等人，2022、Gamarnik 和 Zadik，2022、Cai 和 Wang，2011、Donoho 和 Tanner， 2010], ，重要的扩展在于，这些结果适用于所有信噪比缩放情形和一般的稀疏信号先验。

3 证明思路

低次下界定理 2.2 的证明归结为证明，对于定理指定的范围内的 n，在 MSLR $-$ D（定义 2.1）中， $\chi^2_{\leq D}(\mathbb{P}(\boldsymbol{X}, \boldsymbol{y})\|\mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})) = O(1)$，其中 $\boldsymbol{\beta}_1 = -\boldsymbol{\beta}_2$ 且 $\phi = 1/2$。我们将猜想 1.2 中 $\chi^2_{\leq D}(\mathbb{P}(\boldsymbol{X}, \boldsymbol{y})\|\mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y}))$ 的表达式改写为关于数据 $(\boldsymbol{X}, \boldsymbol{y})$ 的多元厄米多项式形式，次数至多为 D。对于 $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_{np+n}]$，在 $\alpha_i \in \mathbb{N}$ 的条件下， $\boldsymbol{\alpha}$ 阶的归一化厄米多项式记作 $\frac{H_{\boldsymbol{\alpha}}(\boldsymbol{X},\boldsymbol{y})}{\sqrt{\boldsymbol{\alpha}!}}$。该多项式的精确定义见附录 B，但我们将用到的关键事实是， $\left\{\frac{\tilde{H}_{\boldsymbol{\alpha}}}{\sqrt{\boldsymbol{\alpha}!}}\right\}$ 关于定义 2.1 中的零分布构成一个正交归一系统（见附录 B 中的命题 B.5）。利用这一点，我们得到

$$\chi^2_{\leq D}(\mathbb{P}(\boldsymbol{X}, \boldsymbol{y})\|\mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})) + 1 = \mathbb{E}_{(\boldsymbol{X},\boldsymbol{y})\sim\mathbb{P}(\boldsymbol{X})\otimes\mathbb{P}(\boldsymbol{y})}\left(\frac{d\mathbb{P}(\boldsymbol{X}, \boldsymbol{y})}{d(\mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y}))}\right)^2_{\leq D}$$

$$= \sum_{0\le|\boldsymbol{\alpha}|\le D} \frac{1}{\boldsymbol{\alpha}!} \mathop{\mathbb{E}}_{\mathbb{P}(\boldsymbol{X})\otimes\mathbb{P}(\boldsymbol{y})} \left[\frac{d\mathbb{P}(\boldsymbol{X},\boldsymbol{y})}{d(\mathbb{P}(\boldsymbol{X})\otimes\mathbb{P}(\boldsymbol{y}))}\tilde{H}_{\boldsymbol{\alpha}}(\boldsymbol{X},\boldsymbol{y})\right]^2$$
$$= \sum_{0\le|\boldsymbol{\alpha}|\le D} \frac{1}{\boldsymbol{\alpha}!} \mathop{\mathbb{E}}_{\mathbb{P}(\boldsymbol{X},\boldsymbol{y})} \left[\tilde{H}_{\boldsymbol{\alpha}}(\boldsymbol{X},\boldsymbol{y})\right]^2, \tag{4}$$

其中 $\boldsymbol{\alpha}! = \prod_i \alpha_i!$。证明的关键步骤是利用厄米多项式恒等式和多项式定理进行变换，从而对式 (4) 作上界估计，得到关于重叠 $\langle\boldsymbol{\beta}_1^{(1)},\boldsymbol{\beta}_1^{(2)}\rangle$ 的 D 阶矩的加权和，其中 $\boldsymbol{\beta}_1^{(1)},\boldsymbol{\beta}_1^{(2)}$ 是信号 $\boldsymbol{\beta}_1$ 的两个独立同分布副本（见引理 B.11）。这些 D 阶矩中的每一个都可以在 $k \le \sqrt{p}$ 时得到界，因此整个关于$D \simeq$的和中，最小的 $\left\{\frac{k^2}{n\mathtt{SNR}^2}(\mathtt{SNR}+1)^2\right.$ 项可以收敛，从而得到所需结果。定理 2.8 中 SLR 的情形类似，~~但由于简~~化条件 $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$，我们可以令 $D \simeq n$，同时仍使该和收敛，由此得到 SLR 与 SB-MSLR 下界的关键差异。

从检测到恢复的归约 我们遵循 Brennan 和 Bresler [2020b] 概述的平均情形归约流程。我们通过构造针对 k 稀疏信号在 $\{-1,0,1\}^p$ 中的平均情形归约，将 SB-MSLR − D 的计算困难性转移到 SB-MSLR 中的恢复问题。将定理 2.2 的参数范围称为“临界”参数范围。给定临界范围内的任意参数序列 $\mathcal{P}$，我们构造另一个具有以下性质的临界范围参数序列 $\mathcal{P}'$：如果存在一个随机多项式时间算法 $\mathcal{A}'$，能够在参数缩放为 $\mathcal{P}'$ 时求解 PSB-MSLR 中的精确恢复问题，那么我们就能构造一个参数缩放为 $\mathcal{P}$ 的随机多项式时间算法来求解 SB-MSLR − D。这将反过来与定理 2.2 矛盾，从而表明在临界范围内，PSB-MSLR 中的精确恢复具有计算困难性。我们首先在引理 C.1 中给出从 SB-MSLR − D 到 SB-MSLR 中精确恢复的平均情形归约，然后在定理 C.3 中将 SB-MSLR 中的精确恢复归约到 PSB-MSLR 中的精确恢复。

MSLR 的 CORR 算法。对于 MSLR，定理 2.5 和定理 2.6 的证明关键依赖于定理 D.1，该定理表明，如果 n 满足定理 2.6 中的条件，CORR 就能恢复信号的联合支撑 $(\mathcal{S}_1 \cup \mathcal{S}_2)$。为证明定理 D.1，我们分三种情况分析量 $u_j := \frac{\langle \boldsymbol{X}_j, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}$。当 $j \in (\mathcal{S}_1 \cup \mathcal{S}_2)^{\complement}$ 时，由于 $\boldsymbol{X}_j$ 与 $\boldsymbol{y}$ 独立，对于~~$j \in [p]$~~，我们有 u_j 独立同分布$\sim \mathcal{N}(0,1)$。此时 $\max_{j\in[p]} u_j$ 的典型值为 $\sqrt{2\log p}$，并且我们可以利用标准集中界来限制假阳性的概率，详细推导见引理 D.9。当 $j \in \mathcal{S}_1 \cap \mathcal{S}_2$ 时，我们证明，在给定 $\boldsymbol{y}$ 的条件下，$\boldsymbol{z}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, u_j$ 服从均值为

$$\mathbb{E}\left[u_j \mid \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2\right] = \frac{\|\boldsymbol{y}_{\{\boldsymbol{z}=1\}}\|_2^2 \boldsymbol{\beta}_{1,j} + \|\boldsymbol{y}_{\{\boldsymbol{z}=0\}}\|_2^2 \boldsymbol{\beta}_{2,j}}{\|\boldsymbol{y}\|_2 (\|\boldsymbol{\beta}\|_2^2 + \sigma^2)}, \tag{5}$$

且方差小于 1（引理 D.4）。这里，$\|\boldsymbol{y}_{\{\boldsymbol{z}=1\}}\|_2$ 表示由条目 $(y_i 1_{\{z_i=1\}})_{i\in[n]}$构成的向量的范数。对于较大的 n、p 和 $j \in [p]$，上述条件均值的典型值为 $\sqrt{\frac{n}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}}(\phi\boldsymbol{\beta}_{1,j} + (1-\phi)\boldsymbol{\beta}_{2,j})$，对于

$$n \ge \frac{2(1+\epsilon)}{\langle\boldsymbol{\beta}\rangle_{\mathtt{min}}^2}(\|\boldsymbol{\beta}\|_2^2 + \sigma^2)\log 2p \simeq (1+\epsilon)\frac{k(\mathtt{SNR}+1)}{\langle\boldsymbol{\beta}\rangle_{\mathtt{min}}^2 \mathtt{SNR}}\log 2p.$$

剩余情况 $j \in \mathcal{S}_1 \Delta \mathcal{S}_2$ 类似，条件均值可通过令$\boldsymbol{\beta}_{2,\ j} = 0$ 在（5）中得到。定理2.9和2.10中关于SLR的结果也遵循类似的推理，且 $\mathbb{E}\left[u_j \mid \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2\right] = \frac{\boldsymbol{\beta}_j \|\boldsymbol{y}\|_2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}$。

4 讨论

在这项工作中，我们通过低次多项式方法，严格刻画了混合稀疏线性回归（MSLR）的计算困难性。我们证明，在高度对称的 SB-MSLR 区域中，随机多项式时间算法无法以样本复杂度为 $n = \tilde{o}\left(\frac{k^2}{\left(\frac{\mathtt{SNR}+1}{\mathtt{SNR}^2}\right)^2}\right)$ 解决一个相关的检测问题，这揭示了至少为 k 的统计-计算差距。然而，在 SB-MSLR 区域之外，一个简单的多项式时间算法 CORR 能以最小样本复杂度成功完成检测。

我们注意到，即使在无噪声（SB-MSLR 信噪比 $= \infty$）情形下，我们针对 信噪比 $= \infty$ 所得到的低阶统计-计算差距依然存在。近期的研究发现表明，统计到计算的差距并非总能在无噪声设定下成立。例如，在学习奇偶函数时，高斯消元等“脆弱”算法可以“打破”统计-计算差距 [Zadik 等人， 2022]。此外，Zadik 等人 [2022] 最近还发现，源自密码学的 LLL 算法族能够打破某些无噪声聚类问题中预测的统计-计算差距。此外，在 Zadik [2021], 最近的一次演讲中，研究者给出了一种基于格的算法的证明概要，该算法能够在 $\boldsymbol{\beta} \in \mathbb{R}^p$ 稠密无噪声相位恢复中，使用 $p+1$ 次测量进行恢复——这打破了稠密相位恢复中一个被猜想存在的统计-计算差距，但该算法未能捕捉MSLR中存在的稀疏问题结构。据我们所知，针对 SB-MSLR 和稀疏相位恢复的无噪声推断，仍无法在少于数量级为 $k^2 \log p$ 个样本的情况下实现，其中 k-稀疏 $\boldsymbol{\beta} \in \{-1, 0, 1\}^p$。如果能够实现这一点，将是一项有趣且新颖的贡献。

致谢

GA 获得了剑桥信托基金奖学金以及 Invenia Labs 的资助。GA 感谢 Alex Wein、Ilias Zadik、Afonso Bandeira、Nicol`o Grometto 和 Daniil Dmitriev，就稀疏线性回归和相位恢复中的统计-计算差距进行了有益的讨论。

参考文献

Ery Arias-Castro、Emmanuel J. Cand`es 和 Yaniv Plan。稀疏备择假设下的全局检验：方差分析、多重比较与高阶检验。《统计年刊》，39(5):2533–2556，2011。1 Gabriel Arpino。稀疏高维线性回归的计算困难性，2021。苏黎世联邦理工学院硕士论文。5，19 Martin Azizyan、Aarti Singh 和 Larry Wasserman。稀疏均值分离下高维高斯混合模型的极小极大理论。载于《神经信息处理系统进展》，第26卷，2013。1 Afonso S. Bandeira、Dmitriy Kunisky 和 Alexander S. Wein。约束主成分分析问题界的认证所面临的计算困难性。载于第11届理论计算机科学创新会议（ITCS 2020），第151卷，第78:1–78:29页，2020。4，6 Afonso S. Bandeira、Ahmed El Alaoui、Samuel B. Hopkins、Tselil Schramm、Alexander S. Wein 和 Ilias Zadik。高维统计中的 Franz-Parisi 判据与计算权衡。载于《神经信息处理系统进展》，2022。2，4，5，7，8，10，19 Boaz Barak、Samuel B. Hopkins、Jonathan Kelner、Pravesh Kothari、Ankur Moitra 和 Aaron Potechin。植入团问题的近乎紧致平方和下界。载于《第57届 IEEE 计算机科学基础年度研讨会论文集》，FOCS 2016，第428–437页，2016年12月。4，6 Adarsh Barik 和 Jean Honorio。具有保证的稀疏混合线性回归：通过不变凸松弛驾驭棘手问题。载于国际机器学习会议，第162卷，第1627–1646页，2022。1 Quentin Berthet 和 Philippe Rigollet。稀疏主成分检测的复杂性理论下界。载于第26届学习理论年度会议论文集，第1046–1066页，2013。5 St´ephane Boucheron、G´abor Lugosi 和 Pascal Massart。《集中不等式：独立性的非渐近理论》。牛津大学出版社，2013。48，49，55 Matthew Brennan 和 Guy Bresler。学习稀疏混合模型、稳健估计及半随机对手的平均情形下界。技术报告，2020a。arXiv:1908.06130。6 Matthew Brennan 和 Guy Bresler。秘密泄露导致的可约性与统计-计算差距。载于第三十三届学习理论会议论文集，第648–847页，2020b。1，2，3，5，6，8，11，31 Matthew S. Brennan、Guy Bresler、Sam Hopkins、Jerry Li 和 Tselil Schramm。统计查询算法与低次检验几乎等价。载于第三十四届学习理论会议论文集，第774–774页，2021。6 T. Tony Cai 和 Lie Wang。含噪稀疏信号恢复的正交匹配追踪。《IEEE信息论汇刊》，57(7):4680–4688，2011。5，10 T. Tony Cai、Xiaodong Li 和 Zongming Ma。通过阈值化 Wirtinger 流对含噪稀疏相位恢复实现最优收敛速率。《统计年刊》，2016年10月。出版方：数理统计学会。8 E.J. Candes 和 T. Tao。线性规划解码。《IEEE信息论汇刊》，51 (12):4203–4215，2005。2 Emmanuel J. Cand`es、Xiaodong Li 和 Mahdi Soltanolkotabi。通过 Wirtinger 流进行相位恢复：理论与算法。《IEEE信息论汇刊》，61(4):1985–2007，2015。38

Arun Tejasvi Chaganty 和 Percy Liang。《用于估计线性回归混合模型的谱专家》。载于《第30届国际机器学习会议论文集》，第1040–1048页，2013年5月。1 Sitan Chen、Jerry Li 和 Zhao Song。《通过傅里叶矩在亚指数时间内学习线性回归混合模型》。《第52届ACM计算理论年会论文集》，第587–600页，2020年。1 Yudong Chen、Xinyang Yi 和 Constantine Caramanis。《双成分混合回归的凸规划：极小极大最优速率》。载于《第27届学习理论会议论文集》，第560–604页，2014年。1, 9, 31, 50, 51, 54, 55 Amin Coja-Oghlan、Oliver Gebhard、Max Hahn-Klimroth、Alexander S. Wein 和 Ilias Zadik。《群体测试中的统计与计算相变》。载于《第35届学习理论会议论文集》，第4764–4781页，2022年6月。6, 18 Partha Deb 和 Ann M. Holmes。《行为健康保健的使用情况与费用估算：标准模型与有限混合模型的比较》。《健康经济学》，9(6):475–489，2000年9月。1 Ilias Diakonikolas、Weihao Kong 和 Alistair Stewart。《鲁棒线性回归的高效算法与下界》。载于《第三十届ACM-SIAM离散算法年会论文集》，第2745–2754页，美国，2019年1月。5 Yunzi Ding、Dmitriy Kunisky、Alexander S. Wein 和 Afonso S. Bandeira。《稀疏主成分分析的亚指数时间算法》。《计算数学基础》，2023。6, 7, 8 David L. Donoho 和 Jared Tanner。《随机投影超立方体和正交象限的面计数及其应用》。《离散与计算几何》，43(3):522–541，2010年4月。10 D.L. Donoho。《压缩感知》。《IEEE信息论汇刊》，52(4):1289–1306，2006年4月。2

Jianqing Fan、Han Liu、Zhaoran Wang 和 Zhuoran Yang。《异质性诅咒：稀疏混合模型与相位恢复中的计算障碍》。技术报告，arXiv，2018年8月。arXiv:1808.06996。1，2，3，5，8，52，54

Susana Faria 和 Gilda Soromenho。拟合线性回归混合模型。《统计计算与模拟杂志》，80(2):201–225，2010年。1

Vitaly Feldman、Elena Grigorescu、Lev Reyzin、Santosh Vempala 和 Ying Xiao。《统计算法与检测植入团的下界》。载于 第四十五届ACM计算理论研讨会论文集，第655–664页，2013年6月。5David Gamarnik 和 Madhu Sudan。《稀疏随机图上的局部算法的极限》。《概率年刊》，45(4):2353–2376，2017年。5David Gamarnik 和 Ilias Zadik。《二元系数高维回归：平方误差估计与相变》。载于 2017年学习理论会议论文集，第65卷，第948–953页，2017年6月。33David Gamarnik 和 Ilias Zadik。《稀疏高维线性回归：平方误差估计与相变》。《统计年刊》，50(2):880–903，2022年4月。2，4，5，10，31，32Avishek Ghosh 和 Ramchandran Kannan。《交替最小化在混合线性回归中超线性收敛》。载于 第二十三届国际人工智能与统计会议论文集，2020年。1Christophe Giraud。 高维统计学导论。《统计学与应用概率专著》。CRC出版社，第二版，2021年。1

Trevor Hastie、Robert Tibshirani 和 Martin Wainwright。 稀疏统计学习：Lasso及其推广。第143卷。CRC出版社、泰勒与弗朗西斯集团，2015年。1Samuel B. Hopkins 和 David Steurer。《基于少量样本的高效贝叶斯估计：社区检测及相关问题》。载于 2017年IEEE第58届计算机科学基础年度研讨会（FOCS），第379–390页，2017年10月。4，6Samuel B. Hopkins、Pravesh K. Kothari、Aaron Potechin、Prasad Raghavendra、Tselil Schramm 和 David Steurer。《平方和方法在检测隐藏结构中的威力》。载于 2017年IEEE第58届计算机科学基础年度研讨会（FOCS），第720–731页，2017年10月。6Samuel Brink Klevit Hopkins。《统计推断与平方和方法》。2018年。康奈尔大学图书馆。3，4，6，19Yunju Im、Yuan Huang、Jian Huang 和 Shuangge Ma。基于组织病理成像的癌症数据分析的贝叶斯分层有限回归混合模型。《医学统计》，41(6):1009–1022，2022年。1Amnon Jakimovski、Ambikeshwar Sharma 和 J´ozsef Szabados。《埃尔米特插值、埃尔米特-伯克霍夫插值与沃尔什等收敛》。载于 复插值多项式的沃尔什等收敛，第25–54页。施普林格数学专著，2006年。19Lucas Janson、Rina Foygel Barber 和 Emmanuel Cand`es。EigenPrism：高维信噪比的推断。《皇家统计学会杂志：B辑（统计方法论）》，79(4):1037–1065，2017年。1

Adel Javanmard、Simeng Shao 和 Jacob Bien。高维专家混合模型的预测集，2022年10月。arXiv:2210.16710。1

Michael I. Jordan 和 Robert A. Jacobs。分层专家混合模型与期望最大化算法。《神经计算》，6(2)：181–214，1994年3月。1

Karen Kazor 和 Amanda S. Hering。大型空间数据集的回归混合模型。*Technomet-rics*，61(4)：507–523，2019年。1

Abbas Khalili 和 Jiahua Chen。有限回归混合模型中的变量选择。《美国统计学会会刊》，102(479)：1025–1038，2007年。1

Florent Krzakala、Andrea Montanari、Federico Ricci-Tersenghi、Guilhem Semerjian 和 Lenka Zdeborova。随机约束满足问题的吉布斯态与解集。《美国国家科学院院刊》，104：10318–10323，2007年6月。5Dmitriy Kunisky、Alexander S. Wein 和 Afonso S. Bandeira。关于假设检验的计算困难性：利用低次似然比进行预测。载于《数学分析、应用与计算》，施普林格数学与统计学会议论文集，第1–50页，2022年。3，6，18，19Guillaume Lecu´e 和 Shahar Mendelson。相位恢复中极小极大收敛速率与经验风险最小化的性能。《电子概率期刊》，2015年。8Yan Li、Chun Yu、Yize Zhao、Weixin Yao、Robert H. Aseltine 和 Kun Chen。寻找聚类总体建模中的异质性来源。《生物统计学》，78(2)：716–729，2022年。1，8Yuanzhi Li 和 Yingyu Liang。以近乎最优的复杂度学习线性回归混合模型。学习理论会议，第1125–1144页，2018年。1Nathan Linial。分布式图算法中的局部性。《工业与应用数学学会计算杂志》，21(1)：193–201，1992年2月。5

Zhaoqiang Liu、Subhroshekhar Ghosh 和 Jonathan Scarlett。迈向具有稀疏先验与生成式先验的样本最优压缩相位恢复。载于《神经信息处理系统进展》，2021年。2，3，38 Ziye Luo、Xinyue Yao、Yifan Sun 和 Xinyan Fan。利用基于回归的异质性分析识别高维数据中的重叠子群结构。《生物计量学杂志》，64(6)：1109–1141，2022年。1 Ashok Makkuva、Pramod Viswanath、Sreeram Kannan 和 Sewoong Oh。打破专家混合模型中的僵局：一致且高效的算法。载于第36届国际机器学习会议，第4304–4313页，2019年5月。1 Arya Mazumdar 和 Soumyabrata Pal。关于学习具有稀疏参数的混合模型。载于第25届国际人工智能与统计会议论文集，第9182–9213页，2022年5月。3，4，5 Andrea Montanari 和 Alexander S. Wein。秩一矩阵估计中近似消息传递与低次多项式的等价性，2022 年 12 月。arXiv：2212.06996。7 Andrea Montanari、Daniel Reichman 和 Ofer Zeitouni。关于谱方法的局限性：从高斯隐藏团问题到高斯张量的秩一扰动。载于《神经信息处理系统进展》，第28卷，2015年。18 Keith B. Oldham、Jan C. Myland、Jerome Spanier 和 Jan Myland。函数图谱：含赤道及 Atlas函数计算器。《数学与统计学》。Springer US，Springer电子书，纽约州纽约市，2009年。国际标准书号 978-0-387-48807-3。20 Soumyabrata Pal、Arya Mazumdar 和 Venkata Gandikota。从线性测量混合中恢复稀疏信号。载于《神经信息处理系统进展》，2021年。1 Soumyabrata Pal、Arya Mazumdar、Rajat Sen 和 Avishek Ghosh。不可实现情形下线性回归混合模型的学习。载于第39届国际机器学习会议论文集，第17202–17220页，2022年。1 Richard E. Quandt 和 James B. Ramsey。正态分布混合模型与切换回归的估计。《美国统计学会会刊》，73(364)：730–738，1978年12月。1 Galen Reeves、Jiaming Xu 和 Ilias Zadik。稀疏线性回归中的全有或全无现象。载于第三十二届学习理论会议论文集，第2652–2663页，2019年6月。2，5，32 Tselil Schramm 和 Alexander S. Wein。利用低次多项式进行估计的计算障碍。《统计年刊》，50(3)，2022年6月。3，19，20 Yanyao Shen 和 Sujay Sanghavi。混合线性回归的迭代最小截尾平方。载于《神经信息处理系统进展》，第32卷，2019年。1 Nicolas St¨adler、Peter B¨uhlmann 和 Sara van de Geer。混合回归模型的L1惩罚。《TEST》，19(2)：209–256，2010年8月。1 Ramon van Handel。高维概率。技术报告，国防技术信息中心，弗吉尼亚州贝尔沃堡，2014年6月。48 Kert Viele 和 Barbara Tong。线性回归混合模型的建模。《统计与计算》，12：315–330，2002年10月。1 Martin J. Wainwright。高维含噪情形下稀疏性恢复的信息论极限。《IEEE信息论汇刊》，55(12)：5728–5741，2009a。1，5

Martin J. Wainwright. 使用 ℓ_1-约束二次规划（Lasso）实现高维和含噪稀疏性恢复的精确阈值。《IEEE信息论汇刊》，55(5):2183– 2202，2009b年5月。1, 2, 4, 5, 10 Martin J. Wainwright.《高维统计学：非渐近视角》。剑桥大学出版社，第1版，2019年2月。35, 48, 58 Wei Wang, Martin J. Wainwright, and Kannan Ramchandran. 稀疏信号恢复的信息论极限：稠密测量矩阵与稀疏测量矩阵。《IEEE信息论汇刊》，56 (6):2967–2979，2010年6月。2, 5, 32 Michel Wedel and Wagner A Kamakura.《市场细分：概念与方法基础》。施普林格《概念与方法基础》，2000年。1 Fan Wu and Patrick Rebeschini. 面向稀疏相位恢复的 Hadamard Wirtinger Flow。载于《第24届国际人工智能与统计会议论文集》，第982–990页，2021年。2, 3, 8 Xinyang Yi, Constantine Caramanis, and Sujay Sanghavi. 混合线性回归的交替最小化。载于《国际机器学习会议》，第613–621页，2014年。1, 9, 51, 53, 54 Guoshen Yu and Guillermo Sapiro. 高斯混合模型的统计压缩感知。《IEEE信号处理汇刊》，59(12):5842–5858，2011年12月。I SSN 1941-0476。doi: 10.1109/ TSP.2011.2168521。会议名称：《IEEE信号处理汇刊》。9 Ilias Zadik. Lenstra-Lenstra-Lovasz算法在无噪声推断中的惊人威力。Simons Institute IFML/CCSI研讨会，2021年。URL https://youtu.be/QGUEaXII4Kw。12 Ilias Zadik, Min Jae Song, Alexander S. Wein, and Joan Bruna. 基于格的方法在聚类中胜过平方和方法。载于《第35届学习理论会议论文集》，第1247–1248页，2022年6月。12 Lenka Zdeborova and Florent Krzakala. 推断的统计物理：阈值与算法。《物理学进展》，65(5):453–552，2016年9月。5 Yihan Zhang, Marco Mondelli, and Ramji Venkataramanan. 混合广义线性模型中谱方法的精确渐近分析，2022年12月。arXiv:2211.11368。1 Kai Zhong, Prateek Jain, and Inderjit S. Dhillon. 多组分混合线性回归。载于《神经信息处理系统进展》，第2190–2198页，2016年。1

A 关于低次方法的补充背景

本节将补充介绍低次方法、卡方散度及其在低次多项式空间上的正交投影。考虑第 1.3 节中的设定，任务是区分两个概率分布 $\mathbb{P}_N$ 和 $\mathbb{Q}_N$，它们定义在$\mathbb{R}^N$ 上，其中 N 是（可能不断增长的问题规模）。给定一个从 $\boldsymbol{x}$ 或 $\mathbb{P}_N$ 中抽取的样本 $\mathbb{Q}_N$，我们希望通过假设检验判断 $\boldsymbol{x}$ 究竟来自前者还是后者。回顾第 1.3 节中强检测和弱检测的概念。

判断强检测或弱检测是否可能的一种有力方法，是研究卡方散度 $\chi^2(\mathbb{P}_N\|\mathbb{Q}_N)$。事实上，假设 $\mathbb{P}_N$ 关于 $\mathbb{Q}_N$ 绝对连续，并令 $\mathsf{L} = \frac{d\mathbb{P}_N}{d\mathbb{Q}_N}$ 为似然比。我们有：

$$\begin{aligned}\chi^2(\mathbb{P}_N\|\mathbb{Q}_N) &:= \mathbb{E}_{\boldsymbol{x}\sim\mathbb{Q}_N}\mathsf{L}(\boldsymbol{x})^2 - 1\\ &= \sup_{f:\mathbb{R}^p\to\mathbb{R}} \frac{(\mathbb{E}_{\boldsymbol{x}\sim\mathbb{P}_N} f(\boldsymbol{x}))^2}{\mathbb{E}_{\boldsymbol{x}\sim\mathbb{Q}_N} f(\boldsymbol{x})^2} - 1\\ &= \sup_{\substack{f:\mathbb{R}^p\to\mathbb{R}\\ \mathbb{E}_{\boldsymbol{x}\sim\mathbb{Q}_N} f(\boldsymbol{x})=0}} \frac{(\mathbb{E}_{\boldsymbol{x}\sim\mathbb{P}_N} f(\boldsymbol{x}))^2}{\mathbb{E}_{\boldsymbol{x}\sim\mathbb{Q}_N} f(\boldsymbol{x})^2},\end{aligned}$$

其中，等价关系可由标准论证得出（参见 Kunisky 等人 [2022]）。对上述结果的解释是，卡方散度体现了 L^2 意义下的最优性。它与任意函数在 $\mathbb{P}_N$ 下的最大期望值的平方有关，同时该函数在空间 $L^2(\mathbb{Q}_N$中有界。事实上，两个分布之间的卡方散度可以严格刻画它们在检验中的行为：

引理 A.1 (改编自 Montanari 等人 [2015] 的引理 2 以及 Coja-Oghlan 等人 [2022]的引理 7.1)。
若 $\chi^2(\mathbb{P}_N\|\mathbb{Q}_N) = O(1)$（当 $N\to\infty$ 时），则不可能实现强检测。

- 如果当 $N\to\infty$ 时，$\chi^2(\mathbb{P}_N\|\mathbb{Q}_N) = o(1)$，那么弱检测是不可能的。

这一结果很有力，因为它指出卡方散度是在检验中寻找可识别性结果的充分量。不过请注意，这一量并未揭示任何与计算有关的信息。

卡方散度的计算类比是度数-D 卡方散度$\chi^2_{\le D}(\mathbb{P}_N\|\mathbb{Q}_N)$。该量衡量 $\mathbb{P}_N$和 $\mathbb{Q}_N$是否能通过输入 $\boldsymbol{x}$ 的度数-D多项式加以区分。考虑希尔伯特空间 $L^2(\mathbb{Q}_N)$，其中对于函数 f、g：$\mathbb{R}^p\to\mathbb{R}$，其内积 $\langle f$、$g\rangle$ 为$= \mathbb{E}_{\boldsymbol{x}\sim Q_N}[f(\boldsymbol{x})g(\boldsymbol{x})]$，相应的范数为 $\|f\|_{\mathbb{Q}_N} = \sqrt{\langle f}$、$f\rangle_{\mathbb{Q}_N}$。此外，记 $\mathbb{R}[\boldsymbol{x}]_{\le D}$为从 $\mathbb{R}^p$到 $\mathbb{R}$ 的次~~数至多为~~ D的多元多项式空间，并令 $f^{\le D}$表示 f在 $L^2(\mathbb{Q}_N)$中投影到 $R[\boldsymbol{x}]_{\le D}$上的正交投影。于是，我们可以如下定义 $\chi^2_{\le D}(\mathbb{P}_N\|\mathbb{Q}_N)$：

$$\begin{aligned}\chi^2_{\le D}(\mathbb{P}_N\|\mathbb{Q}_N) &:= \mathbb{E}_{\boldsymbol{x}\sim\mathbb{Q}_N}\mathsf{L}^{\le D}(\boldsymbol{x})^2 - 1\\ &= \|\mathsf{L}^{\le D}\|^2_{Q_N} - 1\\ &= \sup_{f\in\mathbb{R}[\boldsymbol{x}]_{\le D}} \frac{(\mathbb{E}_{\boldsymbol{x}\sim\mathbb{P}_N} f(\boldsymbol{x}))^2}{\mathbb{E}_{\boldsymbol{x}\sim\mathbb{Q}_N} f(\boldsymbol{x})^2} - 1\\ &= \sup_{\substack{f\in\mathbb{R}[\boldsymbol{x}]_{\le D}\\ \mathbb{E}_{\boldsymbol{x}\sim\mathbb{Q}_N} f(\boldsymbol{x})=0}} \frac{(\mathbb{E}_{\boldsymbol{x}\sim\mathbb{P}_N} f(\boldsymbol{x}))^2}{\mathbb{E}_{\boldsymbol{x}\sim\mathbb{Q}_N} f(\boldsymbol{x})^2}.\end{aligned} \tag{6}$$

该结果的证明见Hopkins [2018], Kunisky 等人 [2022]。因此，低阶卡方散度可以类比于卡方散度来理解：它量化了任意低阶函数在 $\mathbb{P}_N$ 下所能具有的最大期望，同时仍处于D 的度数- $L^2(\mathbb{Q}_N)$多项式子空间中。于是，对于输入的低次多项式函数，以及猜想中的一般多项式时间算法，我们有引理A.1 的类比结果，即猜想1.2。

在本文中，我们考虑检验以下分布：它们并非简单地由信号加噪声构成，而是由经过线性变换的信号加噪声构成。结合近期的工作 Bandeira 等人 [2022], Arpino [2021], 据我们所知，这是低次方法在此类问题上的首次应用之一；此前人们认为，现有方法无法处理这类问题 [Schramm 和 Wein, 2022]。

B 低次下界的证明

预备知识与记号。关于相关问题低次困难性的所有结果均为 p 中的渐近结果，因为我们首先令 $p \to \infty$ 。我们沿用 Schramm 和 Wein [2022] 中的约定。令 $\mathbb{N} = \{0, 1, 2, \cdots\}$ 和 $[n] = \{1, 2 \cdots n\}$。定义 $0^0 := 1$。我们用粗体表示多重集或向量，因此对于 $\boldsymbol{\alpha} \in \mathbb{N}^n$，我们指的是 $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_n]$ 对于$\alpha_i \in \mathbb{N}, \forall i \in [n]$。对于 $\boldsymbol{\alpha} \in \mathbb{N}^n$，定义 $|\boldsymbol{\alpha}| = \sum_{i=1}^n \alpha_i$，$\boldsymbol{\alpha}! = \prod_{i=1}^n \alpha_i!$ 以及（对于 $\boldsymbol{X} \in \mathbb{R}^n$）$\boldsymbol{X}^{\boldsymbol{\alpha}} = \prod_{i=1}^n X_i^{\alpha_i}$。令 $\mathrm{abs}(\boldsymbol{\alpha})$ 表示向量 $\boldsymbol{\alpha}$ 的逐元素绝对值运算。我们用 $\boldsymbol{\alpha} \geq \boldsymbol{\beta}$ 表示对所有 i 均有 $\alpha_i \geq \boldsymbol{\beta}_i$。运算 $\boldsymbol{\alpha} + \boldsymbol{\beta}$和 $\boldsymbol{\alpha} - \boldsymbol{\beta}$均逐元素进行。对于 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^n$，若 $\boldsymbol{\alpha} \geq \boldsymbol{\beta}$，定义 $\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}} = \prod_{i=1}^n \binom{\alpha_i}{\boldsymbol{\beta}_i}$。我们用下标表示向量或多重集的子集，因此对于 $\boldsymbol{\alpha} \in \mathbb{N}^{n \times (p+1)}$，令 $\boldsymbol{\alpha}_{p+1} := [\alpha_{1,p+1}, \ldots, \alpha_{n,p+1}]$ 表示矩阵 $\boldsymbol{\alpha}$ 的第 $p+1$ 列。令 $\boldsymbol{\alpha}_{,:p}$ 表示通过仅选取至第 p 列而得到的整个 $n \times p$ 子矩阵；令 $\boldsymbol{\alpha}_{i,:p}$ 表示由第 i 行至第 p 列的元素构成的向量。我们用 $[\boldsymbol{A} \ \ \boldsymbol{y}]$ 表示将 $\boldsymbol{y} \in \mathbb{R}^n$ 水平拼接到 $\boldsymbol{A} \in \mathbb{R}^{n \times p}$ 上所形成的矩阵，它是一个 $n \times (p+1)$ 实矩阵。除非另有说明，令 $\|\cdot\| := \|\cdot\|_{\mathbb{Q}_p}$ 且 $\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_{\mathbb{Q}_p}$。我们用 $\mathbf{1}$ 表示指示函数。

单变量厄米多项式 $H_k(x)$ 对于 $k \geq 0$ 由递推关系 $H_0(x) = 1$ 和 $H_{k+1}(x) = xH_k(x) - H_k'(x)$ 定义。对于 $\boldsymbol{\alpha} \in \mathbb{N}^N$，令 $H_{\boldsymbol{\alpha}}$ 表示 $\boldsymbol{\alpha}$ 阶多元厄米多项式，其定义为 $H_{\boldsymbol{\alpha}}(\boldsymbol{u}) = \prod_{i=1}^N H_{\alpha_i}(u_i)$，其中 $\boldsymbol{u} \in \mathbb{R}^N$。对于 $N \in \mathbb{N}$，归一化的 N 元厄米多项式 $\frac{1}{\sqrt{\boldsymbol{\alpha}!}} H_{\boldsymbol{\alpha}}$ 构成 $L^2(\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N))$ 上（多元）多项式的完备正交系统（参见 Kunisky 等人 [2022]）。

接下来，我们将介绍关于厄米多项式的更多基本事实（参见 Kunisky 等人[2022] 以了解更详细的描述），并给出两个辅助组合引理，这些引理将在主要证明中发挥作用。

命题 B.1 （高斯分部积分，Kunisky 等人中的命题 2.10 [2022]）。若 $f : \mathbb{R} \to \mathbb{R}$ 是 k 次连续可微，且 $f(y)$ 及其前 k 阶导数均有界于$\mathcal{O}\left(\exp(|y|^\alpha)\right)$，其中 $\alpha \in (0, 2)$，则

$$\mathbb{E}_{y \sim \mathcal{N}(0,1)} \left[H_k(y) f(y) \right] = \mathbb{E}_{y \sim \mathcal{N}(0,1)} \left[\frac{d^k f}{dy^k}(y) \right].$$

命题 B.2 （厄米多项式导数 [Jakimovski 等人, 2006]）。对于 $n \in \mathbb{N}$, $m \in \mathbb{N}$:

$$H_n^{(m)}(x) = \frac{n!}{(n-m)!} H_{n-m}(x).$$

命题 B.3 （Hermite 求和公式，Schramm 和 Wein [2022]中的命题 3.1）。 对于任意 $k \in \mathbb{N}$和 z, $\mu \in \mathbb{R}$,

$$H_k(z+\mu) = \sum_{l=0}^{k} \binom{k}{l} \mu^{k-l} H_l(z).$$

命题 B.4 (Hermite 乘法公式 Oldham 等人 [2009])。 对于 $\gamma \in \mathbb{R}$,

$$H_n(\gamma x) = \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} \gamma^{n-2i} \left(\gamma^2 - 1\right)^i \binom{n}{2i} \frac{(2i)!}{i!} 2^{-i} H_{n-2i}(x).$$

命题 B.5。考虑零分布 $\mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})$，其分布律为

$$\mathcal{N}(0,1)^{\otimes n \times p} \otimes \mathcal{N}\left(0, \frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} + 1\right)^{\otimes (p+1)}.$$

令 $\boldsymbol{u} = [\boldsymbol{X} \ \boldsymbol{y}] \in \mathbb{R}^{N \times (p+1)}$。相对于此零分布，以 $\boldsymbol{\alpha} \in \mathbb{N}^{n \times (p+1)}$ 为索引的正交归一系统为

$$\frac{1}{\sqrt{\boldsymbol{\alpha}!}} \tilde{H}_{\boldsymbol{\alpha}}(\boldsymbol{u}) := \frac{1}{\sqrt{\boldsymbol{\alpha}!}} \prod_{i=1}^{n} \prod_{j=1}^{p} H_{\alpha_{i,j}}(u_{i,j}) H_{\alpha_{i,p+1}}\left(\frac{u_{i,p+1}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} + 1}}\right).$$

证明。令 $\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)} \in \mathbb{N}^{n \times (p+1)}$。于是,

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})} \frac{1}{\sqrt{\boldsymbol{\alpha}^{(1)}!}} \tilde{H}_{\boldsymbol{\alpha}^{(1)}}(\boldsymbol{u}) \frac{1}{\sqrt{\boldsymbol{\alpha}^{(2)}!}} \tilde{H}_{\boldsymbol{\alpha}^{(2)}}(\boldsymbol{u}) \\
&= \frac{1}{\sqrt{\boldsymbol{\alpha}^{(1)}! \boldsymbol{\alpha}^{(2)}!}} \mathbb{E}_{\mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})} \prod_{i=1}^{n} \prod_{j=1}^{p} H_{\alpha_{i,j}^{(1)}}(u_{i,j}) H_{\alpha_{i,p+1}^{(1)}}\left(\frac{u_{i,p+1}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} + 1}}\right) H_{\alpha_{i,j}^{(2)}}(u_{i,j}) H_{\alpha_{i,p+1}^{(2)}}\left(\frac{u_{i,p+1}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} + 1}}\right) \\
&= \frac{1}{\sqrt{\boldsymbol{\alpha}^{(1)}! \boldsymbol{\alpha}^{(2)}!}} \mathbb{E} \prod_{i=1}^{n} \prod_{j=1}^{p} H_{\alpha_{i,j}^{(1)}}(u_{i,j}) H_{\alpha_{i,j}^{(2)}}(u_{i,j}) \, \mathbb{E} \prod_{i=1}^{n} H_{\alpha_{i,p+1}^{(1)}}\left(\frac{u_{i,p+1}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} + 1}}\right) H_{\alpha_{i,p+1}^{(2)}}\left(\frac{u_{i,p+1}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} + 1}}\right) \\
&= \frac{1}{\sqrt{\boldsymbol{\alpha}^{(1)}! \boldsymbol{\alpha}^{(2)}!}} \mathbb{E} \prod_{i=1}^{n} \prod_{j=1}^{p} H_{\alpha_{i,j}^{(1)}}(u_{i,j}) H_{\alpha_{i,j}^{(2)}}(u_{i,j}) \, \mathbb{E} \prod_{i=1}^{n} H_{\alpha_{i,p+1}^{(1)}}(w_i) \, H_{\alpha_{i,p+1}^{(2)}}(w_i) \\
&= \frac{1}{\sqrt{\boldsymbol{\alpha}^{(1)}! \boldsymbol{\alpha}^{(2)}!}} \sqrt{\boldsymbol{\alpha}^{(1)}! \boldsymbol{\alpha}^{(2)}!} \prod_{i=1}^{n} \prod_{j=1}^{p} \mathbb{I}_{\{\alpha_{i,j}^{(1)} = \alpha_{i,j}^{(2)}\}} \prod_{i=1}^{n} \mathbb{I}_{\{\alpha_{i,p+1}^{(1)} = \alpha_{i,p+1}^{(2)}\}} \\
&= \mathbb{I}_{\boldsymbol{\alpha}^{(1)} = \boldsymbol{\alpha}^{(2)}},
\end{aligned}$$

其中 $w_i \overset{\text{独立同分布}}{\sim} \mathcal{N}(0,1)$ 与 $i \in [n]$ 对应的所有其他变量独立。 □

引理 B.6. 对于 $\boldsymbol{\beta} \in \mathbb{N}$ 偶数:

$$\sum_{\xi=0}^{\frac{\boldsymbol{\beta}}{2}} \binom{\boldsymbol{\beta}}{2\xi} \frac{(2\xi)!}{\xi!} \left(-\frac{1}{2}\right)^{\xi} (\boldsymbol{\beta} - 2\xi - 1)!! = \mathbb{I}_{\boldsymbol{\beta}=0}.$$

证明。我们有：

$$\begin{aligned}\sum_{\xi=0}^{\frac{\boldsymbol{\beta}}{2}}\binom{\boldsymbol{\beta}}{2\xi}\frac{(2\xi)!}{\xi!}\left(-\frac{1}{2}\right)^{\xi}(\boldsymbol{\beta}-2\xi-1)!! &= \sum_{\xi=0}^{\frac{\boldsymbol{\beta}}{2}}\frac{\boldsymbol{\beta}!}{\xi!(\boldsymbol{\beta}-2\xi)!!}\left(\frac{-1}{2}\right)^{\xi}\\ &= \sum_{\xi=0}^{\frac{\boldsymbol{\beta}}{2}}\frac{\boldsymbol{\beta}!}{\xi!\cdot(\frac{\boldsymbol{\beta}}{2}-\xi)!\cdot 2^{\frac{\boldsymbol{\beta}}{2}-\xi}}\left(\frac{-1}{2}\right)^{\xi}\\ &= \frac{\boldsymbol{\beta}!}{(\frac{\boldsymbol{\beta}}{2})!\cdot 2^{\frac{\boldsymbol{\beta}}{2}}}\sum_{\xi=0}^{\frac{\boldsymbol{\beta}}{2}}\binom{\frac{\boldsymbol{\beta}}{2}}{\xi}(-1)^{\xi}\\ &= \frac{\boldsymbol{\beta}!}{(\frac{\boldsymbol{\beta}}{2})!\cdot 2^{\frac{\boldsymbol{\beta}}{2}}}(1+(-1))^{\frac{\boldsymbol{\beta}}{2}}\\ &= \mathbb{I}^{\boldsymbol{\beta}=0}.\end{aligned}$$

□

引理 B.7。对于 $p\geq 4$ 和 $k\leq\sqrt{p}$，有 $\frac{p^k}{4k}!\ \leq\binom{p}{k}$。

证明。注意，$\binom{p}{k}\geq\frac{p^k}{4k}!$ 当且仅当：

$$\prod_{j=1}^{k-1}\left(1-\frac{j}{p}\right)\geq\frac{1}{4}$$

然后应用 $k\leq\sqrt{p}$ 假设：

$$\prod_{j=1}^{k-1}\left(1-\frac{j}{p}\right)\geq\prod_{i=1}^{\lfloor\sqrt{p}\rfloor}\left(1-\frac{j}{p}\right)\geq\left(1-\frac{1}{\sqrt{p}}\right)^{\sqrt{p}}$$

现在注意，对于 $\sqrt{p}\geq 2$，我们有 $(1-\frac{1}{\sqrt{p}})^{\sqrt{p}}\geq\frac{1}{4}$，从而得到所需结果。 □

B.1 MSLR 的低阶分析：一般混合模型

本小节中，我们证明两个技术性引理。第一个（引理 B.8）推导出~似然比在多元厄米多项式 $H_{\boldsymbol{\alpha}}$ 上的投影的表达式，该多项式定义于命题 B.5。第二个引理（引理 B.9）推导出低阶卡方散度的显式表达式。

引理 B.8. 设 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2\sim\mathcal{P}_{\|\boldsymbol{\beta}\|_2}(\mathcal{D})$。设 $\mathsf{L}=\frac{d\mathbb{P}(\boldsymbol{X},\boldsymbol{y})}{d\mathbb{P}(\boldsymbol{X})\otimes\mathbb{P}(\boldsymbol{y})}$ 为似然比，且 $\tilde{H}_{\boldsymbol{\alpha}}$ 为命题 B.5中定义的厄米特多项式。于是，对于 $\boldsymbol{\alpha}\in\mathbb{N}^{n\times(p+1)}$，有

$$\begin{aligned}\langle\mathsf{L},\tilde{H}_{\boldsymbol{\alpha}}\rangle = &\left(\frac{\frac{1}{\sigma}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1}}\right)^{|\boldsymbol{\alpha}_{p+1}|}\|\boldsymbol{\beta}\|_2^{|\boldsymbol{\alpha}_{p+1}|-|\boldsymbol{\alpha}_{\cdot,:p}|}\boldsymbol{\alpha}_{p+1}!\\ &\cdot\prod_{i=1}^{n}\mathbb{I}^{\{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{\cdot,:p}|=0\}}\mathop{\mathbb{E}}_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2}\left[\prod_{i=1}^{n}\prod_{j=1}^{p}(\boldsymbol{\beta}_{1,j}z_i+\boldsymbol{\beta}_{2,j}(1-z_i))^{\alpha_{i,j}}\right].\end{aligned}$$

证明。我们先展开内积：

$$
\begin{aligned}
&\langle \mathsf{L}, \tilde{H}_{\boldsymbol{\alpha}} \rangle \\
&= \mathop{\mathbb{E}}_{\mathbb{P}(\boldsymbol{X})\otimes\mathbb{P}(\boldsymbol{y})} \left[\frac{d\mathbb{P}(\boldsymbol{X}, \boldsymbol{y})}{d\mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})} \tilde{H}_{\boldsymbol{\alpha}}(\boldsymbol{X}, \boldsymbol{y}) \right] \\
&= \mathop{\mathbb{E}}_{\mathbb{P}(\boldsymbol{X},\boldsymbol{y})} \left[\prod_{i=1}^{n} \left(\prod_{j=1}^{p} H_{\alpha_{i,j}}(X_{i,j}) \right) H_{\alpha_{i,p+1}} \left(\frac{\left(\frac{1}{\sigma} \boldsymbol{X}\boldsymbol{\beta}_1 \odot \boldsymbol{z} + \frac{1}{\sigma} \boldsymbol{X}\boldsymbol{\beta}_2 \odot (1-\boldsymbol{z}) + \boldsymbol{w} \right)_i}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} + 1}} \right) \right],
\end{aligned}
$$

并应用高斯分部积分（命题B.1），我们得到

$$
\begin{aligned}
&\langle \mathsf{L}, \tilde{H}_{\boldsymbol{\alpha}} \rangle \\
&= \mathop{\mathbb{E}}_{\boldsymbol{X},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{z},\boldsymbol{w}} \prod_{i=1}^{n} \left(\frac{\frac{1}{\sigma}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} + 1}} \right)^{|\boldsymbol{\alpha}_{i,:p}|} \frac{\alpha_{i,p+1}!}{(\alpha_{i,p+1} - |\boldsymbol{\alpha}_{i,:p}|)!} \\
&\qquad \cdot \prod_{j=1}^{p} \left(\boldsymbol{\beta}_{1,j} z_i + \boldsymbol{\beta}_{2,j}(1-z_i) \right)^{\alpha_{i,j}} H_{\alpha_{i,p+1} - |\boldsymbol{\alpha}_{i,:p}|} \left(\frac{\left(\frac{1}{\sigma} \boldsymbol{X}\boldsymbol{\beta}_1 \odot \boldsymbol{z} + \frac{1}{\sigma} \boldsymbol{X}\boldsymbol{\beta}_2 \odot (1-\boldsymbol{z}) + \boldsymbol{w} \right)_i}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} + 1}} \right) \\
&= \left(\frac{\frac{1}{\sigma}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} + 1}} \right)^{|\boldsymbol{\alpha}_{\cdot,:p}|} \mathop{\mathbb{E}}_{\boldsymbol{X},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{z},\boldsymbol{w}} \prod_{i=1}^{n} \frac{\alpha_{i,p+1}!}{(\alpha_{i,p+1} - |\boldsymbol{\alpha}_{i,:p}|)!} \\
&\qquad \cdot \prod_{j=1}^{p} \left(\boldsymbol{\beta}_{1,j} z_i + \boldsymbol{\beta}_{2,j}(1-z_i) \right)^{\alpha_{i,j}} H_{\alpha_{i,p+1} - |\boldsymbol{\alpha}_{i,:p}|} \left(\frac{\left(\frac{1}{\sigma} \boldsymbol{X}\boldsymbol{\beta}_1 \odot \boldsymbol{z} + \frac{1}{\sigma} \boldsymbol{X}\boldsymbol{\beta}_2 \odot (1-\boldsymbol{z}) + \boldsymbol{w} \right)_i}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} + 1}} \right).
\end{aligned}
$$

接下来，我们应用命题 B.3 和B.4中概述的厄米乘法公式与加法公式：

$$
\begin{aligned}
&\langle \mathsf{L}, \tilde{H}_{\boldsymbol{\alpha}} \rangle \\
&= \left(\frac{\frac{1}{\sigma}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} + 1}} \right)^{|\boldsymbol{\alpha}_{\cdot,:p}|} \mathop{\mathbb{E}}_{\boldsymbol{X},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{z},\boldsymbol{w}} \prod_{i=1}^{n} \frac{\alpha_{i,p+1}!}{(\alpha_{i,p+1} - |\alpha_{i,:p}|)!} \prod_{j=1}^{p} \left(\boldsymbol{\beta}_{1,j} z_i + \boldsymbol{\beta}_{2,j}(1-z_i) \right)^{\alpha_{i,j}} \\
&\cdot \sum_{\xi=0}^{\lfloor \frac{\alpha_{i,p+1} - |\boldsymbol{\alpha}_{i,:p}|}{2} \rfloor} \left(\frac{1}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} + 1}} \right)^{\alpha_{i,p+1} - |\boldsymbol{\alpha}_{i,:p}| - 2\xi} \left(\frac{1}{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} + 1} - 1 \right)^{\xi} \binom{\alpha_{i,p+1} - |\boldsymbol{\alpha}_{i,:p}|}{2\xi} \frac{(2\xi)!}{\xi!} 2^{-\xi} \\
&\cdot H_{\alpha_{i,p+1} - |\boldsymbol{\alpha}_{i,:p}| - 2\xi} \left(\left(\frac{1}{\sigma} \boldsymbol{X}\boldsymbol{\beta}_1 \odot \boldsymbol{z} + \frac{1}{\sigma} \boldsymbol{X}\boldsymbol{\beta}_2 \odot (1-\boldsymbol{z}) + \boldsymbol{w} \right)_i \right) \\
&= \left(\frac{\frac{1}{\sigma}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} + 1}} \right)^{|\boldsymbol{\alpha}_{\cdot,:p}|} \mathop{\mathbb{E}}_{\boldsymbol{X},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{z},\boldsymbol{w}} \prod_{i=1}^{n} \frac{\alpha_{i,p+1}!}{(\alpha_{i,p+1} - |\alpha_{i,:p}|)!} \prod_{j=1}^{p} \left(\boldsymbol{\beta}_{1,j} z_i + \boldsymbol{\beta}_{2,j}(1-z_i) \right)^{\alpha_{i,j}} \\
&\cdot \sum_{\xi=0}^{\lfloor \frac{\alpha_{i,p+1} - |\boldsymbol{\alpha}_{i,:p}|}{2} \rfloor} \left(\frac{1}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} + 1}} \right)^{\alpha_{i,p+1} - |\boldsymbol{\alpha}_{i,:p}| - 2\xi} \left(\frac{1}{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} + 1} - 1 \right)^{\xi} \binom{\alpha_{i,p+1} - |\boldsymbol{\alpha}_{i,:p}|}{2\xi} \frac{(2\xi)!}{\xi!} 2^{-\xi}
\end{aligned}
$$

$$\cdot \sum_{\eta=0}^{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|-2\xi} \binom{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|-2\xi}{\eta} \left(\left(\frac{1}{\sigma}\boldsymbol{X}\boldsymbol{\beta}_1 \odot \boldsymbol{z} + \frac{1}{\sigma}\boldsymbol{X}\boldsymbol{\beta}_2 \odot (1-\boldsymbol{z})\right)_i\right)^{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|-2\xi-\eta} \underbrace{H_\eta(w_i)}_{\neq 0 \text{ only if } \eta=0},$$

注意到 $H_\eta(w_i) \neq 0$ 仅当 $\eta = 0$ 时成立，我们将其化简，得到：

$$\begin{aligned}
&\langle \mathsf{L}, \tilde{H}_{\boldsymbol{\alpha}} \rangle \\
&= \left(\frac{\frac{1}{\sigma}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1}}\right)^{|\boldsymbol{\alpha}_{\cdot,:p}|} \mathop{\mathbb{E}}_{\boldsymbol{X},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{z}} \prod_{i=1}^n \frac{\alpha_{i,p+1}!}{(\alpha_{i,p+1}-|\alpha_{i,:p}|)!} \prod_{j=1}^p (\boldsymbol{\beta}_{1,j} z_i + \boldsymbol{\beta}_{2,j}(1-z_i))^{\alpha_{i,j}} \\
&\cdot \sum_{\xi=0}^{\lfloor \frac{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|}{2} \rfloor} \left(\frac{1}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1}}\right)^{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|-2\xi} \left(\frac{1}{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1}-1\right)^{\xi} \binom{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|}{2\xi} \frac{(2\xi)!}{\xi!} 2^{-\xi} \\
&\cdot \left(\left(\frac{1}{\sigma}\boldsymbol{X}\boldsymbol{\beta}_1 \odot \boldsymbol{z} + \frac{1}{\sigma}\boldsymbol{X}\boldsymbol{\beta}_2 \odot (1-\boldsymbol{z})\right)_i\right)^{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|-2\xi}.
\end{aligned}$$

现在交换求和与乘积的顺序并对各项进行分组，我们得到：

$$\begin{aligned}
&\langle \mathsf{L}, \tilde{H}_{\boldsymbol{\alpha}} \rangle \\
&= \left(\frac{\frac{1}{\sigma}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1}}\right)^{|\boldsymbol{\alpha}_{\cdot,:p}|} \prod_{i=1}^n \frac{\alpha_{i,p+1}!}{(\alpha_{i,p+1}-|\alpha_{i,:p}|)!} \mathop{\mathbb{E}}_{[\boldsymbol{X},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{z}]\sim\mathbb{P}(\boldsymbol{X},y)} \prod_{j=1}^p (\boldsymbol{\beta}_{1,j} z_i + \boldsymbol{\beta}_{2,j}(1-z_i))^{\alpha_{i,j}} \\
&\cdot \sum_{\xi=0}^{\lfloor \frac{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|}{2} \rfloor} \binom{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|}{2\xi} \frac{(2\xi)!}{\xi!} \left(\frac{\left(\frac{1}{\sigma}\boldsymbol{X}\boldsymbol{\beta}_1 \odot \boldsymbol{z} + \frac{1}{\sigma}\boldsymbol{X}\boldsymbol{\beta}_2 \odot (1-\boldsymbol{z})\right)_i}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1}}\right)^{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|-2\xi} \left(\frac{-\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}}{2\left(\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1\right)}\right)^{\xi} \\
&= \left(\frac{\frac{1}{\sigma}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1}}\right)^{|\boldsymbol{\alpha}_{\cdot,:p}|} \frac{\boldsymbol{\alpha}_{p+1}!}{(\boldsymbol{\alpha}_{p+1}-|\boldsymbol{\alpha}_{\cdot,:p}|)!} \mathop{\mathbb{E}}_{[\boldsymbol{X},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{z}]\sim\mathbb{P}(\boldsymbol{X},y)} \sum_{0\leq\boldsymbol{\xi}\leq\lfloor\frac{\boldsymbol{\alpha_{i,p+1}}-|\boldsymbol{\alpha}_{i,:p}|}{2}\rfloor} \prod_{i=1}^n \\
&\cdot \binom{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|}{2\xi_i} \frac{(2\xi_i)!}{\xi_i!} \left(\frac{-\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}}{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1}\right)^{\xi_i} \left(\frac{\frac{1}{\sigma}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1}}\right)^{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|-2\xi_i} \\
&\cdot \left[\prod_{j=1}^p (\boldsymbol{\beta}_{1,j} z_i + \boldsymbol{\beta}_{2,j}(1-z_i))^{\alpha_{i,j}} \left(\boldsymbol{X}\boldsymbol{\beta}_1 \odot \boldsymbol{z} + \boldsymbol{X}\boldsymbol{\beta}_2 \odot (1-\boldsymbol{z})\right)_i^{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|-2\xi_i}\right],
\end{aligned}$$

经过化简和展开后，我们便得到

$$\begin{aligned}
&\langle \mathsf{L}, \tilde{H}_{\boldsymbol{\alpha}} \rangle \\
&= \left(\frac{\frac{1}{\sigma}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1}}\right)^{|\boldsymbol{\alpha}_{\cdot,p+1}|} \frac{\boldsymbol{\alpha}_{p+1}!}{(\boldsymbol{\alpha}_{p+1}-|\boldsymbol{\alpha}_{\cdot,:p}|)!} \sum_{0\leq\boldsymbol{\xi}\leq\lfloor\frac{\boldsymbol{\alpha_{i,p+1}}-|\boldsymbol{\alpha}_{i,:p}|}{2}\rfloor} \\
&\cdot \left(\prod_{i=1}^n \binom{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|}{2\xi_i} \frac{(2\xi_i)!}{\xi_i!} \left(\frac{-\|\boldsymbol{\beta}\|_2^2}{2}\right)^{\xi_i}\right)
\end{aligned}$$

$$\cdot \mathop{\mathbb{E}}_{[\boldsymbol{X},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{z}]\sim\mathbb{P}(\boldsymbol{X},y)} \prod_{i=1}^{n}\left[\prod_{j=1}^{p}(\boldsymbol{\beta}_{1,j}z_i+\boldsymbol{\beta}_{2,j}(1-z_i))^{\alpha_{i,j}}\left(\sum_{j=1}^{p}X_{i,j}\boldsymbol{\beta}_{1,j}z_i+\sum_{j=1}^{p}X_{i,j}\boldsymbol{\beta}_{2,j}(1-z_i)\right)^{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|-2\xi_i}\right].$$

将关于 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2$ 的期望提到外面，我们便得到：

$$\begin{aligned}
&\langle \mathsf{L},\tilde{H}_{\boldsymbol{\alpha}}\rangle\\
&=\left(\frac{\frac{1}{\sigma}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1}}\right)^{|\boldsymbol{\alpha}_{\cdot,p+1}|}\frac{\boldsymbol{\alpha}_{p+1}!}{(\boldsymbol{\alpha}_{p+1}-|\boldsymbol{\alpha}_{\cdot,:p}|)!}\sum_{0\le\boldsymbol{\xi}\le\lfloor\frac{\boldsymbol{\alpha_{i,p+1}}-|\boldsymbol{\alpha_{i,:p}}|}{\mathbf{2}}\rfloor}\mathop{\mathbb{E}}_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2}\left(\prod_{i=1}^{n}\prod_{j=1}^{p}(\boldsymbol{\beta}_{1,j}z_i+\boldsymbol{\beta}_{2,j}(1-z_i))^{\alpha_{i,j}}\right)\\
&\cdot\prod_{i=1}^{n}\binom{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|}{2\xi_i}\frac{(2\xi_i)!}{\xi_i!}\left(\frac{-\|\boldsymbol{\beta}\|_2^2}{2}\right)^{\xi_i}\mathop{\mathbb{E}}_{\boldsymbol{X},\boldsymbol{z}}\left(\sum_{j=1}^{p}X_{i,j}\boldsymbol{\beta}_{1,j}z_i+\sum_{j=1}^{p}X_{i,j}\boldsymbol{\beta}_{2,j}(1-z_i)\right)^{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|-2\xi_i}\\
&=\left(\frac{\frac{1}{\sigma}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1}}\right)^{|\boldsymbol{\alpha}_{\cdot,p+1}|}\frac{\boldsymbol{\alpha}_{p+1}!}{(\boldsymbol{\alpha}_{p+1}-|\boldsymbol{\alpha}_{\cdot,:p}|)!}\sum_{0\le\boldsymbol{\xi}\le\lfloor\frac{\boldsymbol{\alpha_{i,p+1}}-|\boldsymbol{\alpha_{i,:p}}|}{\mathbf{2}}\rfloor}\mathop{\mathbb{E}}_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2}\left(\prod_{i=1}^{n}\prod_{j=1}^{p}(\boldsymbol{\beta}_{1,j}z_i+\boldsymbol{\beta}_{2,j}(1-z_i))^{\alpha_{i,j}}\right)\\
&\cdot\prod_{i=1}^{n}\binom{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|}{2\xi_i}\frac{(2\xi_i)!}{\xi_i!}\left(\frac{-\|\boldsymbol{\beta}\|_2^2}{2}\right)^{\xi_i}\mathop{\mathbb{E}}_{w\sim\mathcal{N}(0,\|\boldsymbol{\beta}\|_2^2)}w^{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|-2\xi_i},
\end{aligned}$$

其中 $\sum_{j=1}^{p}X_{i,j}\boldsymbol{\beta}_{1,j}z_i+\sum_{j=1}^{p}X_{i,j}\boldsymbol{\beta}_{2,j}(1-z_i)\sim\mathcal{N}(0,\|\boldsymbol{\beta}\|_2^2)$，无论边缘地还是在给定 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{z}$ 的条件下都成立，因此与 $\prod_{i=1}^{n}\prod_{j=1}^{p}(\boldsymbol{\beta}_{1,j}z_i+\boldsymbol{\beta}_{2,j}(1-z_i))^{\alpha_{i,j}}$ 独立（因为根据我们的先验，$\boldsymbol{\beta}_1,\boldsymbol{\beta}_2$ 的范数被约束为 $\|\boldsymbol{\beta}\|_2$）。交换求和与乘积后，结合高斯矩的已知公式 $\mathbb{E}_{w\sim\mathcal{N}(0,\|\boldsymbol{\beta}\|_2^2)}\,w^b=(b-1)!!\,\|\boldsymbol{\beta}\|_2^b\,\mathbb{I}\{b\,\mathsf{even}\}$，再进行额外的阶乘化简，并应用引理 B.6，我们得到

$$\begin{aligned}
&\langle \mathsf{L},\tilde{H}_{\boldsymbol{\alpha}}\rangle\\
&=\left(\frac{\frac{1}{\sigma}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1}}\right)^{|\boldsymbol{\alpha}_{p+1}|}\frac{\boldsymbol{\alpha}_{p+1}!}{(\boldsymbol{\alpha}_{p+1}-|\boldsymbol{\alpha}_{\cdot,:p}|)!}\sum_{0\le\boldsymbol{\xi}\le\lfloor\frac{\boldsymbol{\alpha_{i,p+1}}-|\boldsymbol{\alpha_{i,:p}}|}{\mathbf{2}}\rfloor}\\
&\cdot\left(\prod_{i=1}^{n}\binom{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|}{2\xi_i}\frac{(2\xi_i)!}{\xi_i!}\left(\frac{-\|\boldsymbol{\beta}\|_2^2}{2}\right)^{\xi_i}\right)\mathop{\mathbb{E}}_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2}\left(\prod_{i=1}^{n}\prod_{j=1}^{p}(\boldsymbol{\beta}_{1,j}z_i+\boldsymbol{\beta}_{2,j}(1-z_i))^{\alpha_{i,j}}\right)\\
&\cdot\left(\prod_{i=1}^{n}(\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|-2\xi_i-1)!!\cdot\|\boldsymbol{\beta}\|_2^{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|-2\xi_i}\,\mathbb{I}\{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|-2\xi_i\text{ even}\}\right)\\
&=\left(\frac{\frac{1}{\sigma}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1}}\right)^{|\boldsymbol{\alpha}_{p+1}|}\frac{\boldsymbol{\alpha}_{p+1}!}{(\boldsymbol{\alpha}_{p+1}-|\boldsymbol{\alpha}_{\cdot,:p}|)!}\|\boldsymbol{\beta}\|_2^{|\boldsymbol{\alpha}_{p+1}|-|\boldsymbol{\alpha}_{\cdot,:p}|}\\
&\cdot\prod_{i=1}^{n}\sum_{\xi}^{\lfloor\frac{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|}{2}\rfloor}\binom{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|}{2\xi}\frac{(2\xi)!}{\xi!}\left(\frac{-1}{2}\right)^{\xi}\\
&\cdot(\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|-2\xi-1)!!\cdot\,\mathbb{I}\{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|-2\xi\text{ even}\}\mathop{\mathbb{E}}_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2}\left(\prod_{i=1}^{n}\prod_{j=1}^{p}(\boldsymbol{\beta}_{1,j}z_i+\boldsymbol{\beta}_{2,j}(1-z_i))^{\alpha_{i,j}}\right)
\end{aligned}$$

$$= \left(\frac{\frac{1}{\sigma}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1}}\right)^{|\boldsymbol{\alpha}_{p+1}|} \frac{\boldsymbol{\alpha}_{p+1}!}{(\boldsymbol{\alpha}_{p+1}-|\boldsymbol{\alpha}_{\cdot,:p}|)!}\|\boldsymbol{\beta}\|_2^{|\boldsymbol{\alpha}_{p+1}|-|\boldsymbol{\alpha}_{\cdot,:p}|}$$
$$\cdot \prod_{i=1}^{n} \mathbb{I}^{\{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|=0\}} \mathop{\mathbb{E}}_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2}\left[\prod_{i=1}^{n}\prod_{j=1}^{p}(\boldsymbol{\beta}_{1,j}z_i+\boldsymbol{\beta}_{2,j}(1-z_i))^{\alpha_{i,j}}\right]$$
$$= \left(\frac{\frac{1}{\sigma}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1}}\right)^{|\boldsymbol{\alpha}_{p+1}|} \|\boldsymbol{\beta}\|_2^{|\boldsymbol{\alpha}_{p+1}|-|\boldsymbol{\alpha}_{\cdot,:p}|}\boldsymbol{\alpha}_{p+1}!\prod_{i=1}^{n} \mathbb{I}^{\{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{\cdot,:p}|=0\}} \mathop{\mathbb{E}}_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2}\left[\prod_{i=1}^{n}\prod_{j=1}^{p}(\boldsymbol{\beta}_{1,j}z_i+\boldsymbol{\beta}_{2,j}(1-z_i))^{\alpha_{i,j}}\right],$$

由此可得所需结果。 □

引理 B.9. 令 $(\boldsymbol{\beta}_1^{(1)},\boldsymbol{\beta}_2^{(1)})$ 和 $(\boldsymbol{\beta}_1^{(2)},\boldsymbol{\beta}_2^{(2)})$ 为从$\mathcal{P}_{\|\boldsymbol{\beta}\|_2}(\mathcal{D})$ 中抽样得到的信号的两个独立副本，并令 $\boldsymbol{z}^{(1)}$ 和 $\boldsymbol{z}^{(2)}$ 的各个元素分别从伯努利分布(ϕ) 中抽样。则有

$$\chi^2_{\le D}(\mathbb{P}(\boldsymbol{X},\boldsymbol{y})\|\mathbb{P}(\boldsymbol{X})\otimes\mathbb{P}(\boldsymbol{y}))+1 = \mathop{\mathbb{E}}_{\substack{(\boldsymbol{\beta}_1^{(1)},\boldsymbol{\beta}_2^{(1)}),(\boldsymbol{\beta}_1^{(2)},\boldsymbol{\beta}_2^{(2)})\overset{i.i.d.}{\sim}\mathcal{P}\\ z^{(1)},z^{(2)}\overset{i.i.d.}{\sim}Ber(\phi)}} \sum_{\frac{d}{2}=0}^{\lfloor\frac{D}{2}\rfloor}\left(\frac{1}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right)^{\frac{d}{2}}$$
$$\cdot \sum_{|\boldsymbol{\alpha}_{p+1}|=\frac{d}{2}}\prod_{i=1}^{n}\langle\boldsymbol{\beta}_1^{(1)}z_i^{(1)}+\boldsymbol{\beta}_2^{(1)}(1-z_i^{(1)}),\boldsymbol{\beta}_1^{(2)}z_i^{(2)}+\boldsymbol{\beta}_2^{(2)}(1-z_i^{(2)})\rangle^{\alpha_{i,p+1}}.$$

证明. 我们首先应用引理 B.8 来开始证明:

$$\begin{aligned}
&\chi^2_{\le D}(\mathbb{P}(\boldsymbol{X},\boldsymbol{y})\|\mathbb{P}(\boldsymbol{X})\otimes\mathbb{P}(\boldsymbol{y}))+1\\
&=\sum_{0\le|\boldsymbol{\alpha}|\le D}\frac{1}{\boldsymbol{\alpha}!}\langle\mathsf{L},\tilde{H}_{\boldsymbol{\alpha}}\rangle^2\\
&=\sum_{0\le|\boldsymbol{\alpha}|\le D}\frac{1}{\boldsymbol{\alpha}!}\left(\frac{\frac{1}{\sigma}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1}}\right)^{2|\boldsymbol{\alpha}_{p+1}|}\|\boldsymbol{\beta}\|_2^{2|\boldsymbol{\alpha}_{p+1}|-2|\boldsymbol{\alpha}_{\cdot,:p}|}(\boldsymbol{\alpha}_{p+1}!)^2\\
&\cdot\prod_{i=1}^{n}\mathbb{I}^{\{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{\cdot,:p}|=0\}}\left(\mathbb{E}\prod_{i=1}^{n}\prod_{j=1}^{p}(\boldsymbol{\beta}_{1,j}z_i+\boldsymbol{\beta}_{2,j}(1-z_i))^{\alpha_{i,j}}\right)^2\\
&=\sum_{d=0}^{D}\sum_{h=0}^{d}\sum_{|\boldsymbol{\alpha}_{p+1}|=h}\sum_{|\boldsymbol{\alpha}_{\cdot,:p}|=d-h}\frac{1}{\boldsymbol{\alpha}!}\left(\frac{\frac{1}{\sigma}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1}}\right)^{2|\boldsymbol{\alpha}_{p+1}|}\|\boldsymbol{\beta}\|_2^{2|\boldsymbol{\alpha}_{p+1}|-2|\boldsymbol{\alpha}_{\cdot,:p}|}(\boldsymbol{\alpha}_{p+1}!)^2\\
&\cdot\prod_{i=1}^{n}\mathbb{I}^{\{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{\cdot,:p}|=0\}}\left(\mathbb{E}\prod_{i=1}^{n}\prod_{j=1}^{p}(\boldsymbol{\beta}_{1,j}z_i+\boldsymbol{\beta}_{2,j}(1-z_i))^{\alpha_{i,j}}\right)^2.
\end{aligned}$$

接下来，我们将一个平方期望拆分为两个独立随机变量乘积的期望：$(\mathbb{E}_w[w])^2=\mathbb{E}_{w^{(1)}}\left[w^{(1)}\right]\mathbb{E}_{w^{(2)}}\left[w^{(2)}\right]=\mathbb{E}_{w^{(1)},\ w^{(2)}}\left[w^{(1)}w^{(2)}\right]$，其中我们选取 $w^{(1)}$ 和 $w^{(2)}$ 作为两个独立同分布的随机变量。继续这样做，我们得到:

$$
\begin{aligned}
&\chi^2_{\le D}(\mathbb{P}(\boldsymbol{X},\boldsymbol{y})\|\mathbb{P}(\boldsymbol{X})\otimes\mathbb{P}(\boldsymbol{y}))+1\\
&=\sum_{d=0}^{D}\sum_{h=0}^{d}\sum_{|\boldsymbol{\alpha}_{p+1}|=h}\sum_{|\boldsymbol{\alpha}_{\cdot,:p}|=d-h}\frac{1}{\boldsymbol{\alpha}!}\left(\frac{\frac{1}{\sigma}}{\sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1}}\right)^{2|\boldsymbol{\alpha}_{p+1}|}\|\boldsymbol{\beta}\|_2^{2|\boldsymbol{\alpha}_{p+1}|-2|\boldsymbol{\alpha}_{\cdot,:p}|}(\boldsymbol{\alpha}_{p+1}!)^2\\
&\cdot\underbrace{\prod_{i=1}^{n}\mathbb{I}^{\{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|=0\}}}_{\Longrightarrow\sum_{i=1}^n\alpha_{i,p+1}-|\boldsymbol{\alpha}_{\boldsymbol{i},:\boldsymbol{p}}|=0\Longrightarrow 2h-d=0\Longrightarrow d\text{ even}}\\
&\cdot\mathop{\mathbb{E}}_{\substack{(\boldsymbol{\beta}_1^{(1)},\boldsymbol{\beta}_2^{(1)}),(\boldsymbol{\beta}_1^{(2)},\boldsymbol{\beta}_2^{(2)})\overset{\text{i.i.d.}}{\sim}\mathcal{P}\\ z^{(1)},z^{(2)}\overset{\text{i.i.d.}}{\sim}\mathrm{Ber}(\phi)}}\prod_{i=1}^{n}\prod_{j=1}^{p}\left(\boldsymbol{\beta}_{1,j}^{(1)}z_i^{(1)}+\boldsymbol{\beta}_{2,j}^{(1)}(1-z_i^{(1)})\right)^{\alpha_{i,j}}\left(\boldsymbol{\beta}_{1,j}^{(2)}z_i^{(2)}+\boldsymbol{\beta}_{2,j}^{(2)}(1-z_i^{(2)})\right)^{\alpha_{i,j}},
\end{aligned}
$$

注意到 $\sum_{i=1}^n\alpha_{i,\ p+1}-|\boldsymbol{\alpha}_{\boldsymbol{i},\ :\boldsymbol{p}}|=0$ 意味着 d 必须是偶数，简化后得到

$$
\begin{aligned}
&\chi^2_{\le D}(\mathbb{P}(\boldsymbol{X},\boldsymbol{y})\|\mathbb{P}(\boldsymbol{X})\otimes\mathbb{P}(\boldsymbol{y}))+1\\
&=\mathbb{E}\sum_{\frac{d}{2}=0}^{\lfloor\frac{D}{2}\rfloor}\sum_{|\boldsymbol{\alpha}_{p+1}|=\frac{d}{2}}\sum_{|\boldsymbol{\alpha}_{\cdot,:p}|=\frac{d}{2}}\frac{1}{\boldsymbol{\alpha}_{p+1}!\cdot\boldsymbol{\alpha}_{\cdot,:p}!}\left(\frac{1}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right)^{\frac{d}{2}}\|\boldsymbol{\beta}\|_2^0(\boldsymbol{\alpha}_{p+1}!)^2\\
&\quad\cdot\prod_{i=1}^{n}\mathbb{I}^{\{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|=0\}}\\
&\quad\cdot\prod_{i=1}^{n}\prod_{j=1}^{p}\left(\boldsymbol{\beta}_{1,j}^{(1)}z_i^{(1)}+\boldsymbol{\beta}_{2,j}^{(1)}(1-z_i^{(1)})\right)^{\alpha_{i,j}}\left(\boldsymbol{\beta}_{1,j}^{(2)}z_i^{(2)}+\boldsymbol{\beta}_{2,j}^{(2)}(1-z_i^{(2)})\right)^{\alpha_{i,j}}\\
&=\mathbb{E}\sum_{\frac{d}{2}=0}^{\lfloor\frac{D}{2}\rfloor}\left(\frac{1}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right)^{\frac{d}{2}}\\
&\quad\cdot\sum_{|\boldsymbol{\alpha}_{p+1}|=\frac{d}{2}}\frac{(\boldsymbol{\alpha}_{p+1}!)^2}{\boldsymbol{\alpha}_{p+1}!}\sum_{|\boldsymbol{\alpha}_{\cdot,:p}|=\frac{d}{2}}\frac{1}{\boldsymbol{\alpha}_{\cdot,:p}!}\prod_{i=1}^{n}\mathbb{I}^{\{\alpha_{i,p+1}-|\boldsymbol{\alpha}_{i,:p}|=0\}}\\
&\quad\cdot\prod_{i=1}^{n}\prod_{j=1}^{p}\left(\boldsymbol{\beta}_{1,j}^{(1)}z_i^{(1)}+\boldsymbol{\beta}_{2,j}^{(1)}(1-z_i^{(1)})\right)^{\alpha_{i,j}}\left(\boldsymbol{\beta}_{1,j}^{(2)}z_i^{(2)}+\boldsymbol{\beta}_{2,j}^{(2)}(1-z_i^{(2)})\right)^{\alpha_{i,j}},
\end{aligned}
$$

并可以重新排序，以便更清楚地应用多项式定理:

$$
\begin{aligned}
&\chi^2_{\le D}(\mathbb{P}(\boldsymbol{X},\boldsymbol{y})\|\mathbb{P}(\boldsymbol{X})\otimes\mathbb{P}(\boldsymbol{y}))+1\\
&=\mathbb{E}\sum_{\frac{d}{2}=0}^{\lfloor\frac{D}{2}\rfloor}\left(\frac{1}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right)^{\frac{d}{2}}\sum_{|\boldsymbol{\alpha}_{p+1}|=\frac{d}{2}}\boldsymbol{\alpha}_{p+1}!\\
&\quad\cdot\sum_{|\boldsymbol{\alpha}_{1,:p}|=\alpha_{1,p+1}}\frac{1}{\boldsymbol{\alpha}_{1,:p}!}\cdots\sum_{|\boldsymbol{\alpha}_{n,:p}|=\alpha_{n,p+1}}\frac{1}{\boldsymbol{\alpha}_{n,:p}!}\\
&\quad\prod_{i=1}^{n}\prod_{j=1}^{p}\left(\boldsymbol{\beta}_{1,j}^{(1)}z_i^{(1)}+\boldsymbol{\beta}_{2,j}^{(1)}(1-z_i^{(1)})\right)^{\alpha_{i,j}}\left(\boldsymbol{\beta}_{1,j}^{(2)}z_i^{(2)}+\boldsymbol{\beta}_{2,j}^{(2)}(1-z_i^{(2)})\right)^{\alpha_{i,j}}
\end{aligned}
$$

$$= \mathbb{E} \sum_{\frac{d}{2}=0}^{\lfloor \frac{D}{2} \rfloor} \left(\frac{1}{\|\boldsymbol{\beta}\|_2^2 + \sigma^2} \right)^{\frac{d}{2}} \sum_{|\boldsymbol{\alpha}_{p+1}|=\frac{d}{2}} \boldsymbol{\alpha}_{p+1}! \cdot \sum_{|\boldsymbol{\alpha}_{1,:p}|=\alpha_{1,p+1}} \frac{\prod_{j=1}^{p} \left(\boldsymbol{\beta}_{1,j}^{(1)} z_i^{(1)} + \boldsymbol{\beta}_{2,j}^{(1)} (1 - z_i^{(1)}) \right)^{\alpha_{1,j}} \left(\boldsymbol{\beta}_{1,j}^{(2)} z_i^{(2)} + \boldsymbol{\beta}_{2,j}^{(2)} (1 - z_i^{(2)}) \right)^{\alpha_{1,j}}}{\boldsymbol{\alpha}_{1,:p}!} \cdots \sum_{|\boldsymbol{\alpha}_{n,:p}|=\alpha_{n,p+1}} \frac{\prod_{j=1}^{p} \left(\boldsymbol{\beta}_{1,j}^{(1)} z_i^{(1)} + \boldsymbol{\beta}_{2,j}^{(1)} (1 - z_i^{(1)}) \right)^{\alpha_{n,j}} \left(\boldsymbol{\beta}_{1,j}^{(2)} z_i^{(2)} + \boldsymbol{\beta}_{2,j}^{(2)} (1 - z_i^{(2)}) \right)^{\alpha_{n,j}}}{\boldsymbol{\alpha}_{n,:p}!}.$$

然后，我们应用多项式定理得到结果：

$$\chi^2_{\leq D}(\mathbb{P}(\boldsymbol{X}, \boldsymbol{y}) \| \mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})) + 1 = \mathbb{E} \sum_{\frac{d}{2}=0}^{\lfloor \frac{D}{2} \rfloor} \left(\frac{1}{\|\boldsymbol{\beta}\|_2^2 + \sigma^2} \right)^{\frac{d}{2}} \sum_{|\boldsymbol{\alpha}_{p+1}|=\frac{d}{2}} \prod_{i=1}^{n} \langle \boldsymbol{\beta}_1^{(1)} z_i^{(1)} + \boldsymbol{\beta}_2^{(1)} (1 - z_i^{(1)}), \boldsymbol{\beta}_1^{(2)} z_i^{(2)} + \boldsymbol{\beta}_2^{(2)} (1 - z_i^{(2)}) \rangle^{\alpha_{i,p+1}}.$$

□

在接下来的两个小节中，我们通过特化引理 SLR − D (定理2.8 和 SB-MSLR − D (定理 2.2) 来证明其计算下界。B.9。

B.2 特殊情形： SLR − D

定理 2.8 的证明。根据下文的定理 B.10，回忆信噪比 $= \|\boldsymbol{\beta}\|_2^2/\sigma^2$，令 $n = (1-\epsilon)(1-2\theta) \frac{\|\boldsymbol{\beta}\|_2^2}{\boldsymbol{\beta}\|_\infty^2} \frac{\mathtt{SNR}+1}{\mathtt{SNR}}$ log p。于是，定理 B.10 表明，对于所有 $D \leq \frac{2\epsilon}{1-\epsilon} n$，都有 $\chi^2_{\leq D}(\mathbb{P}(\boldsymbol{X}, \boldsymbol{y}) \| \mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})) = O(1)$。应用猜想 1.2，并回忆 $n = \omega(\log\ p)$（根据定理的假设），可知所需运行时间为 exp $\left(\tilde{\Omega}(n) \right)$。

□

定理 B.10（一般 SLR − D 下界）。考虑 SLR − D 的情形（定义 2.1，其中 $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$）。设 $\boldsymbol{\beta} \sim \mathcal{P}_{\|\boldsymbol{\beta}\|_2}(\mathcal{D})$。若 $k = O(p^\theta) \leq \sqrt{p}$ 对某个 $\theta \in (0, 1/2]$，则对任意 $\epsilon \in (0, 1)$，$n \leq (1-\epsilon)(1-2\theta) \left(\frac{\|\boldsymbol{\beta}\|_2^2 + \sigma^2}{\|\boldsymbol{\beta}\|_\infty^2} \right) \log\ p$ 且 $D \leq \frac{2\epsilon}{1-\epsilon} n$，我们有 $\chi^2_{\leq D}(\mathbb{P}(\boldsymbol{X}, \boldsymbol{y}) \| \mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})) = O(1)$。

证明。令 $\mathcal{S}^{(1)}$ 和 $\mathcal{S}^{(2)}$ 分别表示 $\boldsymbol{\beta}^{(1)}$ 和 $\boldsymbol{\beta}^{(2)}$的支撑集。我们应用引理B.9得到：

$$\begin{aligned}
\chi^2_{\leq D}(\mathbb{P}(\boldsymbol{X}, \boldsymbol{y}) \| \mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})) + 1 &= \mathbb{E} \sum_{\frac{d}{2}=0}^{\lfloor \frac{D}{2} \rfloor} \left(\frac{1}{\|\boldsymbol{\beta}\|_2^2 + \sigma^2} \right)^{\frac{d}{2}} \sum_{|\boldsymbol{\alpha}_{p+1}|=\frac{d}{2}} \prod_{i=1}^{n} \langle \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)} \rangle^{\alpha_{i,p+1}} \\
&= \mathbb{E} \sum_{\frac{d}{2}=0}^{\lfloor \frac{D}{2} \rfloor} \left(\frac{1}{\|\boldsymbol{\beta}\|_2^2 + \sigma^2} \right)^{\frac{d}{2}} \sum_{|\boldsymbol{\alpha}_{p+1}|=\frac{d}{2}} \langle \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)} \rangle^{\frac{d}{2}} \\
&\leq \mathbb{E} \sum_{\frac{d}{2}=0}^{\lfloor \frac{D}{2} \rfloor} \left(\frac{1}{\|\boldsymbol{\beta}\|_2^2 + \sigma^2} \right)^{\frac{d}{2}} \binom{\frac{D}{2} + n - 1}{n - 1} \langle \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)} \rangle^{\frac{d}{2}}
\end{aligned}$$

$$
\begin{aligned}
&\leq \mathbb{E} \sum_{\frac{d}{2}=0}^{\lfloor \frac{D}{2} \rfloor} \left(\frac{1}{\|\boldsymbol{\beta}\|_2^2 + \sigma^2} \right)^{\frac{d}{2}} \frac{\left(\frac{D}{2}+n\right)^{\frac{d}{2}}}{\frac{d}{2}!} \langle \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)} \rangle^{\frac{d}{2}} \\
&\leq \mathop{\mathbb{E}}_{\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)} \overset{\text{i.i.d.}}{\sim} \mathcal{P}} \exp\left(\frac{\frac{1}{\sigma^2}}{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}+1} \left(\frac{D}{2}+n\right) \langle \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)} \rangle \right) \\
&\leq \mathop{\mathbb{E}}_{\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)} \overset{\text{i.i.d.}}{\sim} \mathcal{P}} \exp\left(\frac{\langle \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)} \rangle}{\|\boldsymbol{\beta}\|_\infty^2} (1-2\theta) \log p \right).
\end{aligned}
$$

接下来，我们应用引理 B.7 并注意到 $\langle \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)} \rangle \leq \|\boldsymbol{\beta}\|_\infty^2 |\mathcal{S}^{(1)} \cap \mathcal{S}^{(2)}|$ 从而对 $p > 4$ 和 $k \leq p$ 得到：

$$
\begin{aligned}
\chi^2_{\leq D}(\mathbb{P}(\boldsymbol{X}, \boldsymbol{y}) \| \mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})) + 1 &= \mathop{\mathbb{E}}_{\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)} \overset{\text{i.i.d.}}{\sim} \mathcal{P}} \exp\left(\frac{\langle \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)} \rangle}{\|\boldsymbol{\beta}\|_\infty^2} (1-2\theta) \log p \right) \\
&\leq \mathop{\mathbb{E}}_{\langle \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)} \rangle} \exp\left(|\mathcal{S}^{(1)} \cap \mathcal{S}^{(2)}| (1-2\theta) \log p \right) \\
&\leq \sum_{l=0}^{k} \frac{\binom{k}{l}\binom{p-k}{k-l}}{\binom{p}{k}} \exp\left(l(1-2\theta) \log p \right) \\
&\leq \sum_{l=0}^{k} \frac{4k!}{p^k} \frac{k^l}{l!} \frac{(p-k)^{k-l}}{(k-l)!} \exp\left(l(1-2\theta) \log p \right) \\
&\leq 4 \sum_{l=0}^{k} \left(\frac{k^2}{p} \right)^l \exp\left(l(1-2\theta) \log p \right) \\
&= 4 \sum_{l=0}^{k} \left(\frac{k^2}{p^{2\theta}} \right)^l = O(1).
\end{aligned}
$$

□

B.3 特殊情形： SB-MSLR − D

定理 2.2 的证明。该结论由下文的定理 B.12 推出。任取满足 $n \geq k, n = \omega(\log p)$ 的样本量 n，以及 $n = o((\|\boldsymbol{\beta}\|_2^2+\sigma^2)^2/(\|\boldsymbol{\beta}\|_\infty^4 \log p))$，则对于 $D = (\sqrt{2}-1) \min\left\{ \frac{(\|\boldsymbol{\beta}\|_2^2+\sigma^2)^2}{n\|\boldsymbol{\beta}\|_\infty^4}, n \right\}$，有 $\chi^2_{\leq D}(\mathbb{P}(\boldsymbol{X}, \boldsymbol{y}) \| \mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})) = O(1)$。接着，我们调用猜想 1.2 并使用 $\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2} =$ ~~信噪比~~；注意，对于有界幅度的信号，我们有 $k\|\boldsymbol{\beta}\|_\infty/\sigma^2 \gtrsim \sqrt{k\|\boldsymbol{\beta}\|_\infty/\sigma^2} \gtrsim$ 信噪比。 □

引理 B.11。对于 SB-MSLR − D，我们有：

$$
\chi^2_{\leq D}(\mathbb{P}(\boldsymbol{X}, \boldsymbol{y}) \| \mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})) + 1 \leq \mathop{\mathbb{E}}_{\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}} \sum_{\frac{d}{4}=0}^{\lfloor \frac{D}{4} \rfloor} \left(\frac{1}{\|\boldsymbol{\beta}\|_2^2 + \sigma^2} \right)^{\frac{d}{2}} \frac{\left(\frac{D}{4}+n\right)^{\frac{d}{4}}}{\frac{d}{4}!} \langle \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)} \rangle^{\frac{d}{2}},
$$

其中，$\boldsymbol{\beta}^{(1)}$、$\boldsymbol{\beta}^{(2)}$ 是随机变量 $\boldsymbol{\beta} \overset{d}{=} \boldsymbol{\beta}_1$ 的两个独立副本。

证明。 我们首先将假设应用于引理 B.9 ，并利用z_i之间的独立性。注意，在 SB-MSLR的情形下，特别有 $\phi\boldsymbol{\beta}_1 +(1-\phi)\boldsymbol{\beta}_2 = 0$，因此我们代入 $\boldsymbol{\beta}:=\boldsymbol{\beta}_1 = -\frac{\phi}{1-\phi}\boldsymbol{\beta}_2$。

$$
\begin{aligned}
&\chi^2_{\le D}(\mathbb{P}(\boldsymbol{X},\boldsymbol{y})\|\mathbb{P}(\boldsymbol{X})\otimes\mathbb{P}(\boldsymbol{y}))+1\\
&\le \mathbb{E}\sum_{\frac{d}{2}=0}^{\lfloor\frac{D}{2}\rfloor}\left(\frac{1}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right)^{\frac{d}{2}}\sum_{|\boldsymbol{\alpha}_{p+1}|=\frac{d}{2}}\prod_{i=1}^{n}\langle\boldsymbol{\beta}_1^{(1)}z_i^{(1)}+\boldsymbol{\beta}_2^{(1)}(1-z_i^{(1)}),\boldsymbol{\beta}_1^{(2)}z_i^{(2)}+\boldsymbol{\beta}_2^{(2)}(1-z_i^{(2)})\rangle^{\alpha_{i,p+1}}\\
&=\mathop{\mathbb{E}}_{\boldsymbol{\beta}}\sum_{\frac{d}{2}=0}^{\lfloor\frac{D}{2}\rfloor}\left(\frac{1}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right)^{\frac{d}{2}}\sum_{|\boldsymbol{\alpha}_{p+1}|=\frac{d}{2}}\cdot\prod_{i=1}^{n}\Big(\phi^2\langle\boldsymbol{\beta}_1^{(1)},\boldsymbol{\beta}_1^{(2)}\rangle^{\alpha_{i,p+1}}+\phi(1-\phi)\langle\boldsymbol{\beta}_1^{(1)},\boldsymbol{\beta}_2^{(2)}\rangle^{\alpha_{i,p+1}}\\
&\qquad+\phi(1-\phi)\langle\boldsymbol{\beta}_2^{(1)},\boldsymbol{\beta}_1^{(2)}\rangle^{\alpha_{i,p+1}}+(1-\phi)^2\langle\boldsymbol{\beta}_2^{(1)},\boldsymbol{\beta}_1^{(2)}\rangle^{\alpha_{i,p+1}}\Big)\\
&=\mathop{\mathbb{E}}_{\boldsymbol{\beta}}\sum_{\frac{d}{2}=0}^{\lfloor\frac{D}{2}\rfloor}\left(\frac{1}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right)^{\frac{d}{2}}\sum_{|\boldsymbol{\alpha}_{p+1}|=\frac{d}{2}}\\
&\quad\cdot\prod_{i=1}^{n}\left(\phi^2+2\phi(1-\phi)\left(-\frac{\phi}{1-\phi}\right)^{\alpha_{i,p+1}}+(1-\phi)^2\left(\frac{\phi}{1-\phi}\right)^{2\alpha_{i,p+1}}\right)\langle\boldsymbol{\beta}^{(1)},\boldsymbol{\beta}^{(2)}\rangle^{\alpha_{i,p+1}}\\
&=\mathop{\mathbb{E}}_{\boldsymbol{\beta}}\sum_{\frac{d}{2}=0}^{\lfloor\frac{D}{2}\rfloor}\left(\frac{1}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right)^{\frac{d}{2}}\sum_{|\boldsymbol{\alpha}_{p+1}|=\frac{d}{2}}\prod_{i=1}^{n}\left(\phi+(1-\phi)\left(-\frac{\phi}{1-\phi}\right)^{\alpha_{i,p+1}}\right)^2\langle\boldsymbol{\beta}^{(1)},\boldsymbol{\beta}^{(2)}\rangle^{\alpha_{i,p+1}}.
\end{aligned}
$$

现在我们注意到，当且仅当$\phi = 1/2$时，乘积中的项对所有$\alpha_{i、\ p+1}$ 奇数都等于零，而 SB-MSLR 正是这种情况。因此，我们只对偶数项求和，得到:

$$
\begin{aligned}
&\chi^2_{\le D}(\mathbb{P}(\boldsymbol{X},\boldsymbol{y})\|\mathbb{P}(\boldsymbol{X})\otimes\mathbb{P}(\boldsymbol{y}))+1\\
&=\mathop{\mathbb{E}}_{\boldsymbol{\beta}}\sum_{\substack{\frac{d}{2}=0\\ \text{even}}}^{\lfloor\frac{D}{2}\rfloor}\left(\frac{1}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right)^{\frac{d}{2}}\sum_{\substack{|\boldsymbol{\alpha}_{p+1}|=\frac{d}{2}\\ \text{even}}}\prod_{i=1}^{n}\langle\boldsymbol{\beta}^{(1)},\boldsymbol{\beta}^{(2)}\rangle^{\alpha_{i,p+1}}\\
&=\mathop{\mathbb{E}}_{\boldsymbol{\beta}}\sum_{\frac{d}{4}=0}^{\lfloor\frac{D}{4}\rfloor}\left(\frac{1}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right)^{\frac{d}{2}}\sum_{\substack{|\boldsymbol{\alpha}_{p+1}|=\frac{d}{2}\\ \text{even}}}\langle\boldsymbol{\beta}^{(1)},\boldsymbol{\beta}^{(2)}\rangle^{|\boldsymbol{\alpha}_{p+1}|}\\
&=\mathop{\mathbb{E}}_{\boldsymbol{\beta}}\sum_{\frac{d}{4}=0}^{\lfloor\frac{D}{4}\rfloor}\left(\frac{1}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right)^{\frac{d}{2}}\binom{\frac{d}{4}+n-1}{n-1}\langle\boldsymbol{\beta}^{(1)},\boldsymbol{\beta}^{(2)}\rangle^{\frac{d}{2}}\\
&\le\mathop{\mathbb{E}}_{\boldsymbol{\beta}}\sum_{\frac{d}{4}=0}^{\lfloor\frac{D}{4}\rfloor}\left(\frac{1}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right)^{\frac{d}{2}}\frac{\left(\frac{D}{4}+n\right)^{\frac{d}{4}}}{\frac{d}{4}!}\langle\boldsymbol{\beta}^{(1)},\boldsymbol{\beta}^{(2)}\rangle^{\frac{d}{2}}.
\end{aligned}
$$

□

定理 B.12（一般 SB-MSLR − D 下界）。考虑具有联合先验 $\mathcal{P}_{\|\boldsymbol{\beta}\|_2}(\mathcal{D})$ 的 SB-MSLR − D 情形。若 $k \le \sqrt{\frac{p}{e}}$，且 $k \le D \le 2(\sqrt{2}-1)\min\left\{\frac{(\|\boldsymbol{\beta}\|_2^2+\sigma^2)^2}{n\|\boldsymbol{\beta}\|_\infty^4}, n\right\}$，则有$\chi^2_{\le D}(\mathbb{P}(\boldsymbol{X},\boldsymbol{y})\|\mathbb{P}(\boldsymbol{X})\otimes\mathbb{P}(\boldsymbol{y})) = O(1)$。

证明。 我们首先应用引理 B.9 的结果，得到：

$$\chi^2_{\le D}(\mathbb{P}(\boldsymbol{X},\boldsymbol{y})\|\mathbb{P}(\boldsymbol{X})\otimes\mathbb{P}(\boldsymbol{y}))+1 \le \sum_{\frac{d}{4}=0}^{\frac{D}{4}} \left(\frac{1}{(\|\boldsymbol{\beta}\|_2^2+\sigma^2)^2}\right)^{\frac{d}{4}} \frac{\left(\frac{D}{4}+n\right)^{\frac{d}{4}}}{\frac{d}{4}!} \mathop{\mathbb{E}}_{\boldsymbol{\beta}^{(1)},\boldsymbol{\beta}^{(2)}\overset{\text{i.i.d.}}{\sim}\mathcal{P}} \langle\boldsymbol{\beta}^{(1)},\boldsymbol{\beta}^{(2)}\rangle^{\frac{d}{2}}$$

$$\le \sum_{\frac{d}{4}=0}^{\frac{D}{4}} \left(\frac{1}{(\|\boldsymbol{\beta}\|_2^2+\sigma^2)^2}\right)^{\frac{d}{4}} \frac{\left(\frac{D}{4}+n\right)^{\frac{d}{4}}}{\frac{d}{4}!} \mathop{\mathbb{E}}_{\boldsymbol{\beta}^{(1)},\boldsymbol{\beta}^{(2)}\overset{\text{i.i.d.}}{\sim}\mathcal{P}} \langle\mathtt{abs}(\boldsymbol{\beta}^{(1)}),\mathtt{abs}(\boldsymbol{\beta}^{(2)})\rangle^{\frac{d}{2}},$$

其中 $\langle\boldsymbol{\beta}^{(1)},\ \boldsymbol{\beta}^{(2)}\rangle \le \langle\mathtt{abs}(\boldsymbol{\beta}^{(1)}),\ \mathtt{abs}(\boldsymbol{\beta}^{(2)})\rangle$，并且我们回顾一下 $\mathtt{abs}(\boldsymbol{\beta})$ 表示对向量 $\boldsymbol{\beta}$ 逐元素取绝对值的运算。注意，根据引理 B.7 我们有，对于 $p>4,\ k\le\sqrt{p}$，

$$\mathop{\mathbb{E}}_{\boldsymbol{\beta}^{(1)},\boldsymbol{\beta}^{(2)}\overset{\text{i.i.d.}}{\sim}\mathcal{P}} \langle\mathtt{abs}(\boldsymbol{\beta}^{(1)}),\mathtt{abs}(\boldsymbol{\beta}^{(2)})\rangle^{\frac{d}{2}} \le \sum_{l=0}^{k} \frac{\binom{k}{l}\binom{p-k}{k-l}}{\binom{p}{k}} (l\|\boldsymbol{\beta}\|_\infty^2)^{\frac{d}{2}}$$

$$\le 4\sum_{l=0}^{k}\left(\frac{k^2}{p}\right)^l (l\|\boldsymbol{\beta}\|_\infty^2)^{\frac{d}{2}}.$$

于是我们得到：

$$\chi^2_{\le D}(\mathbb{P}(\boldsymbol{X},\boldsymbol{y})\|\mathbb{P}(\boldsymbol{X})\otimes\mathbb{P}(\boldsymbol{y}))+1 \le 4\sum_{\frac{d}{4}=0}^{\frac{D}{4}} \left(\frac{1}{(\|\boldsymbol{\beta}\|_2^2+\sigma^2)^2}\right)^{\frac{d}{4}} \frac{\left(\frac{D}{4}+n\right)^{\frac{d}{4}}}{\frac{d}{4}!} \sum_{l=0}^{k}\left(\frac{k^2}{p}\right)^l (l\|\boldsymbol{\beta}\|_\infty^2)^{\frac{d}{2}}.$$

为了界定右侧，我们施加条件 i)：$(D/4+n)D \le (\|\boldsymbol{\beta}\|_2^2+\sigma^2)^2/\|\boldsymbol{\beta}\|_\infty^4$。交换求和顺序后，得到

$$\chi^2_{\le D}(\mathbb{P}(\boldsymbol{X},\boldsymbol{y})\|\mathbb{P}(\boldsymbol{X})\otimes\mathbb{P}(\boldsymbol{y}))+1 \le 4\sum_{l=0}^{k}\left(\frac{k^2}{p}\right)^l \sum_{\frac{d}{4}=0}^{\frac{D}{4}} \frac{\left(\frac{(\|\boldsymbol{\beta}\|_2^2+\sigma^2)^2\|\boldsymbol{\beta}\|_\infty^4}{D(\|\boldsymbol{\beta}\|_2^2+\sigma^2)^2\|\boldsymbol{\beta}\|_\infty^4}\right)^{\frac{d}{4}}}{\frac{d}{4}!} l^{\frac{d}{2}}$$

$$\le 4\sum_{l=0}^{k}\left(\frac{k^2}{p}\right)^l \exp\left(\frac{l^2}{D}\right).$$

现在我们施加条件 ii)：$k\le D$，从而得到结果，

$$\chi^2_{\le D}(\mathbb{P}(\boldsymbol{X},\boldsymbol{y})\|\mathbb{P}(\boldsymbol{X})\otimes\mathbb{P}(\boldsymbol{y}))+1 \le 4\sum_{l=0}^{k}\left(\frac{k^2}{p}\right)^l \exp(l)$$

$$= 4\sum_{l=0}^{k}\left(\frac{k^2 e}{p}\right)^l = O(1).$$

注意，对于任意 $n>0$ 和 $k\le D\le 2(\sqrt{2}-1)\min\left\{\frac{(\|\boldsymbol{\beta}\|_2^2+\sigma^2)}{n\|\boldsymbol{\beta}\|_\infty^4}^2, n\right\}$，条件 i) 和 ii) 均成立。

□

C 多项式时间归约的证明

考虑 MSLR 问题中的带符号支撑恢复：我们要恢复 $\boldsymbol{\beta}_1$和 $\boldsymbol{\beta}_2$的支撑，以及其中各元素的符号。取 $(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)\sim\mathcal{P}_{\|\beta\|_2}(\{-1,1\})$，并令

$\mathcal{S}_1 := \text{supp}(\boldsymbol{\beta}_1) = \{j \in [p] : \boldsymbol{\beta}_{1,\, j} \neq 0\}$，并对 $\boldsymbol{\beta}_2$作类似定义。我们研究当联合信号分布的两个参数——重叠 ξ和有符号重叠度 τ ——变化时，该问题的计算困难性：

$$\xi = \frac{|\mathcal{S}_1 \cap \mathcal{S}_2|}{k}, \qquad \tau = \frac{\langle \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \rangle}{|\mathcal{S}_1 \cap \mathcal{S}_2|},$$

为常数量级，即不随 n、 p、 k变化。先前工作[Gamarnik 和 Zadik,2022] 研究了稀疏线性回归中的精确支撑恢复，以及由两个相同信号副本的重叠分布所导致的计算困难性。我们通过改变参数 τ 扩展了这一分析，该参数衡量索引相同的 $+1$ 和 -1 条目之间的相对频率。注意，对于 $\xi = 1, \tau = 1$ ，我们得到通常的 SLR 问题；对于 $\xi = 1, \tau = -1$ ，我们处于 SB-MSLR 区域；而重要的是，对于 $\phi = 1/2, \tau \in (-1, 1)$ $\xi > 0$，我们处于 PSB-MSLR 区域。我们将具有联合信号先验 $\mathcal{P}_{\|\beta\|_2}(\{-1, 1\})$、且信号 $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ 的重叠度和有符号重叠度分别受限为 ξ 和 τ 的 MSLR$_{\xi,\, \tau}$和 MSLR $-$ D$_{\xi,\, \tau}$问题，分别记为 MSLR 和 MSLR $-$ D 问题。

利用这些定义，我们首先在引理 C.1 中构造一个多项式时间归约，将 SB-MSLR $-$ D 归约为 SB-MSLR 中的精确有符号支撑恢复，且处于定理 2.2 的尺度范围内。注意，MSLR $-$ D1,$_{-1}$ 等价于 SB-MSLR $-$ D 问题。接下来，我们证明一个多项式时间归约，将 SB-MSLR 中的精确有符号支撑恢复归约为 MSLR$_{\xi,\tau}$中针对 $\tau \in (-1, 1)$ $\xi > 0$ (PSB-MSLR) 的精确有符号支撑恢复，且处于定理 2.2 的尺度范围内。这证明了如果能够在 PSB-MSLR 中实现精确有符号支撑恢复，那么也能够在 SB-MSLR 中实现。

结合上述两项论证可知，解决PSB-MSLR 中的精确有符号支撑恢复问题SB-MSLR $-$ D中的强检测问题，这将与定理 2.2 所述的蕴含关系相矛盾： SB-MSLR $-$ D 无法在多项式时间内解决。因此得到定理 C.3。关于平均情形归约的逻辑背景，参见 [Brennan 和 Bresler, 2020b]。

从 SLR $-$ D 到 SLR 的归约与引理 C.1 中的归约几乎完全相同，只是条件稍微宽松一些，即 信噪比 ≥ 1, 因此省略证明。

在整个证明过程中，我们对线性回归混合模型使用以下恢复误差度量 [Chen等人, 2014]:

$$\rho((\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2), (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)) := \min \left\{ \left\| \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \right\|_2 + \left\| \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \right\|_2, \left\| \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_2 \right\|_2 + \left\| \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_1 \right\|_2 \right\}.$$

该误差度量考虑了重新标记后两个信号的恢复。对于向量 a、$\hat{}b$、$\hat{}a$、 $b \in \mathbb{R}^p$，定义

$$\|(\hat{a}, \hat{b}) - (a, b)\|_\infty := \min \left\{ \|\hat{a} - a\|_\infty + \|\hat{b} - b\|_\infty, \|\hat{b} - a\|_\infty + \|\hat{a} - b\|_\infty \right\}.$$

注意，对于 $\epsilon \in [0, 1)$ 以及信号（$\boldsymbol{\beta}_1$、 $\boldsymbol{\beta}_2$）$\sim \mathcal{P}_{\|\boldsymbol{\beta}\|_2}$（$\{-1, 1\}$），我们有

$$\mathbb{P}\left[\rho((\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2), (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)) > \epsilon\right] \to 0 \iff \mathbb{P}\left[\|(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2) - (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)\|_\infty > \epsilon\right] \to 0,$$

主要陈述 我们首先定义感兴趣的参数范围：

$$\mathcal{C}_1 = \Big\{ (p_i, n_i, k_i, \sigma_i)_{i=1}^\infty \subset \mathbb{N}^4 : p_i = \omega_i(1), k_i = o(\sqrt{p_i}), n_i = \omega(\max\{k_i, \log p_i\}), \\ n_i = o\left((k_i + \sigma_i^2)^2 \cdot \frac{1}{\log p_i} \right) \Big\}. \tag{7}$$

$$\mathcal{C}_2 = \Big\{(p_i, n_i, k_i, \sigma_i)_{i=1}^{\infty} \subset \mathbb{N}^4 : \; p_i = \omega_i(1), k_i = o(\sqrt{p_i}), n_i = \omega(\max\{k_i, \log p_i\}) \,, \\ n_i \gtrsim \frac{k_i \log p_i}{\log(1 + \frac{k_i}{\sigma_i^2})} \Big\} . \tag{8}$$

请注意， $\mathcal{C}_1$， $\mathcal{C}_2$ 二者都处于SB-MSLR $-$ D 面临计算障碍的参数范围内，这与定理2.2一致。以下引理由两个子归约组成，它们共同给出了从SB-MSLR $-$ D 到PSB-MSLR 中的精确恢复的归约论证。

引理 C.1. 设 $(\beta_1, \beta_2) \sim \mathcal{P}_{\|\beta\|_2}(\{-1, 1\})$ $SNR = \omega(1)$。给定一列参数 $\{(p_i, n_i, k_i, \sigma_i)\}_{i=1}^{\infty}$属于 $\mathcal{C}_2$ for SB-MSLR $-$ D 和 SB-MSLR，若对于任意 $\epsilon > 0$，均存在一个用于 SB-MSLR 的随机多项式时间算 $\mathcal{A}$ 法，输出 $(\hat{\beta}_1, \hat{\beta}_2)$，且满足 $\mathbb{P}\left[\|(\hat{\beta}_1, \hat{\beta}_2) - (\beta_1, \beta_2)\|_\infty < \epsilon\right] \to 1$，则 $(i \to \infty)$ 存在一个用于 SB-MSLR $-$ D 的随机多项式时间检测算法 $\mathcal{A}'$，其第一类+和第二类错误随 $i \to \infty$ 趋于零。

引理 C.1的证明见第 C.1节。

固定信号先验为 $\mathcal{P}_{\|\beta\|_2}(\{-1, 1\})$。对于任意一列参数 $\{(p_i', n_i', k_i', \sigma_i')\}_{i=1}^{\infty}$属于 $\mathcal{C}_1$ for PSB-MSLR，其解为 β_1', β_2'，问题实例为 $(\boldsymbol{X}', \boldsymbol{y}')$，存在一列参数 $\{(p_i, n_i, k_i, \sigma_i)\}_{i=1}^{\infty}$ 属于 $\mathcal{C}_1$ for SB-MSLR，其解为 β_1, β_2，问题实例为 $(\boldsymbol{X}, \boldsymbol{y})$，使得对于任意用于 PSB-MSLR 的随机多项式时间算法 $\mathcal{A}'$，若其输出 $(\hat{\beta'}_1, \hat{\beta'}_2)$，且满足

$$\mathbb{P}\left[\|(\hat{\beta}'_1, \hat{\beta}'_2) - (\beta_1', \beta_2')\|_\infty > 0\right] \to 0,$$

^ ^ 我们可以构造另一个用于 PSB-MSLR 的随机多项式时间算法 $\mathcal{A}$，输出 (β_1, β_2)，使得

$$\mathbb{P}\left[\|(\hat{\beta}_1, \hat{\beta}_2) - (\beta_1, \beta_2)\|_\infty > 0\right] \to 0.$$

引理 C.2的证明见第 C.2节。

定理 C.3 （从 SB-MSLR $-$ D 到 PSB-MSLR 中的精确恢复的归约）。考虑具有联合信号先验 PSB-MSLR (2) 的设置，其中联合信号先验为 $\mathcal{P}_{\|\beta\|_2}(\{-1, 1\})$。任何在参数范围 $\mathcal{A}$ 内求解 PSB-MSLR 的随机多项式时间算法 $\mathcal{C}_1 \cap \mathcal{C}_2$ 且满足 $SNR = \omega(1)$，都会与定理 2.2相矛盾。

证明。假设存在一个随机多项式时间算法 $\mathcal{A}$，用于求解PSB-MSLR 中的精确恢复， $\{-1, 0, 1\}^p$ 其信号属于 $\mathcal{C}_1$，且参数范围包含于式（7）所定义的范围内。于是，根据引理C.2，我们将得到一个随机多项式时间算法 $\mathcal{A}'$，用于求解SB-MSLR 中的精确恢复，且参数范围相同。根据引理C.1，进而将得到一个多项式时间算法，用于求解 SB-MSLR $-$ D，其尺度范围为 $\mathcal{C}_1 \cap \mathcal{C}_2$，该范围包含于定理 2.2 的尺度范围内，因此与定理 2.2相矛盾。 □

注释 C.4. 请注意，约束 $k \log p \log(1+\text{SNR})$ 中的下界 $\mathcal{C}_1$ 在 (7) 中使用，以证明定理 C.3 并不苛刻，因为这是在 SLR [中进行支撑集恢复的信息论最小样本复杂度，2019年，Gamarnik 和 Zadik，2022年，Wang 等人， 2010]。

从 SB-MSLR − D 到 SB-MSLR 的归约

我们利用 Gamarnik 和 Zadik [2017] 中的一个定理变体来构造归约，引理C.5。考虑以下关于 $\sigma > 0$ 的优化问题:

$$\begin{aligned} \psi := \min \quad & n^{-\frac{1}{2}} \|\sigma \boldsymbol{w} - \boldsymbol{X}\boldsymbol{\beta}_1 \odot \boldsymbol{z} - \boldsymbol{X}\boldsymbol{\beta}_2 \odot (1-\boldsymbol{z})\|_2 \\ \text{s.t.} \quad & \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \{-1,0,1\}^p, \ \boldsymbol{z} \in \{0,1\}^n \\ & \|\boldsymbol{\beta}_1\|_0 = \|\boldsymbol{\beta}_2\|_0 = k, \end{aligned} \tag{9}$$

其中 $\boldsymbol{X}$ 独立同分布$\sim \mathcal{N}(0,1)$，与 w_i 独立同分布$\sim \mathcal{N}(0,1)$ 相互独立。

引理 C.5。令 ψ 如 (9)中所定义。对于 $\delta > 0$ 我们有:

$$\mathbb{P}\left[\psi \geq e^{-(1+\delta)/2} \exp\left(-\frac{2k(\log p + 1)}{n}\right)\sqrt{k+\sigma^2}\right] \geq 1 - e^{-\frac{\delta}{2}n}.$$

证明与 [Gamarnik 和 Zadik 中定理 3.1 的证明几乎完全相同，2017] 因此省略。

引理 *C.1* 的证明。在整个证明过程中，为方便起见，我们省略参数 $(p_i, n_i, k_i, \sigma_i)$ 中的 i 下标。我们将 SB-MSLR 和 SB-MSLR − D 分别称为 MSLR$_{\xi,\ \tau}$和 MSLR $-$ D$_{\xi,\ \tau}$其中$\xi = 1$、$\tau = -1$。我们令 $\mathrm{P} := \mathbb{P}(\boldsymbol{X}, \boldsymbol{y})$ 表示 SLR − D 表述中的植入测度，令 $\mathrm{Q} := \mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})$ 表示零假设测度。我们强调，由于$\boldsymbol{\beta} \sim \mathcal{P}_{\|\beta\|_2}(\{-1,1\})$，$\|\boldsymbol{\beta}\|_2^2$和 k是可以互换的。按照引理陈述中的规定，假设对于任意 $\epsilon > 0$，$\mathbb{P}\left[\rho((\hat{}\boldsymbol{\beta}_1, \hat{}\boldsymbol{\beta}_2), (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)) < \epsilon\right] \underset{(i\to\infty)}{\to} 1$。

在植入假设 P 下，定义以下两个事件:

$$\tilde{\Omega}_1 := \left\{\{\hat{\boldsymbol{\beta}}_1 = \boldsymbol{\beta}_1, \hat{\boldsymbol{\beta}}_2 = \boldsymbol{\beta}_2\} \cup \{\hat{\boldsymbol{\beta}}_2 = \boldsymbol{\beta}_1, \hat{\boldsymbol{\beta}}_1 = \boldsymbol{\beta}_2\}\right\},$$

and

$$\tilde{\Omega}_2 := \{|w_q| < |\sigma^{-1}\langle \boldsymbol{X}_q, \boldsymbol{\beta}_2 - \boldsymbol{\beta}_1\rangle + w_q|, \ \forall q \in [n]\}.$$

˜注意，根据假设，Ω_1 在 P 下以概率 $1 - o(1)$ 发生。事实上，我们可以在给定算法 $\mathcal{A}$ 的定义中选择 ˆ ˆ$\epsilon < 1$，并且由于 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \{-1,0,1\}^p$，我们得到 $\mathbb{P}\left[\tilde{}\Omega_1\right] \to 1$。我们先考虑植入假设 P。令 ν_q独立同分布$\sim \mathcal{N}(0, (1-\xi\tau)\frac{k}{\sigma^2})$，且$g_q \sim \mathcal{N}(0,1)$，彼此独立，并且与 w_q 独立，其中 $q \in [n]$。在这种情况下，根据对称性，我们有

$$\begin{aligned} \mathrm{P}\left[\tilde{\Omega}_2^{\complement} \middle| \tilde{\Omega}_1\right] &= \mathrm{P}\left[\{|w_q| \geq |\sigma^{-1}\langle \boldsymbol{X}_q, \boldsymbol{\beta}_2 - \boldsymbol{\beta}_1\rangle + w_q|, \ \forall q \in [n]\}\right] \\ &= \int \mathrm{P}\left[\{|w_q| \geq |\nu_q + w_q|, \ \forall q \in [n]\} | w_q\right] \mathrm{P}[dw_q] \\ &= 2\int_0^\infty \mathrm{P}\left[\nu_q \in [-2w_q, 0], \ \forall q \in [n]\} | w_q\right] \mathrm{P}[dw_q] \\ &= 2\int_0^\infty \mathrm{P}\left[\{g_q \in [-2w_q/((1-\xi\tau)k/\sigma^2), 0], \ \forall q \in [n]\} \middle| w_q\right] \mathrm{P}[dw_q], \end{aligned}$$

其中我们有

$$\mathrm{P}\left[\{g_q \in [-2w_q/((1-\xi\tau)k/\sigma^2), 0], \ \forall q \in [n]\} \middle| w_q\right] \to 0 \quad \text{as } k/\sigma^2 \to \infty,$$

且 $\mathrm{P}\left[\{g \in [-2w_q/((1-\xi\tau)k/\sigma^2)\ 0],\ \forall q \in [n]\}\middle|w_q\right] \le 1$，因此我们可以应用控制收敛定理，得到

$$\mathrm{P}\left[\tilde{\Omega}_2^{\complement}\middle|\tilde{\Omega}_1\right] = 2\int_0^{\infty} \mathrm{P}\left[\{g_q \in [-2w_q/((1-\xi\tau)k/\sigma^2), 0],\ \forall q \in [n]\}\middle|w_q\right] \mathrm{P}[dw_q] \to 0 \text{ as } k/\sigma^2 \to \infty.$$

因此，我们有 $\mathrm{P}\left[\tilde{\Omega}_2\middle|\tilde{\Omega}_1\right] = 1 - o(1)$，从而 $\mathrm{P}\left[\left(\tilde{\Omega}_1,\ \tilde{\Omega}_2\right)\right] = 1 - o(1)$。

接下来，注意，在植入假设 P 下，且在联合事件 (Ω_1,Ω_2) 中，对于满足 $z_q = 1$ 的指标 q，我们有：

$$\begin{aligned}
|y_q - \frac{1}{\sigma}\langle \boldsymbol{X}_q, \hat{\boldsymbol{\beta}}_1\rangle| &= |y_q - \frac{1}{\sigma}\langle \boldsymbol{X}_q, \boldsymbol{\beta}_1\rangle| \\
&= |w_q| \\
&< |\frac{1}{\sigma}\langle \boldsymbol{X}_q, \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\rangle + w_q| \\
&= |y_q - \frac{1}{\sigma}\langle \boldsymbol{X}_q, \boldsymbol{\beta}_2\rangle| \\
&= |y_q - \frac{1}{\sigma}\langle \boldsymbol{X}_q, \hat{\boldsymbol{\beta}}_2\rangle|.
\end{aligned}$$

对于满足 $z_q = 0$ 的指标 q，交换 $\boldsymbol{\beta}_1$ 和 $\boldsymbol{\beta}_2$ 后也成立类似的结论。因此，我们有在 (Ω_1, Ω_2) 下，可以使用上述阈值化过程精确估计 $\boldsymbol{z}$，并将此精确估计记作 $\hat{\boldsymbol{z}}$。接着，我们在这种情况下定义检测算法：

$$\mathcal{A}'\left(\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{y}\end{bmatrix}\right) = \begin{cases}\mathtt{p}, & n^{-1/2}\|\boldsymbol{y} - \frac{1}{\sigma}\boldsymbol{X}\hat{\boldsymbol{\beta}}_1 \odot \hat{\boldsymbol{z}} - \frac{1}{\sigma}\boldsymbol{X}\hat{\boldsymbol{\beta}}_2 \odot (1-\hat{\boldsymbol{z}})\| \le \sqrt{5} \\ \mathtt{q}, & n^{-1/2}\|\boldsymbol{y} - \frac{1}{\sigma}\boldsymbol{X}\hat{\boldsymbol{\beta}}_1 \odot \hat{\boldsymbol{z}} - \frac{1}{\sigma}\boldsymbol{X}\hat{\boldsymbol{\beta}}_2 \odot (1-\hat{\boldsymbol{z}})\| > \sqrt{5}\end{cases}.$$

我们将证明 $\mathcal{A}'$ 的第二类错误趋于零。事实上，在 P 下，且在高概率事件 (Ω_1,Ω_2) 下，我们有：

$$\begin{aligned}
&\|\boldsymbol{y} - \frac{1}{\sigma}\boldsymbol{X}\hat{\boldsymbol{\beta}}_1 \odot \hat{\boldsymbol{z}} - \frac{1}{\sigma}\boldsymbol{X}\hat{\boldsymbol{\beta}}_2 \odot (1-\hat{\boldsymbol{z}})\|_2 \\
&= \|\frac{1}{\sigma}\boldsymbol{X}\boldsymbol{\beta}_1 \odot \boldsymbol{z} + \frac{1}{\sigma}\boldsymbol{X}\boldsymbol{\beta}_2 \odot (1-\boldsymbol{z}) + \boldsymbol{w} - \frac{1}{\sigma}\boldsymbol{X}\hat{\boldsymbol{\beta}}_1 \odot \hat{\boldsymbol{z}} - \frac{1}{\sigma}\boldsymbol{X}\hat{\boldsymbol{\beta}}_2 \odot (1-\hat{\boldsymbol{z}})\|_2 \\
&\le \|\boldsymbol{w}\|_2 + \|\frac{1}{\sigma}\boldsymbol{X}\boldsymbol{\beta}_1 \odot \boldsymbol{z} + \frac{1}{\sigma}\boldsymbol{X}\boldsymbol{\beta}_2 \odot (1-\boldsymbol{z}) - \frac{1}{\sigma}\boldsymbol{X}\hat{\boldsymbol{\beta}}_1 \odot \hat{\boldsymbol{z}} - \frac{1}{\sigma}\boldsymbol{X}\hat{\boldsymbol{\beta}}_2 \odot (1-\hat{\boldsymbol{z}})\|_2 \\
&= \|\boldsymbol{w}\|_2.
\end{aligned}$$

因此我们有

$$\begin{aligned}
&\mathrm{P}\left[\mathcal{A}'\left(\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{y}\end{bmatrix}\right) = \mathtt{q}\right] \\
&= \mathrm{P}\left[\mathcal{A}'\left(\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{y}\end{bmatrix}\right) = \mathtt{q}\middle|(\tilde{\Omega}_1,\tilde{\Omega}_2)^{\complement}\right]\cdot \mathrm{P}\left[(\tilde{\Omega}_1,\tilde{\Omega}_2)^{\complement}\right] + \mathrm{P}\left[\left\{\mathcal{A}'\left(\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{y}\end{bmatrix}\right) = \mathtt{q}\right\} \cap (\tilde{\Omega}_1,\tilde{\Omega}_2)\right] \\
&\le \mathrm{P}\left[\mathcal{A}'\left(\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{y}\end{bmatrix}\right) = \mathtt{q}\middle|(\tilde{\Omega}_1,\tilde{\Omega}_2)^{\complement}\right]\cdot \mathrm{P}\left[(\tilde{\Omega}_1,\tilde{\Omega}_2)^{\complement}\right] + \mathrm{P}\left[\left\{\|\boldsymbol{w}\|_2 > \sqrt{5n}\right\} \cap (\tilde{\Omega}_1,\tilde{\Omega}_2)\right] \\
&\le 1\cdot o(1) + \mathrm{P}\left[\left\{\|\boldsymbol{w}\|_2 > \sqrt{5n}\right\}\right] \\
&\le 1\cdot o(1) + e^{-n} = o(1),
\end{aligned}$$

其中，最后一个不等式是利用标准的卡方大偏差尾界得到的（Wainwright [2019]中的例 2.11）：

$$\mathrm{P}\left[\|\boldsymbol{w}\|_2 < \sqrt{n + 2\sqrt{nt} + 2t}\right] \geq 1 - e^{-t},$$

并取 $t = \sqrt{n}$。

现在我们转而证明 $\mathcal{A}'$ 的第一类错误趋于零。在零假设 Q 下，根据定义 ψ，由定义 9 可得

$$\|\boldsymbol{y} - \frac{1}{\sigma}\boldsymbol{X}\hat{\boldsymbol{\beta}}_1 \odot \hat{\boldsymbol{z}} - \frac{1}{\sigma}\boldsymbol{X}\hat{\boldsymbol{\beta}}_2 \odot (1 - \hat{\boldsymbol{z}})\|_2 \geq \psi,$$

这里回顾引理 C.5，可知

$$\mathrm{Q}\left[\psi \geq e^{-3/2} \exp\left(-\frac{2k(\log p + 1)}{n}\right)\sqrt{k + \sigma^2}\right] \geq 1 - e^{-n},$$

因此，只需证明

$$e^{-3/2} \exp\left(-\frac{2k(\log p + 1)}{n}\right)\sqrt{k + \sigma^2} > \sqrt{5}.$$

为此，取 $n^* = \frac{4k}{(\log p+1)\log(1+\frac{k}{\sigma^2}) - \log 5 - 3} = \Theta\left(\frac{4k}{\log p \log(1+\frac{k}{\sigma^2})}\right)$，并注意到，若该不等式对 n^*成立，则由于左边随 n递增，它对所有 $n \geq n^*$ 都成立。代入n^*，得到

$$e^{-3/2} \exp\left(-1/2 \log(1 + \frac{k}{\sigma^2}) + \log\sqrt{5} + 3/2\right)\sqrt{1 + 2\frac{k}{\sigma^2}} = \frac{\sqrt{1 + 2\frac{k}{\sigma^2}}}{\sqrt{1 + \frac{k}{\sigma^2}}}\sqrt{5} > \sqrt{5},$$

因此我们有

$$\begin{aligned}
&\mathrm{Q}\left[\mathcal{A}'\left(\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{y}\end{bmatrix}\right) = \mathtt{q}\right] \\
&= \mathrm{Q}\left[\|\boldsymbol{y} - \frac{1}{\sigma}\boldsymbol{X}\hat{\boldsymbol{\beta}}_1 \odot \hat{\boldsymbol{z}} - \frac{1}{\sigma}\boldsymbol{X}\hat{\boldsymbol{\beta}}_2 \odot (1 - \hat{\boldsymbol{z}})\|_2 > \sqrt{5}\right] \\
&\geq \mathrm{Q}\left[\psi > \sqrt{5}\right] \\
&\geq \mathrm{Q}\left[\psi \geq e^{-3/2} \exp\left(-\frac{2k(\log p + 1)}{n}\right)\sqrt{k + \sigma^2}\right] \\
&\geq 1 - e^{-n}.
\end{aligned}$$

需要特别注意，n 满足 $\mathcal{C}_2$的约束。

□

C.2 从 SB-MSLR 到 PSB-MSLR的归约

引理的证明 *C.2*。 首先，回顾一下 PSB-MSLR制度意味着：

$$\phi = 1/2 \text{ and } \boldsymbol{\beta}_{1,j} = -\boldsymbol{\beta}_{2,j} \text{ for } j \in J \subseteq \mathrm{supp}(\boldsymbol{\beta}_1) \cap \mathrm{supp}(\boldsymbol{\beta}_2) \text{ with } C_1 k \leq |J| \leq C_2 k,$$

对于某些常数 $1 \geq C_1$、$C_2 > 0$。不失一般性，我们可以取 $|J| = Ck$，其中 $0 < C \leq 1$ 为某个常数，后续所有常数均可相应地取下界或上界。据此，PSB-MSLR 对应于 MSLR$_{\xi,\ \tau}$，其中某些 $\tau \in (-1,$ 1) 和某些 $\xi > 0$ 满足这一关系。回顾一下，ξ、τ 为常数量级，即不随 n、p、k 缩放。

为简洁起见，我们将 $\mathfrak{P}' :=$ 记为 PSB-MSLR，将 $\mathfrak{P} :=$ 记为 SB-MSLR。令 $c = 1 - \frac{\tau \cdot \xi + 1}{2} \in$ (0, 1) 表示 $\boldsymbol{\beta}_1$ 与 $\boldsymbol{\beta}_2$ 之间符号相反的匹配非零项所占的比例（直观地说，这对应于信号中“困难”的部分），并注意该比例是固定的。注意，这是因为 $\tau \cdot \xi = \frac{\langle \boldsymbol{\beta}_1,\ \boldsymbol{\beta}_2 \rangle}{k}$ 对于 $\xi > 0$。

给定参数序列 $\{p_i', n_i', k_i', \sigma_i'\}_{i=1}^{\infty} \subseteq \mathcal{C}_1$（其中 $\mathfrak{P}'$），考虑参数序列 $\{(p_i, n_i, k_i, \sigma_i)\}_{i=1}^{\infty} = \{(cp_i', n_i', ck_i',$ $\sigma_i')\}_{i=1}^{\infty}$（其中 $\mathfrak{P}$）。注意，$\{(p_i, n_i, k_i, \sigma_i)\}_{i=1}^{\infty} \subseteq \mathcal{C}_1$，因为（为方便起见，省略下标 i 的记号）：

- $k = ck' = o(cC\sqrt{p'}) = o(c^{3/2}C\sqrt{p}) = o(\sqrt{p})$
- $n = n' = o\left((k' + (\sigma')^2)^2 \cdot \frac{1}{\log p'}\right) = o\left((k + \sigma^2)^2 \cdot \frac{1}{\log p}\right)$,
- $n = n' = \omega(\max\{k', \log p'\}) = \omega(\max\{k, \log p\})$,

因此，$\mathfrak{P}$ 的参数范围也包含于 $\mathcal{C}_1$中。对于 $i \in \mathbb{N}$，令 $J = (\boldsymbol{X}, \boldsymbol{y})$表示 $\mathfrak{P}$ 的一个实例，其参数为（p_i, n_i, k_i, σ_i），其中我们回顾如下：

$$\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X} \\ \frac{1}{\sigma}\boldsymbol{X}\boldsymbol{\beta}_1 \odot \boldsymbol{z} + \frac{1}{\sigma}\boldsymbol{X}\boldsymbol{\beta}_2 \odot (1 - \boldsymbol{z}) + \boldsymbol{w} \end{bmatrix}.$$

现在我们希望证明：给定 $\mathfrak{P}'$ 的一列参数$\{p_i', n_i', k_i'\}_{i=1}^{\infty} \subseteq \mathcal{C}_1$ 以及一个能够求解该问题的随机多项式时间算法 $\mathcal{A}'$，我们可以构造一个沿用上述参数序列 $\{p_i, n_i, k_i\}_{i=1}^{\infty} \subseteq \mathcal{C}_1$、求解 $\mathfrak{P}$ 的随机多项式时间算法 $\mathcal{A}$。我们将通过把 $\mathcal{A}'$ 与一个前处理和后处理步骤组合起来，构造所需的算法 $\mathcal{A}$。具体而言，令 $\mathcal{A} = \mathcal{B} \circ \mathcal{A}' \circ \mathcal{D}$，其中 $\mathcal{B}$ 和 $\mathcal{D}$ 将在下文定义。

首先，令RS 表示这样一个随机变量：根据指标集 p 的均匀随机置换，对给定大小的 p列矩阵的向量或列进行重排 $[p]$。令IRS表示将该重排过程逆转的随机变量，使得IRS $\circ$RS 为恒等操作。令 $\mathbb{I}$ 表示任意大小的全 1 向量（其大小由上下文推断）。令 $\mathbb{I}_1$ 和 $\mathbb{I}_2$ 表示一个 $(1-c)\ k'$ 稀疏向量的两个独立副本，其中 $\{0, 1, -1\}^{(1-c)p'}$。现在我们定义前处理和后处理过程算法1、2，可以看出，它们相对于 p在随机多项式时间内运行。

算法 1：$\mathcal{D}(\boldsymbol{X}, \boldsymbol{y})$，其中 $(\boldsymbol{X}, \boldsymbol{y})$ 是 $\mathfrak{P}(p, n, k, \sigma)$ 的一个实例
数据：$\boldsymbol{V} \in \mathbb{R}^{n \times (1-c)p}$ 具有列 $\overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \boldsymbol{I}_n)\tilde{\boldsymbol{X}} \leftarrow \text{RS}[\boldsymbol{V} \quad \boldsymbol{X}]$返回 $(\boldsymbol{X}, \tilde{\boldsymbol{y}})$

命题 C.6. 对于 $i \in \mathbb{N}$，$(\boldsymbol{X}, \boldsymbol{y})$ 是 $\mathfrak{P}(p_i, n_i, k_i, \sigma_i)$ 的一个实例，而 $(\boldsymbol{X}', \boldsymbol{y}')$ 是$\mathfrak{P}'(p_i', n_i', k_i')$ 的一个实例，我们有 $\mathcal{D}(\boldsymbol{X}, \boldsymbol{y}) \overset{d}{=} (\boldsymbol{X}', \boldsymbol{y}')$。

证明。令 $(\tilde{\boldsymbol{X}}\ \tilde{\boldsymbol{y}} = \mathcal{D}\ \boldsymbol{X}\ \boldsymbol{y}$ $\tilde{\boldsymbol{X}} \overset{d}{=} \boldsymbol{X}'$ $[\boldsymbol{V} \quad \boldsymbol{X}]$ $n \times\ p +,) : (,)$。
$\frac{1-c}{c}p\ = n' \times\ cp' +\ 1 - c\ p'\ = n' \times p'$ $\tilde{\boldsymbol{y}}$
首先注意，由于具有维度 ( ) (( ))。接下来，注意我们可以将其分解如下：

$$\tilde{\boldsymbol{y}} = \boldsymbol{y} + \frac{1}{\sigma}\boldsymbol{V}\ \mathbb{I}$$

算法 2： $\mathcal{B}(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_2)$，其中 $(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_2)$ 均属于 $\{0, 1, -1\}^{p'}$

$\begin{bmatrix} \bar{\mathbb{I}}^1 \\ \boldsymbol{\beta}_1 \end{bmatrix} \leftarrow \mathtt{IRS}(\tilde{\boldsymbol{\beta}}_1)$

$\begin{bmatrix} \bar{\mathbb{I}}^2 \\ \boldsymbol{\beta}_2 \end{bmatrix} \leftarrow \mathtt{IRS}(\tilde{\boldsymbol{\beta}}_2)$

return $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$

$$
\begin{aligned}
&= \frac{1}{\sigma}\boldsymbol{X}\boldsymbol{\beta}_1 \odot \boldsymbol{z} + \frac{1}{\sigma}\boldsymbol{X}\boldsymbol{\beta}_2 \odot (1-\boldsymbol{z}) + \frac{1}{\sigma}\boldsymbol{V}_{\bar{\mathbb{I}}} + w \\
&= \frac{1}{\sigma}\boldsymbol{X}\boldsymbol{\beta}_1 \odot \boldsymbol{z} + \frac{1}{\sigma}\boldsymbol{X}\boldsymbol{\beta}_2 \odot (1-\boldsymbol{z}) + \frac{1}{\sigma}\boldsymbol{V}_{\bar{\mathbb{I}}} \odot \boldsymbol{z} + \frac{1}{\sigma}\boldsymbol{V}_{\bar{\mathbb{I}}} \odot (1-\boldsymbol{z}) + w \\
&= \frac{1}{\sigma}(\mathtt{RS}[\boldsymbol{V} \quad \boldsymbol{X}]) \left(\mathtt{RS}\begin{bmatrix} \bar{\mathbb{I}}^1 \\ \boldsymbol{\beta}_1 \end{bmatrix}\right) \odot \boldsymbol{z} + \frac{1}{\sigma}(\mathtt{RS}[\boldsymbol{V} \quad \boldsymbol{X}]) \left(\mathtt{RS}\begin{bmatrix} \bar{\mathbb{I}}^2 \\ \boldsymbol{\beta}_2 \end{bmatrix}\right) \odot (1-\boldsymbol{z}) + w
\end{aligned}
$$

因此， $\tilde{\boldsymbol{y}}$ 是一个大小为 $cp' + (1-c)\, p' = p'$、具有 $\left(\mathtt{RS}\begin{bmatrix} \bar{\mathbb{I}}^1 \\ \boldsymbol{\beta}_1 \end{bmatrix}\right.$ 个信号的 MSLR 模型的输出；$\left.\mathtt{RS}\begin{bmatrix} \bar{\mathbb{I}}^2 \\ \boldsymbol{\beta}_2 \end{bmatrix}\right)$ 包含来自（$\boldsymbol{\beta}_1$， $\boldsymbol{\beta}_2$）和（$1-c$）的 ck' 个符号相反的非零条目，以及通过追加 $\bar{\mathbb{I}}$ 得到的 k' 个其余非零条目，总支撑大小为 k'。此外，这些信号经过一个独立同分布高斯设计矩阵的线性变换，并且如上所述满足 $n' \times p'$。第一点和第二点合起来意味着， $\tilde{\boldsymbol{y}}$ 是一个 MSLR$_{\xi,\, \tau}$ 模型的输出，并且 $\mathcal{D}(\boldsymbol{X},\ \boldsymbol{y}) = (\tilde{\boldsymbol{X}},\ \tilde{\boldsymbol{y}}) \overset{d}{=} (\boldsymbol{X}',\ \boldsymbol{y}')$。 □

根据命题 C.6，剩下要证明的只有 $\mathbb{P}[\|\mathcal{B} \circ \mathcal{A}' \circ \mathcal{D}(\boldsymbol{X},\ \boldsymbol{y}) - (\boldsymbol{\beta}_1,\ \boldsymbol{\beta}_2)\|_\infty > 0] \to 0$。事实上，以下步骤由命题 C.6 以及 $\mathcal{B}$ 的定义可得：

$$
\begin{aligned}
&\|\mathcal{A}'(\boldsymbol{X}', \boldsymbol{y}') - (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')\|_\infty \\
\overset{d}{=} &\left\|\mathcal{A}'(\mathtt{RS}([\boldsymbol{V} \quad \boldsymbol{X}]), \boldsymbol{y} + \frac{1}{\sigma}\boldsymbol{V}_{\bar{\mathbb{I}}}) - \left(\mathtt{RS}\begin{bmatrix} \bar{\mathbb{I}}^1 \\ \boldsymbol{\beta}_1 \end{bmatrix}, \mathtt{RS}\begin{bmatrix} \bar{\mathbb{I}}^2 \\ \boldsymbol{\beta}_2 \end{bmatrix}\right)\right\|_\infty \\
= &\left\|\mathcal{A}' \circ \mathcal{D}(\boldsymbol{X}, \boldsymbol{y}) - \left(\mathtt{RS}\begin{bmatrix} \bar{\mathbb{I}}^1 \\ \boldsymbol{\beta}_1 \end{bmatrix}, \mathtt{RS}\begin{bmatrix} \bar{\mathbb{I}}^2 \\ \boldsymbol{\beta}_2 \end{bmatrix}\right)\right\|_\infty \\
\geq &\left\|\mathcal{B} \circ \mathcal{A}' \circ \mathcal{D}(\boldsymbol{X}, \boldsymbol{y}) - \mathcal{B}\left(\mathtt{RS}\begin{bmatrix} \bar{\mathbb{I}}^1 \\ \boldsymbol{\beta}_1 \end{bmatrix}, \mathtt{RS}\begin{bmatrix} \bar{\mathbb{I}}^2 \\ \boldsymbol{\beta}_2 \end{bmatrix}\right)\right\|_\infty \\
= &\|\mathcal{A}(\boldsymbol{X}, \boldsymbol{y}) - (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)\|_\infty,
\end{aligned}
$$

因此：

$$
\mathbb{P}\left[\|\mathcal{A}'(\boldsymbol{X}', \boldsymbol{y}') - (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')\|_\infty > 0\right] \to 0 \implies \mathbb{P}\left[\|\mathcal{A}(\boldsymbol{X}, \boldsymbol{y}) - (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)\|_\infty > 0\right] \to 0,
$$

其中， $\mathcal{A}$ 的运行时间相对于输入规模 p 为随机多项式时间，因为它是若干相对于 p为随机多项式时间的过程的组合。 □

C.3 从 $\mathtt{SB\text{-}MSLR} - \mathtt{D}$ 归约到 $\mathtt{SPR}$

在本节中，我们给出一种多项式时间归约：将无噪声 SB-MSLR – D 中的强检测归约为 SPR 中的精确支撑恢复，其中信号的非零条目属于 $\{-1, 0, 1\}^p$。更具体地说，我们考虑以下对称 SLR 问题。

Defi义 C.7（S – SLR）。对于 $\boldsymbol{X} \in \mathbb{R}^{n \times p}$、 $\boldsymbol{w} \in \mathbb{R}^n$ 和 $\boldsymbol{z} \in \mathbb{R}^n$，考虑该模式 l:

$$
\boldsymbol{y} = g(\boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{w},
$$

其中，$\odot$表示向量之间的逐元素乘积，$X_{i,\ j\text{独立同分布}\sim}\ \mathcal{N}(0,1)$，$w_{i\text{独立同分布}\sim}\ \mathcal{N}(0,\sigma^2)$，$\boldsymbol{\beta}\in\mathbb{R}^p$各自为 k-稀疏，且 $g:\ \mathbb{R}^n\to\mathbb{R}^n$ 一个可分离逐项偶函数（对于 $i\in[n]$和 $x\in\mathbb{R}$，$g_i(x)=g_i(-x)$）。给定 $(\boldsymbol{X},\boldsymbol{y})$，目标是估计 $\boldsymbol{\beta}$。

注意，令 $g_i(x)=|x|$ 和 $g_i(x)=x^2$，即可得到稀疏信号相位恢复问题（SPR）的两种标准表述 [Cand`es 等人，2015年，Liu 等人，2021]。我们希望证明，在定理 2.2 的参数尺度范围内，无噪声情形（$\frac{\mathrm{SNR}+1}{\mathrm{SNR}}=1$）下该问题是困难的，因此我们在约束集 $\mathcal{C}_3$内证明一个归约：

$$\mathcal{C}_3=\Big\{(p_i,n_i,k_i,\sigma_i)_{i=1}^{\infty}\subset\mathbb{N}^4:\exists C\in\mathbb{R}_+\ \text{ s.t. }\ p_i=\omega_i(1),k_i=o(\sqrt{p_i}),,\\ n_i=\omega(\max\{k_i,\log p_i\})\,n_i=o\left(k_i^2\cdot\frac{1}{\log p_i}\right)\Big\}. \tag{10}$$

注意，$\mathcal{C}_3$只是$\mathcal{C}_1$式（7）中的情形，但约束更严格，即 $n_i=o\left(k_i^2\cdot\ {}^1\log p_i\right)$。我们先给出一个引理，并用它来启动定理 C.9中的归约。

引理 C.8. 固定信号先验为 $\mathcal{P}_{\|\beta\|_2}(\{-1,1\})$。对于 S $-$ SLR 的任意参数序列 $\{(p_i',n_i',k_i',\sigma_i')\}_{i=1}^{\infty}\in\mathcal{C}_3$，其解为 $\boldsymbol{\beta}'$，问题实例为 $(\boldsymbol{X}',\boldsymbol{y}')$，存在 SB-MSLR（无噪声）的一列参数 $\{(p_i,n_i,k_i,\sigma_i=0)\}_{i=1}^{\infty}\in\mathcal{C}_3$，其解为 $\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2$，问题实例为 $(\boldsymbol{X},\boldsymbol{y})$；因此，对于任意用于 S $-$SLR、并产生 $\hat{}\,\boldsymbol{\beta}'$ 的随机多项式时间算法 $\mathcal{A}'$，满足

$$\mathbb{P}\left[\|\hat{\boldsymbol{\beta}}'-\boldsymbol{\beta}\|_\infty>0\right]\to 0,$$

$\hat{}\,\hat{}$ 我们可以构造第二个用于 SB-MSLR 的随机多项式时间算法 $\mathcal{A}$，输出 $(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)$，使得

$$\mathbb{P}\left[\|(\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2)-(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)\|_\infty>0\right]\to 0.$$

证明。为方便起见，我们省略下标 i。简洁起见，令 $\mathfrak{P}'$:= S $-$ SLR，且 $\mathfrak{P}$:= SB-MSLR。给定 $\{p_i',n_i',k_i',\sigma_i'\}_{i=1}^{\infty}$ 中用于 $\mathfrak{P}'$ 的一列参数 $\mathcal{C}_3$，考虑用于 $\mathfrak{P}$ 的参数序列 $\{(p_i,n_i,k_i,\sigma_i)\}_{i=1}^{\infty}=\{(p_i',n_i',k_i',0)\}_{i=1}^{\infty}$。注意，$\{(p_i,n_i,k_i,\sigma_i)\}_{i=1}^{\infty}$ 属于 $\mathcal{C}_3$。令 $(\boldsymbol{X},\boldsymbol{y})$ 为一个参数为 $\{(p_i,n_i,k_i,\sigma_i)\}_{i=1}^{\infty}$ 的 $\mathfrak{P}$ 问题实例，并回忆一下，对于无噪声 SB-MSLR，观测的形式为 $(\boldsymbol{X},\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}\odot(2\boldsymbol{z}-1))$。我们执行一个预处理步骤，构造 $\tilde{\boldsymbol{y}}:=g(\boldsymbol{y})+\boldsymbol{w}\in\mathbb{R}^n$，其中 $\boldsymbol{w}$ 独立同分布于 $\sim\ \mathcal{N}(0,1)$；该步骤可在随机多项式时间内完成。注意，由于 g 关于符号翻转具有对称性，$(\boldsymbol{X},\tilde{\boldsymbol{y}})$ 现在是 $\mathfrak{P}'$ 的一个实例。随后，我们在 $(\boldsymbol{X},\tilde{\boldsymbol{y}})$ 上运行算法 $\mathcal{A}'$，得到 $\hat{}\,\boldsymbol{\beta}'$，其中 $\mathbb{P}\left[\|\hat{}\,\boldsymbol{\beta}'-\boldsymbol{\beta}\|_\infty>0\right]\to 0$ 如问题陈述所述。不失一般性，令 $(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)=(\boldsymbol{\beta},-\boldsymbol{\beta})$，我们得到一个算法 $\mathcal{A}$，它产生 $\mathbb{P}\left[\|(\hat{}\,\boldsymbol{\beta}_1,\hat{}\,\boldsymbol{\beta}_2)-(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)\|_\infty>0\right]\to 0$。 □

定理 C.9 （从 SB-MSLR $-$ D 到 SPR 中的精确恢复的归约）。考虑具有联合信号先验 SPR 的设定 $\mathcal{P}_{\|\beta\|_2}$（$\{-1,1\}$）。任何在参数范围 $\mathcal{A}$中求解SPR 且满足 $\mathcal{C}_2\cap\mathcal{C}_3$ 和 $\mathit{SNR}=\omega$（1）的随机多项式时间算法都会与定理 2.2矛盾。

证明。我们首先使用引理 C.1，将 SB-MSLR $-$ D 归约为约束集 $\mathcal{C}_2$ (8) 内的 SB-MSLR。接着，根据 SPR 的具体定义，选择 $g(x)=|x|$ 或 $g(x)=x^2$，并使用引理 C.8 在约束集 $\mathcal{C}_3$ (10) 内将无噪声 SB-MSLR 归约为 SPR。整个过程中，我们令 SNR $=\omega(1)$，以满足引理 C.1。假设存在一个随机多项式时间算法 $\mathcal{A}$，能够在 SPR 中对非零条目属于 $\{-1,0,1\}^p$ 的信号进行精确恢复，即 $\mathbb{P}\left[\|\hat{}\,\boldsymbol{\beta}'-\boldsymbol{\beta}\|_\infty>0\right]\to 0$。那么，根据上述归约链，我们就能得到

一个随机多项式时间算法 $\mathcal{A}'$ 能够在以下问题中进行强检测：SB-MSLR $-$ D 非零条目属于 $\{-1,0,1\}^p$ 的信号，其尺度范围为 $\mathcal{C}_2 \cap \mathcal{C}_3$，该范围包含于定理 2.2 所述的参数范围内，因此将与定理 2.2矛盾。

□

为完整起见，我们还给出了一个从 SB-MSLR $-$ D 到 SPR的一种检测变体的归约，见定理C.11。

定义 C.10（检测变体 SPR $-$ D）。对于 $\boldsymbol{X} \in \mathbb{R}^{n\times p}$、$\sigma > 0$ 和 $\boldsymbol{w}^{(1)}$、$\boldsymbol{w}^{(2)} \in \mathbb{R}^n$，考虑以下假设检验问题：

$$\mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y}) : \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X} \\ \sqrt{\frac{\|\boldsymbol{\beta}\|_2^2}{\sigma^2}} |\boldsymbol{w}_1| + \boldsymbol{w}_2 \end{bmatrix}$$
$$\mathbb{P}(\boldsymbol{X}, \boldsymbol{y}) : \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X} \\ \frac{1}{\sigma} |\boldsymbol{X}\boldsymbol{\beta}| + \boldsymbol{w} \end{bmatrix}$$

其中 $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \sim \mathcal{P}_{\|\boldsymbol{\beta}\|_2}(\mathcal{D})$，且 $X_{i,\ j}$ 独立同分布$\sim \mathcal{N}(0,1)$，w_i 独立同分布$\sim \mathcal{N}(0,1)$，z_i 独立同分布$\sim$ 伯努利分布(ϕ)。任务是构造一个函数 f，使其能够强区分 $\mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})$ 与 $\mathbb{P}(\boldsymbol{X}, \boldsymbol{y})$。

定理 C.11。固定信号先验为 $\mathcal{P}_{\|\beta\|_2}(\{-1,1\})$。对于 SPR $-$ D 的任意参数序列 $\{(p_i', n_i', k_i', \sigma_i')\}_{i=1}^{\infty}$，其中信号为 $\boldsymbol{\beta}'$，问题实例为 $(\boldsymbol{X}', \boldsymbol{y}')$，存在一个 SB-MSLR $-$ D（无噪声）的参数序列$\{(p_i, n_i, k_i, \sigma_i = 0)\}_{i=1}^{\infty} \subseteq \mathcal{C}_3$，其信号为 $\boldsymbol{\beta}_1$、$\boldsymbol{\beta}_2$，问题实例为 $(\boldsymbol{X}, \boldsymbol{y})$；因此，对于任意一个能在 $\mathfrak{P}'$ 中求解强检测的随机多项式时间算法 $\mathcal{A}'$，我们都可以构造另一个随机多项式时间算法 $\mathcal{A}$，用于在$\mathfrak{P}$中求解强检测。

证明。为方便起见，我们省略下标 i 记号。简言之，令 $\mathfrak{P}' := \text{SPR} - \text{D}$，并令$\mathfrak{P} := \text{SB-MSLR} - \text{D}$。给定 $\{p_i', n_i', k_i', \sigma_i'\}_{i=1}^{\infty}$ 中针对 $\mathfrak{P}'$的参数序列 $\mathcal{C}_3$，考虑针对 $\mathfrak{P}$的参数序列 $\{(p_i, n_i, k_i, \sigma_i)\}_{i=1}^{\infty} = \{(p_i', n_i', k_i', 0)\}_{i=1}^{\infty}$。注意，$\{(p_i, n_i, k_i, \sigma_i)\}_{i=1}^{\infty}$ 属于 $\mathcal{C}_3$。设 $(\boldsymbol{X}, \boldsymbol{y})$ 是一个 $\mathfrak{P}$ 问题实例，其参数为 $\{(p_i, n_i, k_i, \sigma_i)\}_{i=1}^{\infty}$，并回顾在无噪声 SB-MSLR $-$ D 中，备择假设下的观测具有如下形式 $(\boldsymbol{X}, \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} \odot (2\boldsymbol{z} - 1))$。我们应用预处理步骤来构造 $\tilde{\boldsymbol{y}} := |\boldsymbol{y}| + \boldsymbol{w} \in \mathbb{R}^n$，其中 $\boldsymbol{w}$ 独立同分布$\sim \mathcal{N}(0,1)$，该步骤可在随机多项式时间内完成。注意，在两个假设下，$(\boldsymbol{X}, \tilde{\boldsymbol{y}})$ 都是 $\mathfrak{P}'$的实例，这是因为 g 关于符号翻转是对称的（偶函数）。随后，我们可以在 $(\boldsymbol{X}, \tilde{\boldsymbol{y}})$ 上运行算法 $\mathcal{A}'$，以解决关于 $(\boldsymbol{X}, \boldsymbol{y})$ 的 $\mathfrak{P}$ 假设检验问题。

□

D 高效算法的证明

D.1 CORR 的支撑集恢复 MSLR

定理 D.1（CORR 在 MSLR 中实现联合支撑恢复）。考虑 MSLR 的一般设定，$(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \sim \mathcal{P}_{\|\boldsymbol{\beta}\|_2}(\mathcal{D})$。令 $\epsilon \in (0,1)$ 为 CORR 中使用的参数。则在满足以下条件时

$$n \geq \frac{32(1+\epsilon)}{\min\{\phi^2 \boldsymbol{\beta}_{min}^2, (1-\phi)^2 \boldsymbol{\beta}_{min}^2, \langle \boldsymbol{\beta} \rangle_{min}^2\}} \frac{k(\boldsymbol{SNR}+1)}{\boldsymbol{SNR}} \log 2p,$$

我们有，CORR 以至少$1 - c_1(\frac{k}{p} + ke^{-c_2 n} + \frac{k}{n} + \frac{1}{p^{c_2}})$的概率输出信号 $\boldsymbol{\beta}_1$ 和 $\boldsymbol{\beta}_2$ 的精确联合支撑，其中常数为 c_1、$c_2 > 0$。

证明。令 $\delta > 0$稍后确定，对于两个集合 A 和 B，记对称差$A\Delta B := (A \setminus B) \cup (B \setminus A)$。令 $\tau := \sqrt{2(1+\epsilon/2)\log 2p}$，并定义误差事件

$$\mathcal{E} := \cup_{j\in\mathcal{S}_1\cup\mathcal{S}_2}\left\{\left|\frac{\langle \boldsymbol{X}_j, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| < \tau\right\} \cup \left\{\max_{q\in(\mathcal{S}_1\cup\mathcal{S}_2)^{\complement}}\left|\frac{\langle \boldsymbol{X}_q, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \geq \tau\right\}.$$

从这里开始，我们将索引集合 $j \in \mathcal{S}_1 \cup \mathcal{S}_2$ 划分为两个集合，

$$J_a := \{j \in [p] : j \in \mathcal{S}_1 \cap \mathcal{S}_2\}, \qquad J_b := \{j \in [p] : j \in \mathcal{S}_1 \Delta \mathcal{S}_2\},$$

并据此应用并集界：

$$\begin{aligned}
\mathbb{P}[\mathcal{E}] &\leq \mathbb{P}\left[\cup_{j\in\mathcal{S}_1\cup\mathcal{S}_2}\left\{\left|\frac{\langle \boldsymbol{X}_j, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| < \tau\right\}\right] + \mathbb{P}\left[\max_{q\in(\mathcal{S}_1\cup\mathcal{S}_2)^{\complement}}\left|\frac{\langle \boldsymbol{X}_q, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \geq \tau\right] \\
&\leq \mathbb{P}\left[\cup_{j_a\in J_a}\left\{\left|\frac{\langle \boldsymbol{X}_{j_a}, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| < \tau\right\}\right] + \mathbb{P}\left[\cup_{j_b\in J_b}\left\{\left|\frac{\langle \boldsymbol{X}_{j_b}, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| < \tau\right\}\right] + \mathbb{P}\left[\max_{q\in(\mathcal{S}_1\cup\mathcal{S}_2)^{\complement}}\left|\frac{\langle \boldsymbol{X}_q, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \geq \tau\right] \\
&\leq 2k\mathbb{P}\left[\left|\frac{\langle \boldsymbol{X}_{j_a}, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| < \tau\right] + 2k\mathbb{P}\left[\left\{\left|\frac{\langle \boldsymbol{X}_{j_b}, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| < \tau\right\}\right] + \mathbb{P}\left[\max_{q\in(\mathcal{S}_1\cup\mathcal{S}_2)^{\complement}}\left|\frac{\langle \boldsymbol{X}_q, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \geq \tau\right], \qquad (11)
\end{aligned}$$

其中 $j_a \in J_a$、$j_b \in J_b$ 和 $q \in (\mathcal{S}_1 \cup \mathcal{S}_2)^{\complement}$。分析最后一项时，我们注意到，对于 $q \notin \mathcal{S}_1 \cap \mathcal{S}_2$，有 $\frac{\langle \boldsymbol{X}_q, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2} \overset{\boldsymbol{y}\text{ 独立同分布}}{\sim} \mathcal{N}(0,1)$。因此，我们有，

$$\begin{aligned}
&\mathbb{P}\left[\max_{q\in(\mathcal{S}_1\cup\mathcal{S}_2)^{\complement}}\left|\frac{\langle \boldsymbol{X}_q, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \geq \sqrt{2(1+\epsilon/2)\log 2p}\right] \\
&\overset{i)}{\leq} \mathbb{P}\left[\max_{q\in(\mathcal{S}_1\cup\mathcal{S}_2)^{\complement}}\left|\frac{\langle \boldsymbol{X}_q, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \geq \sqrt{2\log 2p} + \frac{\epsilon}{2\sqrt{8}}\sqrt{\log 2p}\right] \overset{ii)}{\leq} (2p)^{-\frac{1}{16}\left(\frac{\epsilon}{2}\right)^2},
\end{aligned}$$

其中 $i)$ 可由 $\sqrt{2(1+\epsilon/2)}$ 得出 $\geq \sqrt{2} + \frac{\epsilon}{2\sqrt{8}}$，而 $ii)$ 则来自标准高斯变量最大值的尾界（参见 $\mathbb{P}[\Omega_2^{\complement}(t)]$，引理 D.9）。分别应用引理 D.3和 D.2 （并选择足够小的 $\delta > 0$ 以满足这些条件）到式（11）中的前两项，可得存在常数 c_1，$c_2 > 0$，

$$\mathbb{P}[\mathcal{E}] \leq c_1\left(\frac{k}{p} + ke^{-c_2 n} + \frac{k}{n} + \frac{1}{p^{c_2}}\right).$$

□

主要集中引理。

引理 D.2（$j \in \mathcal{S}_1\Delta\mathcal{S}_2$ 的一般界）。考虑 MSLR 的设置。设 $j^* \in \mathcal{S}_1\Delta\mathcal{S}_2$。那么，如果 $n \geq {}^{32}\,\text{(1+ϵ) 最小值}\{\phi^2, (1-\phi)^2\}\beta_{min}^2(\|\boldsymbol{\beta}\|_2^2 + \sigma^2)\log 2p$，对于任意 $\epsilon \in (0,1)$，我们可得

$$\mathbb{P}\left[\left|\frac{\langle \boldsymbol{X}_{j^*}, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \leq \sqrt{2(1+\epsilon/2)\log 2p}\right] \leq \frac{1}{p} + 4e^{-\frac{\delta^2 n}{8}} + \frac{39}{\delta^2 n}.$$

证明。令 $g_i \overset{\text{独立同分布}}{\sim} \mathcal{N}(0,1)$（与 $\boldsymbol{y}$、$\boldsymbol{z}$、$\boldsymbol{\beta}_1$、$\boldsymbol{\beta}_2$ 独立），其中 $i \in [n]$ 和 $\delta > 0$ 将在后文选定。不失一般性，我们可以假设 $j^* \in \mathcal{S}_1 \setminus \mathcal{S}_2$。我们首先考虑固定的 $\boldsymbol{\beta}_1$、$\boldsymbol{\beta}_2$、$\boldsymbol{y}$ 和 $\boldsymbol{z}$ 向量，因此我们关注的初始随机性在设计矩阵 $\boldsymbol{X}$ 中。定义

- $\sigma_i^{(1)} := \frac{y_i}{\|\boldsymbol{y}\|_2}\sqrt{1-\frac{\boldsymbol{\beta}_{1,j^*}^2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}}$,
- $\sigma_i^{(2)} := \frac{y_i}{\|\boldsymbol{y}\|_2}$,
- $\mu_i^{(1)} := \frac{y_i^2}{\|\boldsymbol{y}\|_2}\frac{\boldsymbol{\beta}_{1,j^*}}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}$,
- $\mu_i^{(2)} := 0$,

并注意，对于 $\tau \in \mathbb{R}$,

$$\mathbb{P}\left[X_{j^*,i} \le \tau | \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] = \mathbb{P}\left[z_i\left(\frac{y_i \cdot \boldsymbol{\beta}_{1,j^*}}{\|\boldsymbol{\beta}\|_2^2+\sigma^2} + \sqrt{1-\frac{\boldsymbol{\beta}_{1,j^*}^2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}}g_i\right) + (1-z_i)g_i \le \tau \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, y_i, z_i\right],$$

这可由引理D.4推出。因此我们有

$$\mathbb{P}\left[X_{j^*,i}\frac{y_i}{\|\boldsymbol{y}\|_2} \le \tau \middle| \boldsymbol{y}, \boldsymbol{z}\right] = \mathbb{P}\left[z_i(\mu_i^{(1)}+\sigma_i^{(1)}g_i) + (1-z_i)(\mu_i^{(2)}+\sigma_i^{(2)}g_i) \le \tau \middle| \boldsymbol{y}, \boldsymbol{z}\right],$$

由此可得，对于 $\tau > 0$,

$$\begin{aligned}
&\mathbb{P}\left[\left|\frac{\langle \boldsymbol{X}_{j^*}, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \le \tau \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \\
&= \mathbb{P}\left[\left|\sum_{i=1}^n z_i(\mu_i^{(1)}+\sigma_i^{(1)}g_i) + (1-z_i)(\mu_i^{(2)}+\sigma_i^{(2)}g_i)\right| \le \tau \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \\
&= \mathbb{P}\left[\left|\sum_{i=1}^n (z_i\mu^{(1)} + (1-z_i)\mu^{(2)}) + (z_i\sigma^{(1)} + (1-z_i)\sigma^{(2)})g_i\right| \le \tau \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \\
&\overset{i)}{=} \mathbb{P}\left[\left|\sum_{i=1}^n (z_i\mu^{(1)} + (1-z_i)\mu^{(2)}) + g_1\left(\sum_{i=1}^n (z_i\sigma_i^{(1)} + (1-z_i)\sigma_i^{(2)})^2\right)^{\frac{1}{2}}\right| \le \tau \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \\
&= \mathbb{P}\left[\left\{\sum_{i=1}^n (z_i\mu^{(1)} + (1-z_i)\mu^{(2)}) + g_1\left(\sum_{i=1}^n (z_i\sigma_i^{(1)} + (1-z_i)\sigma_i^{(2)})^2\right)^{\frac{1}{2}} \le \tau\right\}\right. \\
&\quad \left. \cap \left\{-\tau \le \sum_{i=1}^n (z_i\mu^{(1)} + (1-z_i)\mu^{(2)}) + g_1\left(\sum_{i=1}^n (z_i\sigma_i^{(1)} + (1-z_i)\sigma_i^{(2)})^2\right)^{\frac{1}{2}}\right\} \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \\
&\overset{ii)}{\le} \mathbb{P}\left[g_1\left(\sum_{i=1}^n (z_i\sigma_i^{(1)} + (1-z_i)\sigma_i^{(2)})^2\right)^{\frac{1}{2}} \le \tau - \left|\sum_{i=1}^n (z_i\mu^{(1)} + (1-z_i)\mu^{(2)})\right| \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \\
&= \mathbb{P}\left[g_1\left(\sum_{i=1}^n (z_i\sigma_i^{(1)} + (1-z_i)\sigma_i^{(2)})^2\right)^{\frac{1}{2}} \le \tau - \left|\sum_{i=1}^n \frac{y_i^2/\|\boldsymbol{y}\|_2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2} \cdot z_i\boldsymbol{\beta}_{1,j^*}\right| \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \\
&=: \mathbb{P}\left[A | \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right], \qquad (12)
\end{aligned}$$

其中， i）由高斯随机变量在有限和运算下的封闭性成立，而 ii）由高斯 g_1 的对称性成立，且因为 $\mathbb{P}[B_1 \cap B_2] \le \min_{i\in\{1,2\}} \mathbb{P}[B_i]$。

利用引理 $_1(\delta)$、$\Omega_4(\delta)$ 中定义的高概率事件 ΩD.9，我们有

$$\begin{aligned}
&\mathbb{P}\left[\left|\frac{\langle \boldsymbol{X}_{j^*}, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \le \tau\right] = \int_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2}\int_{y,z} \mathbb{P}\left[\left|\frac{\langle X_{j^*,i}, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \le \tau \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] d\mathbb{P}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, z] \\
&\le \int_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2}\int_{\Omega_1(\delta)\cap\Omega_4(\delta)} \mathbb{P}\left[\left|\frac{\langle X_{j^*,i}, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \le \tau \middle| \boldsymbol{y}, \boldsymbol{z}\right] d\mathbb{P}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, z] \\
&\quad + \int_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2} \mathbb{P}\left[(\Omega_1(\delta)\cap\Omega_4(\delta))^{\complement} \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2\right] d\mathbb{P}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2] \\
&= \int_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2}\int_{\Omega_1(\delta)\cap\Omega_4(\delta)} \mathbb{P}\left[A|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] d\mathbb{P}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, z] + \mathbb{P}\left[(\Omega_1(\delta)\cap\Omega_4(\delta))^{\complement}\right] \\
&\le \int_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2}\int_{\Omega_1(\delta)\cap\Omega_4(\delta)} \mathbb{P}\left[A|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] d\mathbb{P}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, z] + \mathbb{P}\left[\Omega_1^{\complement}(\delta)\right] + \mathbb{P}\left[\Omega_4^{\complement}(\delta)\right]. \qquad (13)
\end{aligned}$$

现在令 $\tau = \sqrt{2(1+\epsilon/2)\log 2p}$ 和 $n \ge {}^{32}_{(1+\epsilon)}\overline{\phi^2\beta^2_{\min}}(\|\boldsymbol{\beta}\|_2^2+\sigma^2)\log p$，我们便得到，对于 $(y, z) \in \Omega_1(\delta) \cap \Omega_4(\delta)$:

$$\begin{aligned}
&\mathbb{P}\left[A|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \\
&= \mathbb{P}\left[g_1\left(\sum_{i=1}^n (z_i\sigma_i^{(1)} + (1-z_i)\sigma_i^{(2)})^2\right)^{\frac{1}{2}} \le \tau - \left|\sum_{i=1}^n \frac{y_i^2/\|\boldsymbol{y}\|_2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\cdot z_i\boldsymbol{\beta}_{1,j^*}\right| \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \\
&\le \mathbb{P}\left[g_1\left(\sum_{i=1}^n (z_i\sigma_i^{(1)} + (1-z_i)\sigma_i^{(2)})^2\right)^{\frac{1}{2}} \le \sqrt{2(1+\epsilon/2)\log 2p}\right. \\
&\qquad\qquad \left. - \left|\frac{\|\boldsymbol{y}\|_2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\phi\boldsymbol{\beta}_{1,j^*} - \sqrt{\frac{2(3+\delta)\log 2p}{(1-\delta)(1+\sigma^2/\|\boldsymbol{\beta}\|_2^2)}}\right| \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \\
&\le \mathbb{P}\left[g_1\left(\sum_{i=1}^n (z_i\sigma_i^{(1)} + (1-z_i)\sigma_i^{(2)})^2\right)^{\frac{1}{2}} \le \sqrt{2(1+\epsilon/2)\log 2p}\right. \\
&\qquad\qquad \left. - \left|\sqrt{\frac{n(1-\delta)}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}}\phi\boldsymbol{\beta}_{1,j^*} - \sqrt{\frac{2(3+\delta)\log 2p}{(1-\delta)(1+\sigma^2/\|\boldsymbol{\beta}\|_2^2)}}\right| \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \\
&\le \mathbb{P}\left[g_1\left(\sum_{i=1}^n (z_i\sigma_i^{(1)} + (1-z_i)\sigma_i^{(2)})^2\right)^{\frac{1}{2}} \le \sqrt{2(1+\epsilon/2)\log 2p}\right. \\
&\qquad\qquad \left. - \left|\sqrt{32(1+\epsilon)(1-\delta)\log 2p} - \sqrt{\frac{2(3+\delta)\log 2p}{(1-\delta)(1+\sigma^2/\|\boldsymbol{\beta}\|_2^2)}}\right| \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right], \qquad (14)
\end{aligned}$$

其中，当 $\delta > 0$ 足够小时，我们有

$$\sqrt{2(1+\epsilon/2)} - \left|\sqrt{32(1+\epsilon)(1-\delta)} - \sqrt{2(3+\delta)/(1-\delta)(1+\sigma^2/\|\boldsymbol{\beta}\|_2^2)}\right| \le -\sqrt{2}. \qquad (15)$$

因此，我们得到不等式（14）右侧为负，从而得到以下关于（y、z）$\in \Omega_1$（δ）$\cap\, \Omega_4$（δ）的

$$\mathbb{P}\left[A|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \le \mathbb{P}\left[g_1\left(\sum_{i=1}^n (z_i\sigma_i^{(1)} + (1-z_i)\sigma_i^{(2)})^2\right)^{1/2} \le -\sqrt{2\log 2p} \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right]$$

$$\leq \exp\left(-\frac{2\log 2p}{2\sum_{i=1}^{n}(z_i\sigma_i^{(1)}+(1-z_i)\sigma_i^{(2)})^2}\right)$$
$$\overset{i)}{\leq} \exp\left(-\frac{2\log 2p}{2\sum_{i=1}^{n}\frac{y_i^2}{\|\boldsymbol{y}\|_2^2}}\right) \leq \frac{1}{2p}, \tag{16}$$

其中，$i)$ 几乎必然成立，这是因为 $z_i \in \{0,1\}$ 和 $\left(1 - \frac{\boldsymbol{\beta}_{1,j^*}^2}{\|\boldsymbol{\beta}_1\|_2^2+\sigma^2}\right) \leq 1$。我们将其推广到 $j \in \mathcal{S}_2 \setminus \mathcal{S}_1$的情形，因此考虑类似的结果，其中将 ϕ 和 $\boldsymbol{\beta}_1$ 替换为 $(1-\phi)$ 和 $\boldsymbol{\beta}_2$。将所有结果代入（13），利用引理 D.9 并选择足够小的 $\delta > 0$ 以满足（15），我们得到对于 $n \geq 32 \min\{\phi^2, (1-\phi)^2\}\beta_{\min}^2 (1+\epsilon)(\|\boldsymbol{\beta}\|_2^2+\sigma^2)\log 2p$，

$$\mathbb{P}\left[\left|\frac{\langle \boldsymbol{X}_{j^*}, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \leq \sqrt{2(1+\epsilon/2)\log 2p}\right] \leq \frac{1}{p} + 4e^{-\frac{\delta^2 n}{8}} + \frac{39}{\delta^2 n}. \tag{17}$$

□

引理 D.3（$j \in \mathcal{S}_1 \cap \mathcal{S}_2$的一般界）。考虑 MSLR 的设定。设 $j^* \in \mathcal{S}_1 \cap \mathcal{S}_2$。那么，若 $n \geq 32\frac{(1+\epsilon)(\phi\beta_{1,j^*}+(1-\phi)\beta_{2,j^*})^2}{}(\|\boldsymbol{\beta}\|_2^2+\sigma^2)\log 2p$，对任意 $\epsilon \in (0,1)$，我们可得

$$\mathbb{P}\left[\left|\frac{\langle \boldsymbol{X}_{j^*}, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \leq \sqrt{2(1+\epsilon/2)\log 2p}\right] \leq \frac{1}{p} + 4e^{-\frac{\delta^2 n}{8}} + \frac{39}{\delta^2 n}.$$

证明。证明思路与引理 D.2 相同，主要区别在于 $X_{j^*,i}$ 的条件均值和方差，其中 $i \in [n]$。与之前一样，设 $g_i \overset{\text{独立同分布}}{\sim} \mathcal{N}(0,1)$（与 y_i、z_i、$\boldsymbol{\beta}_1$、$\boldsymbol{\beta}_2$ 独立），其中 $i \in [n]$，并将 $\delta > 0$ 留待后文选取。定义

- $\sigma_i^{(1)} := \frac{y_i}{\|\boldsymbol{y}\|_2}\sqrt{1-\frac{\boldsymbol{\beta}_{1,j^*}^2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}}$,
- $\sigma_i^{(2)} := \frac{y_i}{\|\boldsymbol{y}\|_2}\sqrt{1-\frac{\boldsymbol{\beta}_{2,j^*}^2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}}$,
- $\mu_i^{(1)} := \frac{y_i^2}{\|\boldsymbol{y}\|_2}\frac{\boldsymbol{\beta}_{1,j^*}}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}$,
- $\mu_i^{(2)} := \frac{y_i^2}{\|\boldsymbol{y}\|_2}\frac{\boldsymbol{\beta}_{2,j^*}}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}$,

并注意，对于 $\tau \in \mathbb{R}$,

$$\begin{aligned}
&\mathbb{P}\left[X_{j^*,i} \leq \tau | \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \\
&= \mathbb{P}\left[z_i\left(\frac{y_i\cdot\boldsymbol{\beta}_{1,j^*}}{\|\boldsymbol{\beta}\|_2^2+\sigma^2} + \sqrt{1-\frac{\boldsymbol{\beta}_{1,j^*}^2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}}g_i\right)\right. \\
&\qquad \left. +(1-z_i)\left(\frac{y_i\cdot\boldsymbol{\beta}_{2,j^*}}{\|\boldsymbol{\beta}\|_2^2+\sigma^2} + \sqrt{1-\frac{\boldsymbol{\beta}_{2,j^*}^2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}}g_i\right) \leq \tau \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, y_i, z_i\right],
\end{aligned}$$

这由引理D.4可得。于是我们有

$$\mathbb{P}\left[X_{j^*,i}\frac{y_i}{\|\boldsymbol{y}\|_2} \leq \tau \middle| \boldsymbol{y}, \boldsymbol{z}\right] = \mathbb{P}\left[z_i(\mu_i^{(1)}+\sigma_i^{(1)}g_i) + (1-z_i)(\mu_i^{(2)}+\sigma_i^{(2)}g_i) \leq \tau \middle| \boldsymbol{y}, \boldsymbol{z}\right].$$

然后，采用与（12）中相同的步骤，我们得到，对于 $\tau > 0$:

$$\begin{aligned}
&\mathbb{P}\left[\left|\frac{\langle \boldsymbol{X}_{j^*}, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \le \tau \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \\
&= \mathbb{P}\left[g_1\left(\sum_{i=1}^n (z_i\sigma_i^{(1)} + (1-z_i)\sigma_i^{(2)})^2\right)^{\frac{1}{2}} \le \tau - \left|\sum_{i=1}^n \frac{y_i^2/\|\boldsymbol{y}\|_2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\cdot(z_i\boldsymbol{\beta}_{1,j^*} + (1-z_i)\boldsymbol{\beta}_{2,j^*})\right| \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \\
&=: \mathbb{P}\left[A|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right]. \qquad (18)
\end{aligned}$$

利用引理 $_1(\delta)$ 中定义的高概率事件 $\Omega\Omega_3(\delta)$，根据与 (13) 相同的论证，我们有

$$\begin{aligned}
&\mathbb{P}\left[\left|\frac{\langle \boldsymbol{X}_{j^*}, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \le \tau\right] \\
&\le \int_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2}\int_{\Omega_1(\delta)\cap\Omega_3(\delta)} \mathbb{P}\left[A|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] d\mathbb{P}[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, z] + \mathbb{P}\left[\Omega_1^{\complement}(\delta)\right] + \mathbb{P}\left[\Omega_3^{\complement}(\delta)\right], \qquad (19)
\end{aligned}$$

现在令 $\tau = \sqrt{2(1+\epsilon/2)\log\ 2p}$ 和 $n \ge \frac{32}{\langle\boldsymbol{\beta}\rangle_{\min}^2}(1+\epsilon)(\|\boldsymbol{\beta}\|_2^2+\sigma^2)\log\ 2p$，我们得到，对于 $(y,z) \in \Omega_1(\delta) \cap \Omega_3(\delta)$,

$$\begin{aligned}
&\mathbb{P}\left[A|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \\
&= \mathbb{P}\left[g_1\left(\sum_{i=1}^n (z_i\sigma_i^{(1)} + (1-z_i)\sigma_i^{(2)})^2\right)^{\frac{1}{2}} \le \tau - \left|\sum_{i=1}^n \frac{y_i^2/\|\boldsymbol{y}\|_2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\cdot(z_i\boldsymbol{\beta}_{1,j^*} + (1-z_i)\boldsymbol{\beta}_{2,j^*})\right| \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \\
&\le \mathbb{P}\left[g_1\left(\sum_{i=1}^n (z_i\sigma_i^{(1)} + (1-z_i)\sigma_i^{(2)})^2\right)^{\frac{1}{2}} \le \sqrt{2(1+\epsilon/2)\log 2p}\right. \\
&\qquad\qquad \left. - \left|\frac{\|\boldsymbol{y}\|_2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}(\phi\boldsymbol{\beta}_{1,j^*} + (1-\phi)\boldsymbol{\beta}_{2,j^*}) - \sqrt{\frac{2(3+\delta)\log 2p}{(1-\delta)(1+\sigma^2/\|\boldsymbol{\beta}\|_2^2)}}\right| \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \\
&\le \mathbb{P}\left[g_1\left(\sum_{i=1}^n (z_i\sigma_i^{(1)} + (1-z_i)\sigma_i^{(2)})^2\right)^{\frac{1}{2}} \le \sqrt{2(1+\epsilon/2)\log 2p}\right. \\
&\qquad\qquad \left. - \left|\sqrt{\frac{n(1-\delta)}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}}(\phi\boldsymbol{\beta}_{1,j^*} + (1-\phi)\boldsymbol{\beta}_{2,j^*}) - \sqrt{\frac{2(3+\delta)\log 2p}{(1-\delta)(1+\sigma^2/\|\boldsymbol{\beta}\|_2^2)}}\right| \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \\
&\le \mathbb{P}\left[g_1\left(\sum_{i=1}^n (z_i\sigma_i^{(1)} + (1-z_i)\sigma_i^{(2)})^2\right)^{\frac{1}{2}} \le \sqrt{2(1+\epsilon/2)\log 2p}\right. \\
&\qquad\qquad \left. - \left|\sqrt{32(1+\epsilon)(1-\delta)\log 2p} - \sqrt{\frac{2(3+\delta)\log 2p}{(1-\delta)(1+\sigma^2/\|\boldsymbol{\beta}\|_2^2)}}\right| \middle| \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right], \qquad (20)
\end{aligned}$$

其中

$$\sqrt{2(1+\epsilon/2)} - \left|\sqrt{32(1+\epsilon)(1-\delta)} - \sqrt{2(3+\delta)/(1-\delta)(1+\sigma^2/\|\boldsymbol{\beta}\|_2^2)}\right| \le -\sqrt{2}, \qquad (21)$$

对于足够小的 $\delta > 0$。因此我们得到，(20) 的右侧为负，从而得到关于 $(y,\ z) \in \Omega_1(\delta) \cap \Omega_3(\delta)$ 的以下不等式，该不等式通过与 (16) 相同的步骤得到：

$$\mathbb{P}\left[A|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}, \boldsymbol{z}\right] \leq \frac{1}{2p}.$$

将上述结果结合到式(19)中，利用引理 D.9 并选取 $\delta > 0$ 充分小以满足条件，可得对于 $n \geq \frac{32}{\langle\boldsymbol{\beta}\rangle_{\min}^2}(1+\epsilon)(\|\boldsymbol{\beta}\|_2^2 + \sigma^2)\log\ 2p$, (20)

$$\mathbb{P}\left[\left|\frac{\langle \boldsymbol{X}_{j^*}, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \leq \sqrt{2(1+\epsilon/2)\log 2p}\right] \leq \frac{1}{p} + 4e^{-\frac{\delta^2 n}{8}} + \frac{39}{\delta^2 n}.$$

□

条件化引理。

引理 D.4（一般条件化引理）。考虑 MSLR 的设定。

对于 $j^* \in \mathcal{S}_1 \cap \mathcal{S}_2$ 和 $i \in [n]$，有：

$$X_{j^*,i}|(y_i, \boldsymbol{\beta}_1, z_i = 1) \sim \mathcal{N}\left(\frac{y_i \cdot \boldsymbol{\beta}_{1,j^*}}{\|\boldsymbol{\beta}_1\|_2^2 + \sigma^2}, \left(1 - \frac{\boldsymbol{\beta}_{1,j^*}^2}{\|\boldsymbol{\beta}_1\|_2^2 + \sigma^2}\right)\right), \tag{22}$$

$$X_{j^*,i}|(y_i, \boldsymbol{\beta}_2, z_i = 0) \sim \mathcal{N}\left(\frac{y_i \cdot \boldsymbol{\beta}_{2,j^*}}{\|\boldsymbol{\beta}_1\|_2^2 + \sigma^2}, \left(1 - \frac{\boldsymbol{\beta}_{2,j^*}^2}{\|\boldsymbol{\beta}_1\|_2^2 + \sigma^2}\right)\right). \tag{23}$$

对于 $j^* \in \mathcal{S}_1 \Delta \mathcal{S}_2$（不失一般性地，$j^* \in \mathcal{S}_1 \setminus \mathcal{S}_2$）和 $i \in [n]$，有

$$X_{j^*,i}|(y_i, \boldsymbol{\beta}_1, z_i = 1) \sim \mathcal{N}\left(\frac{y_i \cdot \boldsymbol{\beta}_{1,j^*}}{\|\boldsymbol{\beta}_1\|_2^2 + \sigma^2}, \left(1 - \frac{\boldsymbol{\beta}_{1,j^*}^2}{\|\boldsymbol{\beta}_1\|_2^2 + \sigma^2}\right)\right), \tag{24}$$

$$X_{j^*,i}|(y_i, z_i = 0) \sim \mathcal{N}(0, 1) \tag{25}$$

证明。为证明(22)和(24)，回顾

$$y_i = z_i(\boldsymbol{X}\boldsymbol{\beta}_1)_i + (1 - z_i)(\boldsymbol{X}\boldsymbol{\beta}_2)_i + w_i,$$

因此，给定 $z_i = 1$，我们有

$$y_i = \sum_{j \neq j^* \in \mathcal{S}_1} \boldsymbol{\beta}_{1,j} X_{j,i} + \boldsymbol{\beta}_{1,j^*} X_{j^*,i} + w_i,$$

这意味着

$$X_{j^*,i}\,|\,(y_i, \boldsymbol{\beta}_1, z_i = 1) = X_{j^*,i}\left|\left(\boldsymbol{\beta}_1, y_i = \sum_{j \neq j^* \in \mathcal{S}_1} \boldsymbol{\beta}_{1,j} X_{j,i} + \boldsymbol{\beta}_{1,j^*} X_{j^*,i} + w_i\right)\right..$$

因此，应用推论 D.7，我们得到 $X_{j^*,\ i}|\ (y_i,\ \boldsymbol{\beta}_1,\ z_i = 1) \sim \mathcal{N}\left(\ {}^{y_i\boldsymbol{\beta}_{1,\ j^*}}\!/_{\|\boldsymbol{\beta}_1\|_2^2+\sigma^2},\ \left(1 -\ {}^{\boldsymbol{\beta}_{1,\ j^*}^2}\!/_{\|\boldsymbol{\beta}_1\|_2^2+\sigma^2}\right)\right)$。式（23）中的结果可用同样的方法证明，只需将 $\boldsymbol{\beta}_1$ 替换为 $\boldsymbol{\beta}_2$即可。

对于（25），注意到给定 $z_i = 0$，我们有

$$y_i = \sum_{j \in \mathcal{S}_2} \boldsymbol{\beta}_{2,j} X_{j,i} + w_i,$$

根据定义，它与 $X_{j^*,\ i}$ 独立，因此 $X_{j^*,\ i}|(y_i, z_i = 0)$ 与 $X_{j^*,\ i} \sim \mathcal{N}(0, 1)$ 具有相同的分布。

□

引理 D.5（一般条件化引理）。设 $a \sim \mathcal{N}(0, \Sigma_{k\times k})$，且 $b \in \mathbb{R}^k$ 是一个固定向量。则：

$$a| \left(\sum_{j=1}^{k} b_j a_j = \eta \right) \sim \mathcal{N}\left(\eta v, B\Sigma_{k\times k} B^T\right)$$

其中，令$\mathbb{I}$ 表示全 1 向量，

$$v = \frac{1}{b^T \tilde{\Sigma}_{\mathbb{I}}} \tilde{\Sigma}_{\mathbb{I}}, \quad B = I_{k\times k} - vb^T, \quad \tilde{\Sigma} = \mathbb{E}\left[a(a \odot b)^T\right].$$

证明。设 B是一个确定性矩阵，且 $\eta := \sum_{j=1}^{k} b_j a_j$。那么（$Ba$， η）服从联合正态分布。我们将构造一个固定矩阵 B 和一个固定向量 v ，使得

- Ba 与 η相互独立
- $a = Ba + \eta v$。

如果上述结论成立，那么由独立性可得所需结果，即 $a|\eta \sim \mathcal{N}\left(\eta v, \ B\Sigma B^T\right)$。要使第一点成立，它们的协方差必须为零，这意味着

$$\mathbb{E}\left[Ba\eta\right] = \mathbb{E}\left[Ba(a \odot b)^T{}_{\mathbb{I}}\right] = B\tilde{\Sigma}_{\mathbb{I}} = 0.$$

与此同时，取 $B = I - vb^T$即可满足第二点。综合这两点，即得所需结果。

□

卡罗拉 ry D.6. 考虑 SLR 的情形，令 $j^* \in \mathcal{S}$，且令 $i \in [n]$。则有 t

$$X_{j^*,i} \left| \left(\sum_{j\in\mathcal{S}} X_{j,i} + w_i = \eta \right) \sim \mathcal{N}\left(\frac{\eta}{k+\sigma^2}, \left(1 - \frac{1}{k+\sigma^2}\right)\right)\right.$$

证明。 应用引理 D. 5 $\sum_{j\in\mathcal{S}} X_{j\,i} + w_i = \eta$ $[\boldsymbol{X}_{\mathcal{S}}\ w_i] \sim$ $\mathcal{N}\ 0\ \Sigma$ 在和 , 条件下： ，其中我们回顾 (,)，并且

$$\Sigma = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} \in \mathbb{R}^{(k+1)\times(k+1)}$$

从而得到

$$v = \frac{1}{k+\sigma^2} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ \sigma^2 \end{bmatrix}$$

and

$$B = \begin{bmatrix} (1 - \frac{1}{k+\sigma^2}) & -\frac{1}{k+\sigma^2} & \cdots & -\frac{1}{k+\sigma^2} \\ -\frac{1}{k+\sigma^2} & (1 - \frac{1}{k+\sigma^2}) & \cdots & -\frac{1}{k+\sigma^2} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{k+\sigma^2} & -\frac{1}{k+\sigma^2} & \cdots & (1 - \frac{1}{k+\sigma^2}) \end{bmatrix}.$$

注意到 $(B\Sigma B^T)_{j^*,\ j^*} = 1 - \frac{1}{k+\sigma^2}$，我们得到该结果。 □

推论 D.7。考虑 SLR 的情形，并令 $b \in \mathbb{R}^p$ 固定。令 $j^* \in \mathcal{S}$，且 $i \in [n]$。则有

$$X_{j^*,i} \left| \left(\sum_{j \in \mathcal{S}} b_j X_{j,i} + w_i = \eta \right) \sim \mathcal{N} \left(\frac{\eta \cdot b_{j^*}}{\|b\|_2^2 + \sigma^2}, \left(1 - \frac{b_{j^*}^2}{\|b\|_2^2 + \sigma^2} \right) \right) \right.$$

证明。 通过在和 上应用引理 D.5 进行条件化，即可得到该结果$_i + w_i = \eta$, $_i + w_i = \eta$。因此我们有

$$\tilde{\Sigma} = \mathbb{E} \left[\begin{bmatrix} X_{1,i} \\ X_{2,i} \\ \vdots \\ w_i \end{bmatrix} \begin{bmatrix} b_1 X_{1,i} \\ b_2 X_{2,i} \\ \vdots \\ w_i \end{bmatrix}^T \right] = \begin{bmatrix} b_1 & 0 & \cdots & 0 \\ 0 & b_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} \in \mathbb{R}^{(k+1)\times(k+1)}$$

从而得到

$$v = \frac{1}{\|b\|_2^2 + \sigma^2} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ \sigma^2 \end{bmatrix}$$

and

$$B = \begin{bmatrix} (1 - \frac{b_1^2}{\|b\|_2^2+\sigma^2}) & -\frac{b_1 b_2}{\|b\|_2^2+\sigma^2} & \cdots & -\frac{b_1}{\|b\|_2^2+\sigma^2} \\ -\frac{b_2 b_1}{\|b\|_2^2+\sigma^2} & (1 - \frac{b_2^2}{\|b\|_2^2+\sigma^2}) & \cdots & -\frac{b_2}{\|b\|_2^2+\sigma^2} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\sigma^2 b_1}{\|b\|_2^2+\sigma^2} & -\frac{\sigma^2 b_2}{\|b\|_2^2+\sigma^2} & \cdots & (1 - \frac{\sigma^2}{\|b\|_2^2+\sigma^2}) \end{bmatrix}.$$

注意到 $(B\Sigma B^T)_{j^*,\ j^*} = 1 - \frac{b_{j^*}^2}{\|b\|_2^2+\sigma^2}$，即可得出结果。 □

引理 D.8。 考虑定义 MSLR 1.1 中的设定。则 y 与 z 相互独立。

证明。设 $g \sim \mathcal{N} \in (0, 1)$，且 A 是 σ代数中的一个事件。根据贝叶斯法则，对每个 $i \in [n]$，均有

$$\begin{aligned} \mathbb{P}\left[z_i = 1 | y_i \in A\right] &= \frac{\mathbb{P}\left[y_i \in A | z_i = 1\right] \mathbb{P}[z_i = 1]}{\mathbb{P}[y]} \\ &= \frac{\mathbb{P}\left[y_i \in A | z_i = 1\right] \mathbb{P}[z_i = 1]}{\mathbb{P}\left[y_i \in A | z_i = 1\right] \mathbb{P}[z_i = 1] + \mathbb{P}\left[y_i \in A | z_i = 0\right] \mathbb{P}[z_i = 0]} \\ &= \frac{\mathbb{P}[\sqrt{\|\boldsymbol{\beta}\|_2^2 + \sigma^2} g \in A]\phi}{\mathbb{P}[\sqrt{\|\boldsymbol{\beta}\|_2^2 + \sigma^2} g \in A]\phi + \mathbb{P}[\sqrt{\|\boldsymbol{\beta}\|_2^2 + \sigma^2} g \in A](1 - \phi)} \\ &= \phi \\ &= \mathbb{P}[z_i = 1], \end{aligned}$$

对于情形 $z_i = 0$，类似的结果也成立。回想 y_i 和 z_i 在 i 索引之间独立同分布，即可得出该结果。 □

高概率事件与集中不等式

引理 D.9（高概率事件）。考虑 MSLR 的设定。令 $g_q \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$，对于 $q \in [p]$ 和 $\delta \in (0,1)$，$t > 0$，$j^* \in [p]$。定义以下事件，这些事件对于分析 MSLR 上的 CORR 是必要的：

$$\Omega_1(\delta) := \left\{ (\|\boldsymbol{\beta}\|_2^2 + \sigma^2)n(1-\delta) \le \|\boldsymbol{y}\|^2 \le (\|\boldsymbol{\beta}\|_2^2 + \sigma^2)n(1+\delta) \right\},$$

$$\Omega_2(t) := \left\{ \max_{q \in [p]} |g_q| \le \sqrt{2\log 2p} + \sqrt{2t \log 2p} \right\},$$

$$\Omega_3(\delta) := \left\{ \left| \sum_{i=1}^n (z_i \boldsymbol{\beta}_{1,j^*} + (1-z_i)\boldsymbol{\beta}_{2,j^*}) \frac{y_i^2/\|\boldsymbol{y}\|_2}{\|\boldsymbol{\beta}\|_2^2 + \sigma^2} \right.\right.$$
$$\left.\left. -(\phi \boldsymbol{\beta}_{1,j^*} + (1-\phi)\boldsymbol{\beta}_{2,j^*}) \frac{\|\boldsymbol{y}\|_2}{\|\boldsymbol{\beta}\|_2^2 + \sigma^2} \right| < \sqrt{\frac{2(3+\delta)\log(2p)}{(1-\delta)(1+\sigma^2/\|\boldsymbol{\beta}\|_2^2)}} \right\},$$

$$\Omega_4(\delta) := \left\{ \left| \sum_{i=1}^n z_i \boldsymbol{\beta}_{1,j^*} \frac{y_i^2/\|\boldsymbol{y}\|_2}{\|\boldsymbol{\beta}\|_2^2 + \sigma^2} - \phi \boldsymbol{\beta}_{1,j^*} \frac{\|\boldsymbol{y}\|_2}{\|\boldsymbol{\beta}\|_2^2 + \sigma^2} \right| < \sqrt{\frac{2(3+\delta)\log(2p)}{(1-\delta)(1+\sigma^2/\|\boldsymbol{\beta}\|_2^2)}} \right\},$$

$$\Omega_5(\delta) := \left\{ \left| \|\boldsymbol{y}\|_4^4 - 3n(\|\boldsymbol{\beta}\|_2^2 + \sigma^2)^2 \right| < \delta n (\|\boldsymbol{\beta}\|_2^2 + \sigma^2)^2 \right\}.$$

那么，上述事件均以高概率发生。具体而言，

- $\mathbb{P}\left[\Omega_1^{\complement}(\delta)\right] \le 2e^{-\frac{\delta^2 n}{8}}$ *as per Example 2.11 in Wainwright [2019] noting that* $y_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \|\boldsymbol{\beta}\|_2^2 + \sigma^2)$ *both marginally and conditionally on* z *(similarly for the setting of* SLR*).*
- $\mathbb{P}\left[\Omega_2^{\complement}(t)\right] \le \frac{1}{(2p)^t}$ *as per Lemma 5.2 in van Handel [2014].*
- $\mathbb{P}\left[\Omega_3^{\complement}(\delta) \middle| \Omega_1(\delta), \Omega_5(\delta)\right] \le \frac{1}{2p}$ *as per Lemma D.11 and consequently* $\mathbb{P}\left[\Omega_3^{\complement}(\delta)\right] \le \frac{1}{2p} + 2e^{-\frac{\delta^2 n}{8}} + \frac{39}{\delta^2 n}$.
- $\mathbb{P}\left[\Omega_4^{\complement}(\delta) \middle| \Omega_1(\delta), \Omega_5(\delta)\right] \le \frac{1}{2p}$ *as per Lemma D.11 with* $\boldsymbol{\beta}_{2,j^*} = 0$ *and consequently* $\mathbb{P}\left[\Omega_4^{\complement}(\delta)\right] \le \frac{1}{2p} + 2e^{-\frac{\delta^2 n}{8}} + \frac{39}{\delta^2 n}$.
- $\mathbb{P}\left[\Omega_5^{\complement}(\delta)\right] \le \frac{39}{\delta^2 n}$ *as per Lemma D.10.*

引理 D.10（快速 ℓ_4 范数集中）。设 $\boldsymbol{y} \in \mathbb{R}^n$ 服从 $y_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$，其中 $i \in [n]$。那么，对于 $\delta > 0$，我们有：

$$\mathbb{P}\left[\left| \|\boldsymbol{y}\|_4 - 3n(\sigma^2)^2 \right| \ge \delta n (\sigma^2)^2\right] \le \frac{39}{\delta^2 n}.$$

证明。该证明直接应用标准的切比雪夫不等式即可。首先回顾一下，p^{th} 中心高斯矩为（方差）$^{\frac{p}{2}}$ $(p-1)!!$，因此由独立性可得 $\mathbb{E}\left[\|\boldsymbol{y}\|_4^4\right] = 3n(\sigma^2)^2$。然后对Boucheron 等人 [2013]的 $t > 0$应用切比雪夫不等式：

$$\mathbb{P}\left[\left| \|\boldsymbol{y}\|_4^4 - \mathbb{E}\left[\|\boldsymbol{y}\|_4^4\right] \right| \ge t\right] \le \frac{n\mathrm{Var}(y_1^4)}{t^2}$$

$$
\begin{aligned}
&= \frac{n\left(\mathbb{E}y_1^8 - \left(\mathbb{E}y_1^4\right)^2\right)}{t^2} \\
&= \frac{n(\sigma^2)^4((7-1)!! - 9)}{t^2} = \frac{39n(\sigma^2)^4}{t^2}.
\end{aligned}
$$

令 $t = n(\sigma^2)^2$，可得

$$
\mathbb{P}\left[\left|\|\boldsymbol{y}\|_4^4 - \mathbb{E}\left[\|\boldsymbol{y}\|_4^4\right]\right| \geq n(\sigma^2)^2\right] \leq \frac{39n(k+\sigma^2)^4}{n^2(\sigma^2)^4} = \frac{39}{n}.
$$

□

Lemma D.11. 考虑 MSLR 的情形。对于 $t > 0$、$\delta > 0$、$j^* \in [p]$，我们有

$$
\begin{aligned}
&\mathbb{P}\left[\left|\sum_{i=1}^n (z_i\boldsymbol{\beta}_{1,j^*} + (1-z_i)\boldsymbol{\beta}_{2,j^*})\frac{y_i^2/\|\boldsymbol{y}\|_2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2} - (\phi\boldsymbol{\beta}_{1,j^*} + (1-\phi)\boldsymbol{\beta}_{2,j^*})\frac{\|\boldsymbol{y}\|_2^2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right| \geq t \middle| \Omega_1(\delta), \Omega_5(\delta)\right] \\
&\leq \exp\left(-\frac{(1-\delta)}{2(3+\delta)}(1+\sigma^2/\|\boldsymbol{\beta}\|_2^2)t^2\right)
\end{aligned}
$$

证明。考虑 $y \in \Omega_1(\delta) \cap \Omega_5(\delta)$ 引理 D.9中的情况。我们首先应用 Hoeffding 不等式（参见 Boucheron 等人 [2013]），然后利用 $\Omega_1(\delta)$ 和 $\Omega_5(\delta)$ 的定义，得到：

$$
\begin{aligned}
&\mathbb{P}\left[\left|\sum_{i=1}^n (z_i\boldsymbol{\beta}_{1,j^*} + (1-z_i)\boldsymbol{\beta}_{2,j^*})\frac{y_i^2/\|\boldsymbol{y}\|_2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2} - (\phi\boldsymbol{\beta}_{1,j^*} + (1-\phi)\boldsymbol{\beta}_{2,j^*})\frac{\|\boldsymbol{y}\|_2^2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right| \geq t \middle| \boldsymbol{y}\right] \\
&\leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n(\boldsymbol{\beta}_{1,j^*} - \boldsymbol{\beta}_{2,j^*})^2\frac{y_i^4/\|\boldsymbol{y}\|_2^2}{(\|\boldsymbol{\beta}\|_2^2+\sigma^2)^2}}\right) \\
&= \exp\left(-\frac{2t^2}{(\boldsymbol{\beta}_{1,j^*} - \boldsymbol{\beta}_{2,j^*})^2\frac{\|\boldsymbol{y}\|_4^4}{\|\boldsymbol{y}\|_2^2}\frac{1}{(\|\boldsymbol{\beta}\|_2^2+\sigma^2)^2}}\right) \\
&\leq \exp\left(-\frac{2t^2}{(\boldsymbol{\beta}_{1,j^*} - \boldsymbol{\beta}_{2,j^*})^2\frac{3n(\|\boldsymbol{\beta}\|_2^2+\sigma^2)^2+\delta n(\|\boldsymbol{\beta}\|_2^2+\sigma^2)^2}{\|\boldsymbol{y}\|_2^2}\frac{1}{(\|\boldsymbol{\beta}\|_2^2+\sigma^2)^2}}\right) \\
&\leq \exp\left(-\frac{2t^2}{(\boldsymbol{\beta}_{1,j^*} - \boldsymbol{\beta}_{2,j^*})^2\frac{3n(\|\boldsymbol{\beta}\|_2^2+\sigma^2)^2+\delta n(\|\boldsymbol{\beta}\|_2^2+\sigma^2)^2}{n(\|\boldsymbol{\beta}\|_2^2+\sigma^2)(1-\delta)}\frac{1}{(\|\boldsymbol{\beta}\|_2^2+\sigma^2)^2}}\right) \\
&= \exp\left(-\frac{2(\|\boldsymbol{\beta}\|_2^2+\sigma^2)t^2}{(\boldsymbol{\beta}_{1,j^*} - \boldsymbol{\beta}_{2,j^*})^2\frac{3+\delta}{1-\delta}}\right) \\
&\overset{i)}{\leq} \exp\left(-\frac{2(\|\boldsymbol{\beta}\|_2^2+\sigma^2)t^2}{4\|\boldsymbol{\beta}\|_2^2\frac{3+\delta}{1-\delta}}\right) \\
&= \exp\left(-\frac{(1-\delta)}{2(3+\delta)}(1+\sigma^2/\|\boldsymbol{\beta}\|_2^2)t^2\right),
\end{aligned}
$$

其中，$i)$ 由 $(\boldsymbol{\beta}_{1,\,j^*} - \boldsymbol{\beta}_{2,\,j^*})^2 \leq \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\|_2^2$ 以及三角不等式推出。对上述结果应用全概率定律，即可得到所需结论。 □

D.2 CORR 对 MSLR − D

定理 2.5 的证明。首先注意到，由于 $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \sim \mathcal{P}_{\|\boldsymbol{\beta}\|_2}(\mathcal{D})$，这两个信号的范数相等。因此，SB-MSLR 区域的补集可以等价地表示为条件 $\phi\boldsymbol{\beta}_1 + (1-\phi)\boldsymbol{\beta}_2 \neq 0$。

令 CORR$(\boldsymbol{X}, \boldsymbol{y})$ 表示在输入 $\boldsymbol{X}, \boldsymbol{y}$ 上运行 CORR 的输出。考虑检验函数

$$g\left(\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{y}\end{bmatrix}\right) := \begin{cases} \mathtt{p} & \mathtt{CORR}(\boldsymbol{X}, \boldsymbol{y}) \neq \emptyset \\ \mathtt{q} & \mathtt{CORR}(\boldsymbol{X}, \boldsymbol{y}) = \emptyset \end{cases}.$$

令 $\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{y}\end{bmatrix} \sim \mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})$。回想一下，CORR 输出以下集合

$$\mathtt{CORR}(\boldsymbol{X}, \boldsymbol{y}) = \left\{ j \in [p] : \left| \frac{\langle \boldsymbol{X}_j, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|_2} \right| \geq \sqrt{2(1+\epsilon/2)\log 2p} \right\}.$$

如定理 D.1 的证明中所示，我们注意到

$$\mathbb{P}\left[\max_{q \in [p]} \left| \frac{\langle \boldsymbol{X}_q, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|_2} \right| \geq \sqrt{2(1+\epsilon/2)\log 2p} \right] \leq \mathbb{P}\left[\max_{q \in [p]} \left| \frac{\langle \boldsymbol{X}_q, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|_2} \right| \geq \sqrt{2\log 2p} + \frac{\epsilon}{2\sqrt{8}} \sqrt{\log 2p} \right]. \quad (26)$$

注意到，对于 $\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{y}\end{bmatrix} \sim \mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})$，我们有 $\frac{\langle \boldsymbol{X}_q, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|_2} \overset{\boldsymbol{y} \text{ 独立同分布}}{\sim} \mathcal{N}(0, 1)$，应用引理 D.9 于 (26) 并得到

$$\mathbb{P}\left[\max_{q \in [p]} \left| \frac{\langle \boldsymbol{X}_j, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|_2} \right| \geq \sqrt{2(1+\epsilon/2)\log 2p} \right] \leq (2p)^{-\frac{1}{8}\left(\frac{\epsilon}{2}\right)^2},$$

因此，在 $\mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})$ 下，以概率 $1 - o(1)$ 有 $g\left(\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{y}\end{bmatrix}\right) = \mathtt{q}$。

反过来，设 $\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{y}\end{bmatrix} \sim \mathbb{P}(\boldsymbol{X}, \boldsymbol{y})$。令 $J \subseteq \mathrm{supp}(\boldsymbol{\beta}_1) \cup \mathrm{supp}(\boldsymbol{\beta}_2)$，使得 $\phi\boldsymbol{\beta}_{1,j} + (1-\phi)\boldsymbol{\beta}_{2,j} \neq 0$对于 $j \in J$。由引理 D.3，可得：

$$\begin{aligned}
\mathbb{P}\left[|\mathtt{CORR}(\boldsymbol{X}, \boldsymbol{y})| = 0\right] &= \mathbb{P}\left[\cap_{j \in [p]} \left\{ \left| \frac{\langle \boldsymbol{X}_j, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|_2} \right| \leq \sqrt{2(1+\epsilon/2)\log 2p} \right\} \right] \\
&\leq \mathbb{P}\left[\cap_{j \in J} \left\{ \left| \frac{\langle \boldsymbol{X}_j, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|_2} \right| \leq \sqrt{2(1+\epsilon/2)\log 2p} \right\} \right] \\
&\leq \mathbb{P}\left[\left| \frac{\langle \boldsymbol{X}_{J_1}, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|_2} \right| \leq \sqrt{2(1+\epsilon/2)\log 2p} \right] \\
&\leq \left(\frac{1}{p} + 6e^{\frac{\delta^2 n}{w}} + \frac{39}{\delta^2 n} \right) = o(1),
\end{aligned}$$

因此，在 $\mathbb{P}(\boldsymbol{X}, \boldsymbol{y})$ 条件下，$g\left(\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{y}\end{bmatrix}\right) = \mathtt{p}$ 的概率为 $1 - o(1)$。

□

D.3 MSLR 的恢复算法

MSLR 的通用恢复算法 MSLR MSLR 的恢复算法 MSLR在无噪声和平衡区域下见定理 2.6。我们按照 MSLR 中的定义来度量恢复误差，参见 Chen 等人 [2014]，

$$\rho(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) := \min\left\{ \left\| \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \right\|_2 + \left\| \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \right\|_2, \left\| \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_2 \right\|_2 + \left\| \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_1 \right\|_2 \right\},$$

^ ^ ^其中 $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ 和 $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$。

定理的证明 2.6。 在 $\sigma = 0$, $\phi \neq 1/2$ (无噪声)的情况下，我们先应用 CORR ，然后应用 Yi 等人 [2014] 的交替最小化（ AM ）算法。定理 D.12 表明，该方法在我们关注的范围内能够成功。

在 $\phi = 1/2$ （平衡）的情况下，SNR $= \Omega(k)$，我们首先应用 CORR，然后应用 Chen 等人提出的算法 [2014]。定理 D.20 证明，在高信噪比区域 SNR $= \Omega(1)$ 中，我们有 $\rho(\hat{\theta}, \theta) = \Theta\left(\sigma\sqrt{\frac{k}{n}}\right)$。为了对所有允许的有限 n 实现精确恢复，我们需要 $\sqrt{\frac{\sigma^2 k}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}(\,)\log p} \to 0$，这在 SNR $= \|\boldsymbol{\beta}\|_2^2/\sigma^2 = \Omega(k)$ 时成立，意味着相对于支撑集存在非消失的信噪比。根据假设，这一条件成立，因此 CORR 与算法 6 一起，能够在所关注的区域内成功实现渐近精确恢复（允许对$\boldsymbol{\beta}_1$ 和 $\boldsymbol{\beta}_2$重新标记）。 □

无噪声、不平衡情况下的恢复算法 MSLR 本节将展示CORR 算法可以将 MSLR 问题归约为一个稠密问题，其中 n, $k = \Theta(p)$，且不假设信号是次线性稀疏的。对于这种稠密混合线性回归情形，最先进的算法随后可以从 $\boldsymbol{\beta}_1$ 中推断出 $\boldsymbol{\beta}_2$。在 $\phi \neq 1/2$的无噪声情形下，我们可以将 CORR与现有的多项式时间交替最小化（AM）算法结合使用，该算法由 Yi 等人 [2014] 提出，从而通过常数步运算彻底解决 MSLR 问题。回顾一下，定理 D.1 仅对联合信号的支撑集恢复提供了保证，而现在，在缩减后的联合支撑集上执行 AM 算法后，可以从 $\boldsymbol{\beta}_1$中推断出 $\boldsymbol{\beta}_2$。定义混合比例：

$$\frac{n_1}{n} := \frac{\sum_{i=1}^n z_i}{n}$$
$$\frac{n_2}{n} := \frac{\sum_{i=1}^n (1-z_i)}{n}.$$

我们先陈述定理。

定理 D.12（无噪声 MSLR 中 CORR+ AM 的成功）。考虑具有参数 $p, n, k, \sigma = 0$, $\phi \neq 1/2$, $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \sim \mathcal{P}_{\|\boldsymbol{\beta}\|_2}(\mathcal{D})$ 的一般 MSLR 设置。假设

$$n \geq \frac{32}{\min\{\phi^2\boldsymbol{\beta}_{min}^2, (1-\phi)^2\boldsymbol{\beta}_{min}^2, \langle\boldsymbol{\beta}\rangle_{min}^2\}}(1+\epsilon)\|\boldsymbol{\beta}\|_2^2 \log 2p$$

用于 $\epsilon \in (0,1)$ 在 CORR 中。则以至少 $1 - c_1(\frac{k}{p} + ke^{-c_2 n} + \frac{k}{n} + \frac{1}{p^{c_2}})$ 的概率，对于常数ˆ ˆˆc_1, $c_2 > 0$, $\theta = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ 的 CORR+ 算法 3 + 算法 5 满足

$$\rho(\hat{\theta}, \theta) = 0.$$

定理 D.12 的证明。 根据定理 D.1，运行 CORR，即可恢复信号的联合支撑集（大小至多为 $2k$），$\boldsymbol{\beta}_1\boldsymbol{\beta}_2$，概率至少为 $1 - c_1(\frac{k}{p} + ke^{-c_2 n} + \frac{k}{n} + \frac{1}{p^{c_2}})$。

运行 CORR 并识别出 $< 2k$ 联合支撑集索引后，我们通过从设计矩阵 $\boldsymbol{X}$ 中删除所有其他列，将回归问题限制在这些索引上（因为这些列不对应于 $\boldsymbol{\beta}_1$ 或 $\boldsymbol{\beta}_2$的支撑索引，故不会影响输出 y ）。随后，我们需要求解一个双成分回归混合模型，其中包含 n 个样本，且信号维度介于 k 和 $2k$ 之间（重要的是，维度不再是 p）。

从这一步开始，思路是使用算法 $\boldsymbol{\beta}_1^{(0)}$ ， $\boldsymbol{\beta}_2^{(0)}$ 进行谱初始化；命题 3 D.14 提供相应的保证。将 ($\boldsymbol{\beta}_1^{(0)}$ ， $\boldsymbol{\beta}_2^{(0)}$) 输入算法 5 ， 定理 D.15

在给定此初始化的情况下，该算法可保证几何误差衰减，并运行算法 5 ，迭代有限次；命题 D.16对此给出了保证。

命题 n的样本量条件D.14 已满足，因为假设 $n \gtrsim \|\boldsymbol{\beta}\|_2^2 \log 2p \geq \boldsymbol{\beta}_{\min}^2 k \log 2p$ 以及 $k \log 2p = \omega(k \log^2 k)$。定理 D.15 的条件由命题 D.14的结果满足。命题 D.16 的条件可通过运行算法 5(重采样)，使用 $O(k\log^2 k)$ 个样本来满足（见备注 D.17）。

还需证明 $n_1 \neq n_2$，因为备注 D.13要求如此。回顾一下 $n_1 = \sum_{i=1}^n z_i$和 $n_2 = \sum_{i=1}^n (1 - z_i)$，以及 z_i 相互独立，且服从伯努利分布(ϕ)。不失一般性，假设$\phi < 1/2$，则存在一个(0,1) 使得，

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^n z_i \geq 1/2\right] \leq \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^n z_i \geq (1+\delta)\phi\right] \leq e^{-\Theta(n)},$$

应用标准的 Chernoff界后即可得到。该高概率结论可并入 CORR 提供的 $1 - c_1($ $\frac{k}{p} + ke^{-c_2 n} + \frac{k}{n} + \frac{1}{p^{c_2}})$ 高概率结论中，并选择调整后的常数 $c_1, c_2 > 0$。

因此我们得出结论，运行CORR 后接算法 3 再接算法 5 ，我们可以在有限 $\rho(\theta^{(t)}, \theta) = 0$ 内以至少 t 的概率 $1 - c_1(\frac{k}{p} + ke^{-c_2 n} + \frac{k}{n} + \frac{1}{p^{c_2}})$。 □

他们的初始化算法基于以下半正定矩阵：

$$M := \frac{1}{n}\sum_{i=1}^n y_i^2 \boldsymbol{x}_i \otimes \boldsymbol{x}_i$$

它是一个矩阵的无偏估计量，该矩阵的两个最大特征向量张成的空间与 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ 张成的空间相同。

注释D.13。 Fan 等人 [2018] 指出，当混合频率彼此相等时（$n_1 = n_2$）， $\mathbb{E}M$ 的前两个特征向量将不会是 $\boldsymbol{\beta}_1$， $\boldsymbol{\beta}_2$。因此，他们的算法仅适用于 $\phi \neq 1/2$ 且 $\sigma = 0$ （无噪声）的情况。

除 $\phi = 1/2$之外，当混合比例已知时，可以通过算法 3以闭式形式计算$\boldsymbol{\beta}_1$， $\boldsymbol{\beta}_2$ ，其中

$$sign(b) = \begin{cases} 1, & b = 1 \\ -1, & b = 2. \end{cases}$$

接下来，我们陈述Fan 等人 [2018] 提出的迭代算法及其保证 .

算法 3：利用比例信息初始化

数据：输入： n_1, n_2, 样本 $\{(y_i, \boldsymbol{x}_i)\ i = 1, 2, ..., n\}M \leftarrow \frac{1}{N}\sum_{i=1}^N y_i^2 \boldsymbol{x}_i \otimes \boldsymbol{x}_i$计算前 2 个特征向量和特征值 $(v_b, \lambda_b)\ b = 1,2$ of $(M - I)/2$计算 $\boldsymbol{\beta}_b^{(0)} = \sqrt{\frac{1-\Delta_b}{2}} v_b + sign(b)$ $\sqrt{\frac{1+\Delta_b}{2}} v_{-b}$, 其中 $\Delta_b = \frac{(\lambda_b - \lambda_{-b})2}{2+n_b^2 - n_{-b}^2}(\lambda_{-b} - \lambda_b)n_b$, $b = 1, 2$ 返回 $\boldsymbol{\beta}_1^{(0)}, \boldsymbol{\beta}_2^{(0)}$

算法4：AM

数据：初始 $\boldsymbol{\beta}_1^{(0)}$，$\boldsymbol{\beta}_2^{(0)}$，迭代次数 t_0，样本 $\{(y_i, \boldsymbol{x}_i)\ i = 1,\ 2,\ ...,$ 对 $t = 0 \cdots\ t_0 - 1$

执行

$J_1, J_2 \leftarrow \emptyset$

for $i = 1, 2, \cdots, N$ **do**

if $\left|y_i - \langle \boldsymbol{x}_i, \boldsymbol{\beta}_1^{(t)} \rangle\right| < \left|y_i - \langle \boldsymbol{x}_i, \boldsymbol{\beta}_2^{(t)} \rangle\right|$ **then**

$J_1 \leftarrow J_1 \cup \{i\}$

else

$J_2 \leftarrow J_2 \cup \{i\}$

end

end

$\boldsymbol{\beta}_1^{(t+1)} \leftarrow \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^k} \|\boldsymbol{y}_{J_1} - \boldsymbol{X}_{J_1}\boldsymbol{\beta}\|_2$

$\boldsymbol{\beta}_2^{(t+1)} \leftarrow \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^k} \|\boldsymbol{y}_{J_2} - \boldsymbol{X}_{J_2}\boldsymbol{\beta}\|_2$

end

返回 $\boldsymbol{\beta}_1^{(t_0)}$ $\boldsymbol{\beta}_2^{(t_0)}$

命题D.14。 *[Yi 等人， 2014]* 考虑算法 3中的初始化方法。给定任意常数 $\widehat{c} < 1/2$，以至少 $1 - \frac{1}{p^2}$ 的概率，该方法会生成一个满足以下条件的初始化 $(\boldsymbol{\beta}_1^{(0)}, \boldsymbol{\beta}_2^{(0)})$：

$$\rho(\theta^{(0)}, \theta) \ \leq\ \widehat{c}\, \min\{n_1/n, n_2/n\}\, \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\|_2,$$

if

$$n \geq c_1 \left(\frac{1}{\widetilde{\delta}}\right)^2 p\, \log^2 p.$$

这里， c_1 是一个取决于 $\widehat{c}$ 的常数。并且

$$\sqrt{\widetilde{\delta}} = \widehat{c}\sqrt{\min\{n_1/n, n_2/n\}}^3 \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\|_2 (\sqrt{1-\kappa})\kappa,$$

其中 $\kappa = \sqrt{1 - 4(1 - \langle \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \rangle^2)\frac{n_1}{n}\frac{n_2}{n}}$。

算法 5：带重采样的AM

数据：初始 $\boldsymbol{\beta}_1^{(0)}, \boldsymbol{\beta}_2^{(0)}$，迭代次数 t_0，样本 $\{(y_i, \boldsymbol{x}_i)\ i = 1、2、...、\ N\}$

将样本 $\{(y_i, \boldsymbol{x}_i)\}$ 划分为 t_0 个互不相交的集合： $\mathcal{S}_1$、...、 $\mathcal{S}_{t_0}$

对于 $t = 1, \cdots, t_0$， 执行

使用 $\mathcal{S}_t$ 运行算法 4 ， 以 $(\boldsymbol{\beta}_1^{(t-1)}, \boldsymbol{\beta}_2^{(t-1)})$ 作为初始值，并返回 $(\boldsymbol{\beta}_1^t, \boldsymbol{\beta}_2^t)$结束

返回 $\boldsymbol{\beta}_1^{(t_0}$， $\boldsymbol{\beta}_2^{(t_0)}$

定理 D.15。 Yi 等人 *[2014]* 考虑算法 5 中的一次迭代。对于固定的 $(\boldsymbol{\beta}_1^{(t-1)}$， $\boldsymbol{\beta}_2^{(t-1)})$，存在绝对常数 $\widetilde{c}$， c_1， c_2 ，使得如果

$$\rho(\theta^{(t-1)}, \theta) \ \leq\ \widetilde{c}\, \min\{n_1/n, n_2/n\}\, \|\boldsymbol{\beta}_1^* - \boldsymbol{\beta}_2^*\|_2,$$

并且，如果该次迭代中的样本数满足

$$|\mathcal{S}_t| \ \geq\ \left(\frac{c_1}{\min\{n_1/n, n_2/n\}}\right) p,$$

那么，以大于 $1-\exp(-c_2 p)$ 的概率，下一阶段的误差会几何级数式递减，即

$$\rho(\theta^{(t)},\theta) \le \frac{1}{2}\rho(\theta^{(t-1)},\theta)$$

命题 D.16。（精确恢复）Yi 等人 *[2014]* 存在绝对常数 c_1，c_2 使得如果

$$\rho(\theta^{(t-1)},\theta) \le \frac{c_1}{p^2}\|\boldsymbol{\beta}_1-\boldsymbol{\beta}_2\|_2$$

and

$$\frac{1}{\min\{n_1/n, n_2/n\}}p < |\mathcal{S}_t| < c_2 p,$$

那么，以大于 $1-\frac{1}{p}$ 的概率，

$$\rho(\theta^{(t)},\theta) = 0.$$

注释 D.17。很容易看出，且 Fan 等人 [2018]（第 11–12 页，命题 4 的上下文）指出，使用定理 5 中给出的保证运行算法，需要 O（p 对数^2 p）个样本才能获得 ρ（$\theta^{(t-1)}$，θ）$\le \frac{c_1}{p^2}\|\boldsymbol{\beta}_1^*-\boldsymbol{\beta}_2^*\|_2$，其中常数 $c_1>0$ 取为适用于命题D.16的值。

平衡情形下的 MSLR 算法 考虑 Chen 等人 [2014]，所述的线性回归混合模型（MLR），其中 n、$k=\Theta(p)$ 和 σ^2 已知。此外，考虑平衡区域，其中 $\phi=1/2$，但对于任意 $j\in\mathcal{S}_1\cap\mathcal{S}_2$（$\langle\boldsymbol{\beta}\rangle^2_{\min}\neq 0$），$\boldsymbol{\beta}_{1,j}\neq-\boldsymbol{\beta}_{2,j}$。我们声称，在该区域内，可以通过使用 CORR 恢复 $\boldsymbol{\beta}_1$、$\boldsymbol{\beta}_2$ 的联合支撑，然后运行 Chen 等人 [2014] 提出的适用于一般混合线性回归的算法，来解决上述 MSLR 问题。后者即算法 6，概述如下。正如 Chen 等人 [的动机所示， 2014], 该算法

算法 6：估计 $\boldsymbol{\beta}$ 的 [Chen 等人， 2014]数据：$(\boldsymbol{X},\boldsymbol{y})\in\mathbb{R}^{n\times p}\times\mathbb{R}^n$令 $(\hat{\boldsymbol{K}},\hat{\boldsymbol{g}})$：$=\arg\min_{\boldsymbol{K},\,\boldsymbol{g}}\sum_{i=1}^n\left(-\langle\boldsymbol{x}_i\boldsymbol{x}_i^\top,\boldsymbol{K}\rangle+2y_i\langle\boldsymbol{x}_i,\boldsymbol{g}\rangle-y_i^2+\sigma^2\right)^2+\lambda\|\boldsymbol{K}\|_*$计算矩阵 $\hat{\boldsymbol{J}}=\hat{\underline{\boldsymbol{g}}}\hat{\boldsymbol{g}}^\top-\hat{\boldsymbol{K}}$，及其第一个特征值-特征向量对 $\hat{\lambda}$ 和 $\hat{\boldsymbol{v}}$计算 $\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\beta}}_2=\hat{\boldsymbol{g}}\pm\sqrt{\hat{\lambda}}\hat{\boldsymbol{v}}$ 返回 $(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)$

$\hat{}$ 通过凸惩罚最小二乘优化来确定矩阵和向量 $(\boldsymbol{K},\hat{\boldsymbol{g}})$，其中 $\hat{\boldsymbol{g}}$ $\hat{}$ 是 $\boldsymbol{\beta}_1$ 的朴素估计，而 $\boldsymbol{J}$ 的主特征向量-特征值则是必要的校正。我们定义该值为 $\alpha:=\frac{\|\boldsymbol{\beta}_1-\boldsymbol{\beta}_2\|_2^2}{\|\boldsymbol{\beta}_1\|_2^2+\|\boldsymbol{\beta}_2\|_2^2}$。

定义 D.18。 [Chen 等人， 2014] 令 $n_1=\{i\in[n]:\ z_i=1\}$ 表示从 $\boldsymbol{\beta}_1$中获得的样本数，且 n_2 表示对应的 $\boldsymbol{\beta}_2$。我们定义以下正则性条件，以供后续证明使用：

1. $\boldsymbol{X}$ 是一个独立同分布的标准高斯矩阵，
2. $\alpha\ge c_3$，
3. $\min\{n_1,n_2\}\ge c_4 p$，
4. $\lambda=\Theta(\sigma(\|\boldsymbol{\beta}_1\|_2+\|\boldsymbol{\beta}_2\|_2+\sigma)\sqrt{np\log^3 n})$，

5. $n \geq c_3 p \log^8 n$,

6. $|n_1 - n_2| = O\left(\sqrt{n \log n}\right)$,

对于某些常数 $0 < c_3 < 2$ 和 c_4。

定理 D.19。 Chen 等人 *[2014]* 假设定义 *D.18* 中的条件成立。存在ˆ ˆˆ常数 c_1、 c_2、$c_4 > 0$ 使得以至少 $1 - c_1 n^{-c_2}$的概率，算法 $\theta = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ 的算法 3 满足

$$\rho(\hat{\theta}, \theta) \leq c_4 \sigma \sqrt{\frac{p}{n}} \log^4 n + c_4 \min \left\{ \frac{\sigma^2}{\|\boldsymbol{\beta}_1\|_2 + \|\boldsymbol{\beta}_2\|_2}, \sigma \left(\frac{p}{n}\right)^{1/4} \right\} \log^4 n.$$

定理 D.19 的结果表明，在高信噪比区域 $\|\boldsymbol{\beta}\|_2^2/\sigma^2 = \Omega(1)$ 中，有$\rho(\hat{\theta}, \theta) = \Theta(\sigma\sqrt{\frac{p}{n}})$。这是因为 $\|\boldsymbol{\beta}_1\|_2 + \|\boldsymbol{\beta}_2\|_2 = 2\sqrt{k}$。

定理 D.20 (CORR 的 CORR + 算法 6 在 MSLR上的成功)。 考虑具有参数 MSLR的一般情形 p、 n、 k、$\phi = 1/2$、$(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \sim \mathcal{P}_{\|\boldsymbol{\beta}\|_2}(\mathcal{D})$。假设定义 *D.18*中的条件成立。存在常数 c_1、 c_2、 $c_4 > 0$ 使得，只要

$$n \geq \frac{32}{\min\{\phi^2 \boldsymbol{\beta}_{min}^2, (1-\phi)^2 \boldsymbol{\beta}_{min}^2, \langle \boldsymbol{\beta} \rangle_{min}^2\}} (1+\epsilon)(\|\boldsymbol{\beta}\|_2^2 + \sigma^2) \log 2p$$

对于 $\epsilon \in (0, 1)$ 中使用的 CORR，以至少 $1 - c_1(\frac{k}{p} + ke^{-c_2 n} + \frac{k}{n} + \frac{1}{p^{c_2}})$的概率，ˆ ˆˆ$\theta = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ 满足 CORR + 算法 6 的输出

$$\rho(\hat{\theta}, \theta) \leq c_4 \sigma \sqrt{\frac{2k}{n}} \log^4 n + c_4 \min \left\{ \frac{\sigma^2}{\|\boldsymbol{\beta}_1\|_2 + \|\boldsymbol{\beta}_2\|_2}, \sigma \left(\frac{2k}{n}\right)^{1/4} \right\} \log^4 n.$$

定理 *D.20* 的证明。 根据定理 D.1，我们可以以至少 $2k$ 的概率，通过运行 CORR，恢复信号 $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$) 的联合支撑集（大小至多为 $1 - c_1(\frac{k}{p} + ke^{-c_2 n} + \frac{k}{n} + \frac{1}{p^{c_2}})$ CORR。

运行 CORR 并确定 $< 2k$ 联合支撑集索引后，我们通过从设计矩阵 $\boldsymbol{X}$ 中删除所有其他列，将回归问题限制在这些索引上（因为这些列不会影响输出 y，因为它们不对应 $\boldsymbol{\beta}_1$ 或 $\boldsymbol{\beta}_2$的支撑索引）。接下来，我们需要解决一个双分量回归混合问题，其中包含 n 个样本，信号维度介于 k 和 $2k$ 之间。我们在这个简化后的回归问题上运行算法 3，以得到 ˆ ˆ$(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$。

剩下需要证明的是，定理 D.19 的假设在将 p 替换为 $2k$后仍然成立。事实上，条件 1。成立，因为不失一般性地，我们可以假设 $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$，否则问题设定将属于 SLR。条件 3、4、5 根据 MSLR 问题的定义，以及关于 λ和 c_3的自由性而成立。

条件 2。 以至少 $1 - \exp(-\Theta(n))$ 的概率成立；事实上，根据 Hoeffding 不等式[Boucheron 等人, 2013]我们有

$$\begin{aligned} \mathbb{P}\left[n_1 < 2c_4 k\right] &= \mathbb{P}\left[\sum_{i=1}^{n} z_i < 2c_4 k\right] \\ &= \mathbb{P}\left[\sum_{i=1}^{n} (z_i - \phi) < 2c_4 k - \phi n\right] \end{aligned}$$

$$= \mathbb{P}\left[\sum_{i=1}^{n}(-z_i+\phi) \geq \phi n - 2c_4 k\right]$$
$$\leq \exp\left(-\Theta(n)\right),$$

其中 $\phi n - 2c_4 k$ 对于某个 $c_4 > 0$ 为正，因为根据假设 $n \geq$ $^{32}\frac{(1+\epsilon)\min\{(2\phi-1)^2,\ \phi^2,\ (1-\phi)}{^2\}(1+\epsilon)(k+\sigma^2)\log p}$ 以及 $(k+\sigma^2)\log p = \omega(k)$，对 n_2 也可类似得出。

在 $\phi = 1/2$（因此 $2\phi - 1 = 0$）的情况下，根据 Hoeffding 不等式，条件 6. 以概率 $\exp(-\Theta(\log n))$ 成立，

$$\mathbb{P}\left[\left|\sum_{i=1}^{n} z_i - \sum_{i=1}^{n}(1-z_i)\right| \geq \Theta(\sqrt{n\log n})\right]$$
$$\leq \mathbb{P}\left[\sum_{i=1}^{n}(2z_i-1) \geq \Theta(\sqrt{n\log n})\right] + \mathbb{P}\left[\sum_{i=1}^{n}(-2z_i+1) \geq \Theta(\sqrt{n\log n})\right]$$
$$\leq 2\exp\left(-\frac{2\Theta(n\log n)}{2n}\right) = \exp\left(-\Theta(\log n)\right).$$

上述事件发生的概率至少为 $1-\exp(-\Theta(\log n))$，而 CORR成功的事件发生的概率为 $1 - c_1(\frac{k}{p} + ke^{-c_2 n} + \frac{k}{n} + \frac{1}{p^{c_2}})$。这两个事件同时发生的概率至少为 $1 - c_1(\frac{k}{p} + ke^{-c_2 n} + \frac{k}{n} + \frac{1}{p^{c_2}})$，其中常数 $c_1, c_2 > 0$经过了调整。应用定理 D.19，即可得到结论。

□

D.4 CORR 用于 SLR 中的带符号支撑恢复

定理 2.9 的证明。 证明过程与定理 2.5 的证明类似。令 $\texttt{CORR}(\boldsymbol{X}, \boldsymbol{y})$ 表示在输入 CORR 上运行所得的输出 $\boldsymbol{X}, \boldsymbol{y}$。考虑检验函数

$$g\left(\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{y}\end{bmatrix}\right) := \begin{cases}\mathtt{p} & \texttt{CORR}(\boldsymbol{X},\boldsymbol{y}) \neq \emptyset\\ \mathtt{q} & \texttt{CORR}(\boldsymbol{X},\boldsymbol{y}) = \emptyset\end{cases}.$$

令 $\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{y}\end{bmatrix} \sim \mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})$。回顾一下，CORR 输出以下集合

$$\texttt{CORR}(\boldsymbol{X},\boldsymbol{y}) = \left\{j \in [p] : \left|\frac{\langle \boldsymbol{X}_j, \boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \geq \sqrt{2(1+\epsilon/2)\log 2p}\right\}.$$

与定理2.10的证明中一样，我们注意到，对于 $\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{y}\end{bmatrix} \sim \mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})$，有 $\langle \boldsymbol{X}_q, \underbrace{\frac{\boldsymbol{y}}{\|\boldsymbol{y}\|_2}}_{\boldsymbol{y}\text{ 独立同分布}}\rangle \sim \mathcal{N}(0,\ 1)$。将引理D.9 应用于（27），可得:

$$\mathbb{P}\left[\max_{q\in[p]}\left|\frac{\langle \boldsymbol{X}_q,\boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \geq \sqrt{2(1+\epsilon/2)\log 2p}\right] \leq \mathbb{P}\left[\max_{q\in[p]}\left|\frac{\langle \boldsymbol{X}_q,\boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \geq \sqrt{2\log 2p} + \frac{\epsilon}{2\sqrt{8}}\sqrt{\log 2p}\right], \quad (27)$$
$$\leq p^{-\frac{1}{16}(\frac{\epsilon}{2})^2} = o(1),$$

因此我们有，在 $\mathbb{P}(\boldsymbol{X}) \otimes \mathbb{P}(\boldsymbol{y})$ 条件下，$g\left(\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{y}\end{bmatrix}\right) = \mathtt{q}$ 以概率 $1 - o(1)$ 成立。

反之，令 $\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{y}\end{bmatrix} \sim \mathbb{P}(\boldsymbol{X},\boldsymbol{y})$。接着应用定理 2.10，推得 $\texttt{CORR}(\boldsymbol{X},\boldsymbol{y}) = \text{supp}(\boldsymbol{\beta}) \neq \emptyset$ 以至少 $1-\left(\frac{k}{p}+2ke^{-c_2 n}+\frac{1}{p^{c_2}}\right)$ 的概率成立，其中 $c_2>0$为某个常数。因此，在 $\mathbb{P}(\boldsymbol{X},\boldsymbol{y})$ 条件下，$g\left(\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{y}\end{bmatrix}\right) = \mathsf{p}$ 以概率 $1-o(1)$ 成立。

□

接下来，我们考虑式（CORR算法（3）的修改版本，以估计带符号的支撑 $\boldsymbol{\beta}$。

*Pr*定理*2.10*的证明。 令 $\mathcal{S}$ 表示 $\boldsymbol{\beta}$的支撑集。定义误差事件

$$\mathcal{E} = \cup_{j\in\mathcal{S}}\left\{\left|\frac{\langle \boldsymbol{X}_j,\boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| < \sqrt{2(1+\epsilon/2)\log 2p}\right\} \cup \left\{\max_{q\in\mathcal{S}^{\complement}}\left|\frac{\langle \boldsymbol{X}_q,\boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \geq \sqrt{2(1+\epsilon/2)\log 2p}\right\}.$$

通过证明 $\mathbb{P}[\mathcal{E}] = o(1)$，即可得出该定理的结论。基于此，我们应用并集界

$$\begin{aligned}\mathbb{P}[\mathcal{E}] &\leq k\mathbb{P}\left[\left|\frac{\langle \boldsymbol{X}_{j^*},\boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| < \sqrt{2(1+\epsilon/2)\log 2p}\right] + \mathbb{P}\left[\max_{q\in\mathcal{S}^{\complement}}\left|\frac{\langle \boldsymbol{X}_q,\boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \geq \sqrt{2(1+\epsilon/2)\log 2p}\right]\\ &:= k\nu_1+\nu_2\end{aligned}$$

其中 $j^*\in\mathcal{S}$。我们首先关注 ν_2，注意到 $\frac{\langle \boldsymbol{X}_q,\boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2} \overset{\text{独立同分布}}{\sim} \mathcal{N}(0,1)$ 对于 $q\in\mathcal{S}^{\complement}$。应用引理 D.9，我们可推得

$$\begin{aligned}\nu_2 &= \mathbb{P}\left[\max_{q\in\mathcal{S}^{\complement}}\left|\frac{\langle \boldsymbol{X}_q,\boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \geq \sqrt{2(1+\epsilon/2)\log 2p}\right]\\ &\leq \mathbb{P}\left[\max_{q\in\mathcal{S}^{\complement}}\left|\frac{\langle \boldsymbol{X}_q,\boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \geq \sqrt{2\log 2p}+\frac{\epsilon}{2\sqrt{8}}\sqrt{\log 2p}\right]\\ &\leq (2p)^{-\frac{1}{16}\left(\frac{\epsilon}{2}\right)^2},\end{aligned}$$

由于 $\sqrt{2(1+\epsilon/2)} \geq \sqrt{2}+\frac{\epsilon}{2\sqrt{8}}$。 令 $n \geq {}^{8}{}_{(1+\epsilon)\beta^2_{\min}}(\|\boldsymbol{\beta}\|_2^2+\sigma^2)\log 2p$，其中 $\epsilon\in(0,1)$，则对 ν_1 的界可由应用下述引理 D.21 得到:

$$\nu_1 \leq \frac{1}{2p}+2e^{-\frac{\delta^2 n}{8}}.$$

综合以上 ，我们得到 $\mathbb{P}[\mathcal{E}] \leq \frac{k}{2p}+2ke^{-c_2 n}+\frac{1}{p^{c_2}}$，其中 $c_2>0$ 为某个常数。 □

引理 D.21（SLR 的集中界）。考虑 $j^*\in\mathcal{S}$ 的 SLR 设置。那么，对于 $n \geq {}^{8(1+\epsilon)}{}_{\beta^2_{\min}}(\|\boldsymbol{\beta}\|_2^2+\sigma^2)$ log $2p$，我们有

$$\mathbb{P}\left[\left|\frac{\langle \boldsymbol{X}_{j^*},\boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \leq \sqrt{2(1+\epsilon/2)\log 2p}\right] \leq \frac{1}{p}+2e^{-\frac{\delta^2 n}{8}}$$

证明。令 $\nu := \mathbb{P}\left[\left|\frac{\langle \boldsymbol{X}_{j^*},\boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2}\right| \leq \sqrt{2(1+\epsilon/2)\log 2p}\right]$ 和 $\delta>0$稍后确定。我们首先对引理 D.9 中的事件 $\Omega_1 := \Omega_1(\delta)$ 进行条件化，并应用引理 D.5（针对 $z_i=1$ 的情形），将 $g_i \overset{\text{独立同分布}}{\sim} \mathcal{N}(0,1)$ 记为 $i\in[n]$，它们是相互独立的（也独立于 y, z）标准正态高斯变量，从而得到

$$
\begin{aligned}
\nu &= \int \mathbb{P}\left[\left|\frac{\boldsymbol{X}_{j^*}^T\xi}{\|\xi\|_2}\right| < \sqrt{2(1+\epsilon/2)\log 2p} \,\middle|\, \boldsymbol{y}\right] d\mathbb{P}[y] \\
&= \int \mathbb{P}\left[\left|\sum_{i=1}^n g_i \frac{y_i}{\|\boldsymbol{y}\|_2}\sqrt{1-\frac{\boldsymbol{\beta}_{j^*}^2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}} + \frac{\boldsymbol{\beta}_{j^*}\cdot y_i^2/\|\boldsymbol{y}\|_2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right| < \sqrt{2(1+\epsilon/2)\log 2p} \,\middle|\, \boldsymbol{y}\right] d\mathbb{P}[y] \\
&= \int \mathbb{P}\left[\left|g_1\sqrt{1-\frac{\boldsymbol{\beta}_{j^*}^2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}} + \frac{\boldsymbol{\beta}_{j^*}\cdot\|\boldsymbol{y}\|_2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right| < \sqrt{2(1+\epsilon/2)\log 2p} \,\middle|\, \boldsymbol{y}\right] d\mathbb{P}[y] \\
&\le \mathbb{P}\left[\left|g_1\sqrt{1-\frac{\boldsymbol{\beta}_{j^*}^2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}} + \frac{\boldsymbol{\beta}_{j^*}\cdot\|\boldsymbol{y}\|_2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right| < \sqrt{2(1+\epsilon/2)\log 2p} \,\middle|\, \Omega_1\right] + \mathbb{P}\left[\Omega_1^{\complement}\right] \\
&= \mathbb{P}\left[\left\{g_1\sqrt{1-\frac{\boldsymbol{\beta}_{j^*}^2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}} + \frac{\boldsymbol{\beta}_{j^*}\cdot\|\boldsymbol{y}\|_2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2} < \sqrt{2(1+\epsilon/2)\log 2p}\right\}\right. \\
&\qquad\qquad \left.\cap\left\{-\sqrt{2(1+\epsilon/2)\log 2p} \le g_1\sqrt{1-\frac{\boldsymbol{\beta}_{j^*}^2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}} + \frac{\boldsymbol{\beta}_{j^*}\cdot\|\boldsymbol{y}\|_2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right\} \,\middle|\, \Omega_1\right] + \mathbb{P}\left[\Omega_1^{\complement}\right] \\
&\le \mathbb{P}\left[g_1\sqrt{1-\frac{\boldsymbol{\beta}_{j^*}^2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}} < \sqrt{2(1+\epsilon/2)\log 2p} - \left|\frac{\boldsymbol{\beta}_{j^*}\cdot\|\boldsymbol{y}\|_2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right| \,\middle|\, \Omega_1\right] + \mathbb{P}\left[\Omega_1^{\complement}\right] \\
&\le \mathbb{P}\left[g_1\sqrt{1-\frac{\boldsymbol{\beta}_{j^*}^2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}} < \sqrt{2(1+\epsilon/2)\log 2p} - \left|\frac{\boldsymbol{\beta}_{j^*}\cdot\sqrt{n(\|\boldsymbol{\beta}\|_2^2+\sigma^2)(1-\delta)}}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right| \,\middle|\, \Omega_1\right] + \mathbb{P}\left[\Omega_1^{\complement}\right] \\
&= \mathbb{P}\left[g_1\sqrt{1-\frac{\boldsymbol{\beta}_{j^*}^2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}} < \sqrt{2(1+\epsilon/2)\log 2p} - \left|\boldsymbol{\beta}_{j^*}\sqrt{\frac{n(1-\delta)}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}}\right| \,\middle|\, \Omega_1\right] + \mathbb{P}\left[\Omega_1^{\complement}\right].
\end{aligned}
$$

现在令 $n \ge {}^{8}\!/_{(1+\epsilon)\boldsymbol{\beta}_{\min}^2}(\|\boldsymbol{\beta}\|_2^2+\sigma^2)\log 2p$ 并应用标准的次高斯界（参见Wainwright [2019]），得到

$$
\begin{aligned}
\nu &\le \mathbb{P}\left[g\sqrt{1-\frac{\boldsymbol{\beta}_{j^*}^2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}} < (\sqrt{2(1+\epsilon/2)} - \sqrt{8(1+\epsilon)(1-\delta)})\sqrt{\log 2p}\right] + \mathbb{P}\left[\Omega_1^{\complement}\right] \\
&\le \exp\left(-\frac{(\sqrt{2(1+\epsilon/2)} - \sqrt{8(1+\epsilon)(1-\delta)})^2\log 2p}{2\left(1-\frac{\boldsymbol{\beta}_{j^*}^2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right)}\right) + 2e^{-\frac{\delta^2 n}{8}},
\end{aligned}
$$

其中

$$
\sqrt{2(1+\epsilon/2)} - \sqrt{8(1+\epsilon)(1-\delta)} < -\sqrt{2} \tag{28}
$$

对于 $\delta > 0$ 足够小。因此，选择 δ 使其满足上面的（28），我们便可得到，当 k 足够大时，

$$
\nu \le \exp\left(-\frac{2\log 2p}{2\left(1-\frac{\boldsymbol{\beta}_{j^*}^2}{\|\boldsymbol{\beta}\|_2^2+\sigma^2}\right)}\right) + 2e^{-\frac{\delta^2 n}{8}}
$$

$$\leq \exp\left(-\frac{2\log 2p}{2}\right) + 2e^{-\frac{\delta^2 n}{8}}$$
$$\leq \frac{1}{2p} + 2e^{-\frac{\delta^2 n}{8}}.$$

□