arXiv:2601.00329v1 [cs.GT] 1 Jan 2026

稀疏概率联盟结构生成：贝叶斯贪婪追踪与 ℓ_1松弛

Angshul Majumdar[1], *
[1]IIIT Delhi，印度。*
angshul@iiitd.ac.in。

关键词： 联盟结构生成、概率多参与者系统、稀疏回归、贝叶斯贪婪追踪

我们研究联盟价值并非已知、而是必须从回合观测中学习时的联盟结构生成（CSG）。我们将每个回合建模为一个稀疏线性回归问题，其中实现收益 Y_t 是少量联盟贡献的带噪线性组合。这构成了一个概率联盟结构生成框架：规划者首先从 T 回合中估计稀疏价值函数，然后在推断出的联盟集合上运行 CSG 求解器。我们分析了两种估计方案。第一种是贝叶斯贪心联盟追踪（BGCP），这是一种模仿正交匹配追踪的贪心过程。在相干性条件和最小信号假设下，当 $T \gtrsim K \log m$时，BGCP 以高概率恢复真实的有利可图联盟集合，从而得到福利最优结构。第二种方案采用 ℓ_1惩罚估计量；在受限特征值条件下，我们推导出ℓ_1 和预测误差界，并将其转化为福利差距保证。我们将这两种方法与概率基线方法进行比较，并找出稀疏概率 CSG 占优的情形，以及经典最小二乘方法具有竞争力的稠密情形。

1 引言

联盟结构生成（CSG）是合作博弈论和多智能体系统中的一个核心优化问题：给定参与者集合 $N = \{1, \ldots, n\}$ 和特征函数 $v : 2^N \to \mathbb{R}$，目标是将 N 划分为互不相交的联盟，以便

最大化总社会福利 $\sum_{C\in\mathcal{P}} v\,(C)$ 在所有联盟结构 $\mathcal{P}$中。搜索空间的大小由贝尔数给出，其增长速度超指数级，随n增长；即使在最简单的可转移效用设置下，联盟结构生成也是 NP 困难，并且需要复杂的算法技术 [1, 2]。过去二十年来，大量研究致力于开发动态规划、分支定界和随时算法，以精确或近似地生成联盟结构，并着重关注最坏情况下的复杂性以及联盟结构搜索空间的剪枝 [3, 4, 5, 6, 7, 2]。近期发表于 Algorithmica 等相关刊物的研究进一步探讨了联盟结构的结构与复杂性问题 [8, 9], ，包括对划分的偏好和更精细的效率概念。

与此同时，多智能体系统和 JAAMAS 社群围绕不确定性下的联盟形成发展出了丰富的研究文献。在这类研究中，联盟价值并非先验已知，而必须从数据中学习或推断。一个重要的研究方向采用贝叶斯强化学习的视角：参与者反复形成联盟、观察噪声收益，并更新对潜在伙伴类型或能力的信念，由此产生了贝叶斯核心和序贯最优联盟形成策略等新概念 [10, 11]。这些模型明确承认，在许多应用中，联盟价值只有通过阶段性交互才能揭示，并为如何在重复联盟形成中平衡探索与利用提供了合理的方法。更广泛而言，合作型和多目标多智能体系统领域的综述与案例研究强调，联盟形成是分布式决策和资源分配的关键机制，其应用范围涵盖能源系统、通信网络和机器人学等领域 [12, 13]。

关于概率联盟结构生成（PCSG）的研究提供了一种互补的形式化框架，其中不确定性源于参与者的随机出席，而非噪声收益。在 [14], 的 PCSG 模型中，每个参与者在给定实例中以特定概率处于活跃状态，目标是在期望概率出席模型下最大化社会福利的联盟结构。后续研究设计了近似算法，并更详细地探讨了 PCSG 的复杂性，包括 JAAMAS 文献中针对 PCSG 及相关变体提出的两种近似方案 [15, 16]。这些贡献确立了 PCSG 与经典联盟结构生成并列的基础性组合优化问题地位，但其重点主要在于如何在概率出席模型本身的框架下，通过算法学计算出良好的联盟结构。

然而，在许多现实场景中，不确定性既非纯粹的组合性不确定性（随机出席），也非认知性不确定性，即联盟价值未知但固定。相反，系统设计者通常会观察到情节式数据，其形式为

$$(X_t, Y_t), \qquad t = 1, \ldots, T,$$

其中，$X_t \in \{0,1\}^m$ 编码了回合 t中哪些 m 候选联盟（或联盟特征）处于激活状态，而 $Y_t \in \mathbb{R}$ 是由底层特征函数生成的含噪标量奖励。示例包括在线广告和赞助搜索，每一轮都会展示不同的广告主子集；物流和运输，承运商子集会协作处理共享货物；以及机器人或传感团队，每个回合中实际实例化的只有少数联盟模板。在这些

领域中，人们很少能观察到所有 $C \subseteq N$ 的完整取值 $v(C)$；相反，人们观察到的是一系列含噪的聚合收益，因此必须同时学习底层的合作结构，并利用该结构来计算高福利联盟结构。这促使我们构建一种概率联盟结构生成框架，它明确以数据驱动为基础，并具有统计学特性。

从建模角度看，利用回合观测中的结构，一种自然的方法是假设只有少数联盟对福利的边际贡献不可忽略。形式上，若我们将一组 m 候选联盟编为索引 $j \in [m]$，并用 $\theta_j^\star$ 表示联盟 j的（未知）贡献，那么只有其中 $K \ll m$ 个联盟真正相关这一假设，就对应于一个稀疏参数向量 $\theta^\star \in \mathbb{R}^m$。于是，观测到的回合收益可以建模为

$$Y_t = X_t^\top \theta^\star + \varepsilon_t,$$

其中噪声 ε_t服从次高斯分布，这与压缩感知和统计学习中常见的高维稀疏线性模型完全一致。从这一视角看，概率联盟结构生成成为一个两层问题：（1）统计地，我们必须根据 T 个回合恢复 $\theta^\star$ 的支撑集和幅值，理想情况下所需的 T 仅随m对数增长；（2）组合地，随后必须在估计器所建议的缩减后联盟全集上运行一个（可能代价高昂的）CSG 求解器。

现有的 PCSG 和联盟形成文献只能为此类情节式高维问题提供有限的指导。经典联盟结构生成算法假设 v（C）要么是完全已知的，要么可以以单位成本进行查询；它们关注的是剪枝联盟结构空间，而不是从数据中学习 v [2, 3, 7, 5, 6,4, 17]。反过来，PCSG 算法在已知出席概率的情况下优化期望福利，并未处理回合级收益的含噪观测 [14, 15, 16]。用于联盟形成的贝叶斯强化学习模型明确考虑了不确定性和重复性，但通常处理相对较小的联盟空间，重点在于学习稳定联盟或均衡，而不是学习高维稀疏价值函数 [10, 11]。最后，Algorithmica 及相关期刊中关于享乐型博弈和投票博弈中的联盟结构的算法研究，强调稳定性、公平性和帕累托效率 [8, 9]，而不是从情节数据中以高样本效率最大化福利。

本文以一种不同且明确的概率 和高维 视角研究联盟结构生成。我们假设联盟层面的贡献由稀疏线性模型编码，且规划者在确定联盟结构之前会观测到有限数量的带噪回合。随后，我们研究两类概率 CSG 方案：(i) 贝叶斯贪心联盟追踪（BGCP）算法，该算法仿效稀疏回归中正交匹配追踪的贝叶斯版本；(ii) 一种受高维 Lasso 理论启发的 ℓ_1-惩罚回合式估计流程。对于这两类方案，我们都给出了非渐近福利保证，该保证在稀疏情形下是紧的；同时，我们明确比较了它们与自然的回合式插件方案及稠密最小二乘基线在统计和计算方面的权衡。通过这一研究，我们将高维统计与稀疏近似的思想同算法联盟结构生成文献联系起来，并将概率联盟结构生成定位为一种

独立的统计计算问题，它有别于基于出席的 PCSG 和经典合作博弈论表述，但与二者互为补充。

2 概率模型与稀疏表述

本节将明确联盟结构生成的概率性。我们首先回顾经典的（确定性）表述，然后介绍一个用于描述联盟价值和可用性的随机模型，最后说明如何由此得到一种稀疏（ℓ_0）优化视角，并将利用该视角设计我们的算法。

2.1 简述经典确定性联盟结构生成

设 $N = \{1, \dots, n\}$ 为有限参与者集合。联盟 是任意非空子集$S \subseteq N$，而联盟结构 是 N的一个划分 $\mathcal{P} = \{S_1, \dots, S_k\}$，由若干互不相交的联盟组成。在经典设定中，给定一个确定性特征函数

$$v : 2^N \setminus \{\emptyset\} \to \mathbb{R},$$

目标是找到一个联盟结构 $\mathcal{P}$，使总价值最大化

$$W(\mathcal{P}) = \sum_{S \in \mathcal{P}} v(S).$$

这一联盟结构生成（CSG）问题通常被认为是 NP 困难，并且与集合划分和集合打包密切相关；关于最坏情况保证、算法技术以及 CSG 文献综述，可参见 Sandholm 等人和 Rahwan 等人的研究。 [1, 2]在这种确定性模型中，可行划分和联盟价值 $v(S)$均为固定值，并假定在优化时已知。

2.2 概率视角：随机价值与随机参与

在许多应用中，联盟价值具有不确定性，或者只能通过带噪声的样本观测到。例如，在重复市场、任务分配场景和基于学习的环境中，联盟所产生的价值取决于随机的外部条件或未观测到的状态。为此，我们在概率空间 $(\Omega\ \mathcal{F},\ \mathbb{P})$ 上开展研究，并将联盟表现和可行性建模为随机对象。

随机联盟表现。
对于每个非空联盟 $S \subseteq N$，我们引入一个实值随机变量

$$V(S) : \Omega \to \mathbb{R},$$

表示联盟 S 在单个回合 中的实现收益（例如，一场拍卖、一个任务实例或一个时间段）。我们记

$$\mu(S) \coloneqq \mathbb{E}\big[V(S)\big]$$

表示联盟 S 的潜在期望值。经典 CSG 模型是这样一种特殊情况： $V(S) \equiv v(S)$ 几乎必然成立，因此 $\mu(S) = v(S)$，且不存在随机性。

随机可用性／参与。
在许多情境中，并非所有联盟在每个回合都可行（例如，由于资源约束、参与者退出或可行性取决于具体情境）。因此，我们引入二元随机变量

$$A(S) \in \{0, 1\}, \qquad S \subseteq N,\ S \neq \emptyset,$$

其中 $A(S) = 1$ 表示联盟 S 在当前回合是否可用或处于活跃状态。集合 $A = \big(A(S)\big)_S$ 是潜在联盟的随机子集；其分布对可行性或参与的不确定性进行建模。

观测数据。
我们假设观测到 T 个独立同分布轮次

$$\{(A_t, V_t)\}_{t=1}^{T},$$

其中 $A_t(S)$ 是可用性指示变量，而 $V_t(S$ 是联盟的实现收益 S，对应于回合 t。在实践中，我们可能每个回合只观测到部分联盟（例如，拍卖中实际提交的竞标）；将 $A_t(S) = 0$设为在回合 t 中不可行或未被观测到的联盟对应的取值，即可自然地对此进行建模（参见组合拍卖与获胜者确定，例如 [18, 19]）。

给定这一概率模型，概率联盟结构生成 是指利用观测到的回合 $\{(A_t,\ V_t)\}_{t=1}^{T}$，选择一个联盟结构 $\mathcal{P}$，使其期望福利最高

$$W(\mathcal{P}) \coloneqq \sum_{S \in \mathcal{P}} \mu(S)$$

在 $(A,\ V)$ 的联合分布下。与经典联盟结构生成的关键区别在于，期望值 $\mu(S)$ 并非直接给出：必须根据数据进行统计推断，而且潜在的有利可图的联盟空间可能极其庞大。

2.3 参数化稀疏估值模型

为了建立与稀疏优化的联系，我们引入一种联盟价值的参数化表示，该表示受到组合拍卖和稀疏线性模型 [18, 19, 20, 21] 的启发。

设 $\mathcal{C} = \{C_1, \ldots, C_m\}$ 为联盟空间中候选联盟（或原子）的索引集。例如， $\mathcal{C}$ 可以包含所有规模不超过某个固定值 $k_{\max}$ 的联盟，也可以是一个通过领域知识或预处理步骤精心构建的“有趣”联盟库。

我们假设，对于每个回合 t，我们都会观测到一个总福利

$$Y_t = \sum_{j=1}^{m} X_t(C_j)\,\theta_j + \varepsilon_t,$$

其中

- $X_t(C_j) \in \mathbb{R}$ 表示第t回合中联盟 C_j 的激活水平（可能是二元的）；在许多应用中，我们有 $X_t(C_j) \in \{0,1\}$，但也允许使用更一般的实值设计，并且这在分析中会很有用；
- $\theta_j \in \mathbb{R}$ 是一个未知参数，表示联盟C_j 对福利的潜在贡献；
- ε_t 是一个噪声项，用于捕捉未建模的随机性（例如，环境波动或近似误差）。

收集 T 个回合后，我们得到线性模型

$$Y = X\theta + \varepsilon,$$

其中 $Y \in \mathbb{R}^T$， $X \in \mathbb{R}^{T\times m}$（在我们的应用中通常为二元元素 $X_{tj} \in \{0,1\}$），且 $\theta \in \mathbb{R}^m$。对于设计矩阵 X的适当选择，这涵盖了经典的集合打包／获胜者确定表述（每一行都编码了参与者之间联盟的一种可行分配，如组合拍卖中所示 [18, 19]），也涵盖了随机或基于样本的情境，其中每一行对应一个观测到的回合或场景。

我们将 θ 视为决定联盟贡献的参数向量，其未知的“真实”取值为 $\theta^\star \in \mathbb{R}^m$，对应的支撑集为

$$S^\star := \operatorname{supp}(\theta^\star) = \{j \in \{1,\dots,m\} : \theta_j^\star \neq 0\}.$$

稀疏恢复问题是根据$(X,\ Y)$的联合分布推断 $S^\star$（以及相应的系数）。概率要素现在明确了： X、 Y和 ε 都是随机变量，而 $\theta^\star$ 是一个有待推断的未知参数。

2.4 有利可图联盟的稀疏先验

我们关注的情形是：真正能大规模带来收益的联盟只有少数。这既是一种建模假设（只有少数结构化合作模式具有重要意义），也是计算上的必要性（我们无力对数量呈指数级增长的活跃联盟进行推理）。因此，我们假设 $\theta^\star$ 是稀疏的：

- 支撑集 $\operatorname{supp}(\theta^\star) = S^\star$ 满足 $|S^\star| \le K \ll m$。

从贝叶斯视角来看，一种自然的编码方式是采用伯努利–高斯（尖峰与宽松先验）先验，如贝叶斯追踪算法 [22] 所示。对于每个 j,

$$Z_j \sim \text{Bernoulli}(p_j), \qquad \theta_j \mid Z_j = 1 \sim \mathcal{N}(0,\tau^2), \qquad \theta_j \mid Z_j = 0 \equiv 0,$$

在各个 j 之间相互独立。这里，潜在指示变量 Z_j 用于选择哪些联盟在全局范围内处于“激活”状态，而高斯宽松先验则控制其贡献的大小。

在此先验下，可以证明，$(Z,\ \theta)$ 的最大后验估计在小噪声极限下可化为如下形式的 ℓ_0 正则化最小二乘问题

$$\min_{\theta\in\mathbb{R}^m}\ \frac{1}{2}\,\|Y - X\theta\|_2^2 + \lambda\|\theta\|_0,$$

在适当选择 $\lambda > 0$ 的情况下，该选择由伯努利参数 $\{p_j\}$ 和噪声方差决定。这正是压缩感知和稀疏近似文献中研究的标准稀疏回归目标函数（例如 [20, 21]）。

因此，在最大后验概率层面，概率联盟结构生成可归结为：在能产生 ℓ_0 惩罚优化问题的贝叶斯模型下，从含噪样本中识别出一组稀疏的全局有利可图联盟。

2.5 从稀疏参数到联盟结构

$\hat{}$ 一旦推断出估计值 θ，相应的候选联盟集合为

$$\widehat{\mathcal{C}} \coloneqq \{C_j \in \mathcal{C} : \hat{\theta}_j \neq 0\}.$$

$\widehat{}$在这个缩减后的集合 $\mathcal{C}$ 上，我们现在可以求解一个组合联盟结构生成问题

$$\max_{\mathcal{P}}\ \sum_{S\in\mathcal{P}} \hat{\mu}(S) \quad \text{subject to} \quad \mathcal{P} \text{ is a partition of } N,\ S \in \widehat{\mathcal{C}} \text{ for all } S \in \mathcal{P},$$

$\hat{}$其中，$\hat{}\mu(S)$ 是对 $\mu(S)$ 的估计，通常在 $S = C_j$ 时取为 θ_j。这个组合问题可以通过标准动态规划、分支定界或联盟结构生成专用算法来解决，不过此时是在稀疏化联盟空间 [1, 2] 上进行。

从这个意义上说，概率联盟结构生成包含两个相互耦合的组成部分：

$\widehat{}$1. 统计阶段 （稀疏学习）：在回合的概率模型下，通过求解稀疏估计问题，推断出有前景的联盟集合 $\mathcal{C}$；2. 组合阶段 （结构生成）：仅使用 $\mathcal{C}$ 中的联盟，计算一个高福利的$\widehat{}$联盟结构。

在后续章节中，我们将为统计阶段提供贝叶斯贪婪（OMP 风格）程序和 ℓ_1-正则化凸松弛，并在对随机设计矩阵 X 的假设下给出概率恢复保证。这些假设的精神与压缩感知中的受限等距性或不相干条件类似（例如 [20, 21, 22]）。这些结果给出了如下形式的高概率结论：对于采样回合，只要概率至少为 $1-\delta$，我们的算法就能在给定计算预算内恢复真正的全局有利可图联盟集合（或一个 ε-最优的超集）。

3 贝叶斯贪婪追踪与高概率保证

本节将第 2 节的概率稀疏模型具体化为一种用于估计稀疏参数向量的贪婪算法 $\theta^\star$，进而估计全局有利可图联盟集合。该算法是一种正交匹配追踪（OMP）型程序 [23, 24, 25, 26], ，我们将其解释为第 2 节伯努利–高斯先验下最大后验估计量的贪婪近似（参见 [22]）。随后，我们将证明一个高概率支集恢复结果，该结果针对我们的概率联盟结构生成设定进行了精细定制。

3.1 数据模型、支撑集与相干性

我们沿用第2节引入的线性模型。对于 T 个回合，我们观测到

$$Y = X\theta^\star + \varepsilon, \qquad Y \in \mathbb{R}^T,\ X \in \mathbb{R}^{T\times m},\ \theta^\star \in \mathbb{R}^m, \tag{1}$$

其中， X_j 的每一列 X 对应于联盟 $C_j \in \mathcal{C}$在各回合中的激活模式（第2节），而 $\varepsilon \in \mathbb{R}^T$ 表示噪声。

参数的真实支撑集为

$$S^\star \coloneqq \operatorname{supp}(\theta^\star) = \{j \in \{1,\dots,m\} : \theta^\star_j \neq 0\}, \qquad |S^\star| = K, \tag{2}$$

因此， K 是全局有利可图联盟的数量。我们采用列归一化

$$\|X_j\|_2^2 = T \quad \text{for all } j = 1,\dots,m, \tag{3}$$

并将 X 的互相干性定义为

$$\mu(X) \coloneqq \max_{\substack{i,j\in\{1,\dots,m\}\\ i\neq j}} \frac{|X_i^\top X_j|}{\|X_i\|_2\,\|X_j\|_2}. \tag{4}$$

直观而言， $\mu(X)$ 衡量不同联盟的激活模式之间最坏情况下的相关性。

3.2 最大后验目标与贝叶斯贪心联盟追踪

在第2节的伯努利–高斯先验下，噪声方差为 σ^2时，负对数后验（除去与 θ无关的常数）可写为

$$\mathcal{L}(\theta) \propto \frac{1}{2\sigma^2}\|Y - X\theta\|_2^2 + \lambda\|\theta\|_0, \qquad \lambda > 0, \tag{5}$$

其中 $\|\theta\|_0 = |\mathrm{supp}(\theta)|$。因此，最大后验估计量求解如下形式的 ℓ_0 惩罚最小二乘问题

$$\min_{\theta\in\mathbb{R}^m}\left\{\frac{1}{2}\|Y - X\theta\|_2^2 + \lambda_0\|\theta\|_0\right\}, \tag{6}$$

对于合适的 $\lambda_0 > 0$ 先验所确定的 [22]。直接求解（6）是组合优化问题，在大规模情况下不可行。

因此，我们考虑一种贪心近似：每一步添加一个坐标j，其列向量 X_j 与当前残差的相关性最高，然后在选定支撑集上重新计算最小二乘拟合，如经典的正交匹配追踪[23, 24, 25]所示。

定义 3.1（贝叶斯贪心联盟追踪（BGCP））。给定 X、Y、稀疏度预算 $K_{\max}$ 和残差阈值 $\eta \geq 0$，BGCP 算法按如下步骤进行。

1. 初始化：设定 $k \leftarrow 0$、$S^0 \leftarrow \emptyset$、$r^0 \leftarrow Y$。2. 迭代：当 $k < K_{\max}$ 且 $\|r^k\|_2 > \eta$时：（a）（选择）

对于每个 $j \in \{1,\dots,m\}$，定义

$$c_j^k \;:=\; \frac{1}{T}\,X_j^\top r^k.$$

选择

$$j_k \in \arg\max_{j\in\{1,\dots,m\}\setminus S^k} |c_j^k|.$$

更新支撑集

$$S^{k+1} \;\leftarrow\; S^k \cup \{j_k\}.$$

(b)（投影）构造 $X_{S^{k+1}}$，即 X 中由 S^{k+1} 索引的列所组成的子矩阵，并计算最小二乘估计

$$\hat\theta^{k+1}_{S^{k+1}} \;:=\; \arg\min_{u\in\mathbb{R}^{|S^{k+1}|}} \|Y - X_{S^{k+1}}u\|_2^2, \qquad \hat\theta^{k+1}_j = 0 \;\text{ for } j \notin S^{k+1}. \tag{7}$$

更新残差

$$r^{k+1} \;\leftarrow\; Y - X\hat\theta^{k+1}.$$

(c)（递增）设置 $k \leftarrow k+1$。

3. 输出：最终支撑集 $S^{\mathrm{BGCP}} \leftarrow S^k$ 和系数估计 $\hat\theta^{\mathrm{BGCP}} \leftarrow \hat\theta^k$。

我们用 $\widehat{\mathcal{C}}^{\mathrm{BGCP}} := \{C_j \in \mathcal{C} : j \in S^{\mathrm{BGCP}}\}$ 表示 BGCP 选出的联盟集合。

当 $K_{\max} = K$ 且 $\eta = 0$时，BGCP 恰好执行 K 次迭代，并生成一个K-稀疏估计。从最大后验概率的角度来看，它可以被解释为针对（6）低成本解的贪心坐标搜索，类似于稀疏信号恢复中的贝叶斯追踪算法 [22]。

3.3 假设与噪声集中

现在，我们明确 BGCP 以高概率恢复真实支撑集 $S^\star$ 的条件。这些条件是 OMP 分析中的标准条件 [24, 25, 26], , 并已适配为适用于我们的记号。

假设 3.2 (设计矩阵、噪声与最小信号)。 考虑支撑集 $S^\star$ 大小为 K 的模型 (1) 。假设：

(A1) (列归一化与相干性) X 的列满足 (3)，且互相干性 (4) 满足

$$\mu(X) \;<\; \frac{1}{2K-1}. \tag{8}$$

(A2) (次高斯噪声) ε 的 ε_t 个分量相互独立，均值为零，且为方差代理为 σ^2 的次高斯随机变量；也就是说，存在常数 $c_1 > 0$ ，使得对于所有 $t > 0$ 和所有 $j \in \{1, \ldots, m\}$,

$$\mathbb{P}\left(\frac{1}{T}\,|X_j^\top \varepsilon| > t\right) \;\le\; 2\exp\left(-c_1 \frac{Tt^2}{\sigma^2}\right). \tag{9}$$

(A3) (最小信号强度) $\theta^\star$ 的非零分量满足

$$\min_{j \in S^\star} |\theta_j^\star| \;\ge\; \theta_{\min} \;\coloneqq\; C_0 \sigma \sqrt{\frac{\log m}{T}}, \tag{10}$$

其中常数 $C_0 > 0$ 将在后文给出。

假设 3.2将设计矩阵上的几何条件（$\mu(X)$ 较小）与概率噪声模型和信噪比要求结合起来。相干性界（8）是 OMP 在无噪声情形下的经典精确恢复条件 [24, 25]，而（9）–（10）确保噪声不会掩盖由 $\theta^\star$ 诱导的相关性。

我们首先对噪声诱导相关性的最大值进行界定。

引理 3.3 (噪声相关性的统一界)。 在假设 *3.2(A2)* 下，存在常数 C_1， $C_2 > 0$ (仅取决于 c_1)，使得

$$\mathbb{P}\left(\max_{1 \le j \le m} \frac{1}{T}\,|X_j^\top \varepsilon| \;>\; C_1 \sigma \sqrt{\frac{\log m}{T}}\right) \;\le\; m^{-C_2}. \tag{11}$$

*Proof*固定 j并取 9 式，其中 $t = u\sigma\sqrt{(\log m)/T;}$ 我们得到

$$\mathbb{P}\left(\frac{1}{T}\left|X_j^\top \varepsilon\right| > u\sigma\sqrt{\frac{\log m}{T}}\right) \;\le\; 2\exp\left(-c_1 u^2 \log m\right) \;=\; 2m^{-c_1 u^2}.$$

由对 $j = 1,\dots,m$ 的联合界，

$$\mathbb{P}\left(\max_{1\le j\le m}\frac{1}{T}\left|X_j^\top \varepsilon\right| > u\sigma\sqrt{\frac{\log m}{T}}\right) \;\le\; 2m^{1-c_1 u^2}.$$

选取 u 足够大，使得 $c_1 u^2 \ge 2$，我们得到

$$\mathbb{P}\left(\max_{j}\frac{1}{T}\left|X_j^\top \varepsilon\right| > C_1\sigma\sqrt{\frac{\log m}{T}}\right) \;\le\; 2m^{-1} \;\le\; m^{-C_2},$$

对于适当的常数 $C_1, C_2 > 0$。 □

我们用

$$\mathcal{E}_{\text{noise}} \;\coloneqq\; \left\{\max_{1\le j\le m}\frac{1}{T}\left|X_j^\top \varepsilon\right| \le C_1\sigma\sqrt{\frac{\log m}{T}}\right\} \tag{12}$$

表示噪声相关性一致较小的高概率事件。

3.4 BGCP轨迹上的相关性界

现在我们分析 BGCP选择规则中使用的相关性。记 S^k 和 r^k分别为迭代 k开始时的支撑集和残差。

下面的引理利用最小二乘更新的正交性（ r^k ，将其表示为剩余真实列与噪声的线性组合。7）。

引理 3.4 (残差分解)。 假设 $S^k \subseteq S^\star$。则存在一个向量 $\alpha^k \in \mathbb{R}^{|S^\star \setminus S^k|}$ ，使得

$$r^k \;=\; X_{S^\star \setminus S^k}\alpha^k + P_{(X_{S^k})^\perp}\varepsilon, \tag{13}$$

其中 $P_{(X_{S^k})^\perp}$ 表示投影到 X_{S^k}的列空间的正交补上的正交投影算子。此外，

$$\|\alpha^k\|_\infty \;\ge\; \min_{j\in S^\star}|\theta_j^\star|. \tag{14}$$

Proof 根据（θ^k 中的 $\hat{}$ 的定义7），

$$\hat{\theta}^k_{S^k} \;=\; (X_{S^k}^\top X_{S^k})^{-1}X_{S^k}^\top Y, \qquad \hat{\theta}^k_j = 0 \text{ for } j \notin S^k.$$

利用模型（1），

$$Y \;=\; X_{S^\star}\theta^\star_{S^\star} + \varepsilon.$$

将其代入最小二乘解，并记作 $S^\star = S^k \cup (S^\star \setminus S^k)$，可得

$$\hat{\theta}^k_{S^k} \;=\; \theta^\star_{S^k} + (X_{S^k}^\top X_{S^k})^{-1}X_{S^k}^\top X_{S^\star\setminus S^k}\theta^\star_{S^\star\setminus S^k} + (X_{S^k}^\top X_{S^k})^{-1}X_{S^k}^\top \varepsilon.$$

因此，残差可写为

$$r^k \;=\; Y - X\hat{\theta}^k \;=\; X_{S^\star}\theta^\star_{S^\star} + \varepsilon - X_{S^k}\hat{\theta}^k_{S^k}.$$

重新整理可得

$$r^k \;=\; \underbrace{\big(I - P_{X_{S^k}}\big)X_{S^\star\setminus S^k}\theta^\star_{S^\star\setminus S^k}}_{\text{signal part}} + \underbrace{\big(I - P_{X_{S^k}}\big)\varepsilon}_{\text{noise part}},$$

其中 $P_{X_{S^k}}$ 是投影到 X_{S^k}的列空间上的正交投影算子。由于 $(I - P_{X_{S^k}})X_{S^\star\setminus S^k}$ 与 $X_{S^\star\setminus S^k}$具有相同的列空间，因此存在向量 α^k 使得 $(I-P_{X_{S^k}})X_{S^\star\setminus S^k}\theta^\star_{S^\star\setminus S^k} = X_{S^\star\setminus S^k}\alpha^k$，由此得到 (13)。此外，投影只会减小列之间的相关性，因此 α^k 的各项仍至少为 $\min_{j\in S^\star}|\theta^\star_j|$；特别地，我们可以取 $\|\alpha^k\|_\infty \geq \min_{j\in S^\star}|\theta^\star_j|$。（关于相干性情形的详细论证，见 [25, 26]。）□

下一个引理量化了与真实原子和虚假原子的相关性之间的间隔，条件是残差具有 (13) 的形式。

引理 3.5 (真实与虚假相关性界)。 假设 *3.2* 成立，并令 $\mathcal{E}_{\text{noise}}$ 为 (12)中定义的噪声事件。固定 $k \in \{0,\dots,K-1\}$ 并假设$S^k \subseteq S^\star$。在 $\mathcal{E}_{\text{noise}}$ 与引理 *3.4*中的事件的交集上，相关性$c^k_j = T^{-1}X_j^\top r^k$ 满足

$$\min_{j\in S^\star\setminus S^k}|c^k_j| \;\geq\; \big(1-(2K-1)\mu(X)\big)\,\theta_{\min} - C_1\sigma\sqrt{\frac{\log m}{T}}, \tag{15}$$

$$\max_{j\notin S^\star}|c^k_j| \;\leq\; (2K-1)\mu(X)\,\theta_{\min} + C_1\sigma\sqrt{\frac{\log m}{T}}. \tag{16}$$

Proof 令 $j \in S^\star \setminus S^k$。利用分解式（13）和归一化条件（3），

$$c^k_j \;=\; \frac{1}{T}\,X_j^\top X_{S^\star\setminus S^k}\alpha^k + \frac{1}{T}\,X_j^\top P_{(X_{S^k})^\perp}\varepsilon.$$

第一项可展开为

$$\frac{1}{T}\,X_j^\top X_{S^\star\setminus S^k}\alpha^k \;=\; \alpha^k_j + \sum_{i\in S^\star\setminus(S^k\cup\{j\})}\frac{X_j^\top X_i}{T}\,\alpha^k_i.$$

根据相干性界（4）和稀疏性 $|S^\star| = K$，

$$\left|\sum_{i\in S^\star\setminus(S^k\cup\{j\})}\frac{X_j^\top X_i}{T}\,\alpha^k_i\right| \;\leq\; (K-1)\mu(X)\,\|\alpha^k\|_\infty.$$

将其与式（14）以及 $\theta_{\min}$ 在式（10）中的定义相结合，得到

$$\left|\frac{1}{T}\,X_j^\top X_{S^\star\setminus S^k}\alpha^k\right| \;\geq\; \theta_{\min} - (K-1)\mu(X)\,\theta_{\min}.$$

类似地，对于任意 $j \notin S^\star$，

$$\left|\frac{1}{T}\,X_j^\top X_{S^\star\setminus S^k}\alpha^k\right| \;\leq\; K\mu(X)\,\theta_{\min}.$$

噪声项满足，对所有 j，

$$\left|\frac{1}{T}\, X_j^\top P_{(X_{S^k})^\perp}\varepsilon\right| \;\le\; \frac{1}{T}\,|X_j^\top \varepsilon| \;\le\; C_1\sigma\sqrt{\frac{\log m}{T}}$$

上 $\mathcal{E}_{\text{noise}}$，因为正交投影不会增大内积的绝对值。结合这些不等式，并注意到 $(K-1)+K \le 2K-1$，可得 (15) 和 (16)。 □

接下来的推论表明，对于 (C_0 10) 中常数 的适当取值，真实相关性与虚假相关性之间存在严格分离。

推论 3.6 (相关性分离)。 假设 *3.2* 成立，且$\mathcal{E}_{\text{noise}}$ 发生。如果最小信号 $\theta_{\min}$ 满足

$$\theta_{\min} \;\ge\; \frac{2C_1\sigma}{1-2(2K-1)\mu(X)}\sqrt{\frac{\log m}{T}},$$

那么对于每个满足 $S^k \subseteq S^\star$ 的 $k \in \{0,\dots,K-1\}$，

$$\min_{j\in S^\star\setminus S^k} |c_j^k| \;>\; \max_{j\notin S^\star} |c_j^k|. \tag{17}$$

Proof 根据引理 3.5，对于任意这样的 k，

$$\min_{j\in S^\star\setminus S^k} |c_j^k| - \max_{j\notin S^\star} |c_j^k| \;\ge\; \big(1-2(2K-1)\mu(X)\big)\,\theta_{\min} - 2C_1\sigma\sqrt{\tfrac{\log m}{T}}.$$

若 $\theta_{\min}$ 满足上述下界，则右侧严格为正，这意味着 (17)。 □

3.5 BGCP 的精确支集恢复

现在我们可以证明，BGCP 运行恰好 K 次迭代时，以高概率恢复真实支撑集 $S^\star$ 。

定理 3.7 (BGCP 的高概率支集恢复)。 设假设 *3.2*成立，并假设 *BGCP*（定义*3.1*）使用 $K_{\max} = K$ 和 $\eta = 0$，处理由 (1)生成的数据。则存在常数 c_2，$c_3 > 0$，它们仅依赖于假设 *3.2* 中的常数，满足

$$\mathbb{P}\big(S^{\text{BGCP}} = S^\star\big) \;\ge\; 1 - c_2 m^{-c_3}. \tag{18}$$

特别地，以至少 $1-c_2m^{-c_3}$ 的概率，*BGCP* 在第 k 次迭代中为每个 $k=0$ 选择一个索引 $j_k \in S^\star \setminus S^k$，……，$K-1$，因此经过 K 步后， $S^{\text{BGCP}} = S^\star$。

Proof 根据引理 3.3， $\mathbb{P}(\mathcal{E}_{\text{noise}}) \ge 1 - m^{-C_2}$。我们在事件 $\mathcal{E}_{\text{noise}}$ 上进行讨论，并对 k 作归纳，证明 $S^k \subseteq S^\star$，且 BGCP 在第 k 次迭代中选择。

对于 $k=0$，根据定义有 $S^0 = \emptyset \subseteq S^\star$。假设 $S^k \subseteq S^\star$ 对于某个 $k \le K-1$成立。根据引理 3.4，残差 r^k 可以写成形式（13），因此引理 3.5 适用。

在推论 3.6所述的附加信号强度条件下，通过在（10）中将 C_0 取得足够大即可满足该条件，我们得到真实索引与虚假索引的相关性之间的分离（17）。由于 BGCP 选择使 j_k 在 $j \notin S^k$上取最大值的索引 $|c_j^k|$，因此可得 $j_k \in S^\star \setminus S^k$。于是 $S^{k+1} = S^k \cup \{j_k\} \subseteq S^\star$，归纳法得证。

根据归纳法，对于所有 $k = 0, \ldots\ldots, K-1$， $S^k \subseteq S^\star$ 且 $j_k \in S^\star \setminus S^k$。经过 K 次迭代后，已从 $S^\star$ 中恰好选出 K 个不同索引，因此 $S^K = S^\star$。此外， $S^\star$ 之外的索引永远不会被选中。因此，在 $\mathcal{E}_{\text{noise}}$ 上，BGCP 可以恢复精确支持集。

最后，将此结果与概率界 $\mathbb{P}(\mathcal{E}_{\text{noise}}^c) \le m^{-C_2}$ 结合，并将常数吸收到 c_2中， c_3 即可得到尾界（18）。□

注 3.8 （样本复杂度与概率设计）。 定理3.7 是一个条件性结果：它假设固定的设计矩阵 X 满足相干性界（8）和列归一化（3。在第 2 节的概率联盟结构生成场景中，矩阵 X 本身由随机回合产生。对于许多受压缩感知启发的结构化随机设计，[20, 21, 25], 可以证明，只要 $\mu(X)$ 满足（8）的概率就很高，只要 $T \gtrsim K \log m$。将此类设计侧结果与定理3.7 结合，可得如下完全概率性陈述：对于随机回合和噪声，以高概率，BGCP 能够从 $S^\star$ 个 $T = O(K \log m)$ 样本中恢复真实的有利可图联盟集合。

3.6 对概率联盟结构生成的启示

一旦 $S^\star$ 被恢复，就可以将全局有利可图联盟的集合识别为 $\{C_j \in \mathcal{C} : j \in S^\star\}$。随后，我们可以进入联盟结构生成的组合阶段（第2节），并使用任意精确算法或近似算法 [1, 2]，在这个稀疏化联盟集上求解确定性 CSG 问题。

因此，定理3.7 在概率稀疏学习公式与经典 CSG 设置之间搭建了一座严谨的桥梁：在假设3.2成立的条件下，对于采样回合，以高概率，BGCP 能够准确识别有利可图的联盟集合，而福利最大化联盟结构必须以此集合为基础构建。在第4 节中，我们将看到，另一种互补的 ℓ_1 松弛也具有类似的高概率保证，但其偏差–方差权衡有所不同。

4 贝叶斯 ℓ_1 松弛与凸分析

第3 节分析了与第2节伯努利–高斯先验相对应的最大后验估计量的一种贪心（OMP 型）近似。现在，我们考虑一种互补的、完全凸的松弛：通过将（ ℓ_1惩罚项（Lasso）替换为其在单位立方体上的凸包络，得到的 ℓ_0 惩罚估计量，见（6）。我们始终采用与第3节相同的数据模型和记号，并在受限特征值条件下推导非渐近误差和支撑集控制保证。

4.1 从 ℓ_0 最大后验概率到 ℓ_1惩罚估计量

回顾第3节中的线性模型:

$$Y = X\theta^\star + \varepsilon, \qquad Y \in \mathbb{R}^T,\ X \in \mathbb{R}^{T\times m},\ \theta^\star \in \mathbb{R}^m, \tag{19}$$

具有真实支撑集 $S^\star = \operatorname{supp}(\theta^\star)$，其大小为 $|S^\star| = K$，如式（2）所示。我们保留列归一化（3）和次高斯噪声假设（9）。

根据第 2 节的伯努利–高斯先验，最大后验估计量求解以下ℓ_0惩罚最小二乘问题

$$\hat\theta^{(0)} \in \arg\min_{\theta\in\mathbb{R}^m} \left\{ \frac{1}{2}\|Y - X\theta\|_2^2 + \lambda_0\|\theta\|_0 \right\}, \tag{20}$$

对于某个 $\lambda_0 > 0$；参见式（6）。在高维情形下，直接求解式（20）是不可行的。

一种经典的凸替代是将 $\|\theta\|_0$ 替换为 ℓ_1范数 $\|\theta\|_1 = \sum_{j=1}^m |\theta_j|$，从而得到 Lasso 估计量 [27]

$$\hat\theta^{\ell_1} \in \arg\min_{\theta\in\mathbb{R}^m} \left\{ \frac{1}{2T}\|Y - X\theta\|_2^2 + \lambda\|\theta\|_1 \right\}, \qquad \lambda > 0. \tag{21}$$

从贝叶斯角度看，式（21）是在 θ的各坐标上采用独立同分布的拉普拉斯先验时的最大后验估计量；但在这里，我们将其解释为：在对（X，Y）采用相同概率模型的条件下，对 ℓ_0最大后验概率问题（20）进行的凸松弛。

为便于后续使用，我们定义估计误差

$$\Delta^{\ell_1} := \hat\theta^{\ell_1} - \theta^\star. \tag{22}$$

我们的目标是界定 $\|\Delta^{\ell_1}\|_2$ 和 $\|\Delta^{\ell_1}\|_1$，并量化 $\operatorname{supp}(\hat\theta^{\ell_1}) \setminus S^\star$ 中伪索引的数量。

4.2 受限特征值条件与正则化水平

我们回顾设计矩阵 X上的一个标准受限特征值（RE）条件。该条件弱于全局特征值条件，非常适用于稀疏高维回归 [28, 29]。

定义 4.1（受限特征值常数）。设 $S \subseteq \{1,\dots,m\}$ 和 $\alpha \geq 1$。设计矩阵 X 的受限特征值常数 $\kappa(S,\alpha)$ 定义为

$$\kappa(S,\alpha) := \inf_{\substack{u\in\mathbb{R}^m\setminus\{0\} \\ \|u_{S^c}\|_1 \leq \alpha\|u_S\|_1}} \frac{\|Xu\|_2}{\sqrt{T}\,\|u_S\|_2}, \tag{23}$$

其中，S^c是 $\{1,\dots,m\}$中 S的补集，u_S表示将 u限制在索引 S上的结果。

我们假设 X 在真实支撑集 $S^\star$ 上满足受限特征值条件，并且λ 的取值足够大，能够压制噪声诱导的相关性，思路与第3节类似。

假设 4.2 (受限特征值条件和正则化水平)。 在具有大小为 K的支撑集 $S^\star$ 的模型(19) 下，假设：

(B1) (受限特征值) 存在 $\kappa_\star > 0$ ，使得

$$\kappa(S^\star, 3) \;\geq\; \kappa_\star. \tag{24}$$

(B2) (噪声与调参) 噪声向量 ε 满足次高斯尾界(9)，且(21)中的正则化参数选为

$$\lambda \;=\; c_0 \sigma \sqrt{\frac{\log m}{T}}, \tag{25}$$

其中常数 $c_0 > 0$充分大。

假设 4.2(B1) 是标准的受限特征值条件，其中 $\alpha = 3$ 用于 Lasso 分析 [28, 29]。条件 (B2) 确保 λ通过匹配噪声事件 $\mathcal{E}_{\text{noise}}$ 在 (12) 中的尺度，支配最大的噪声诱导相关性。

现在我们将对此进行精确说明。

引理4.3 (噪声事件蕴含对偶可行性)。 设假设 *4.2(B2)* 成立，并设 $\mathcal{E}_{\text{noise}}$ 为 (12)中定义的事件。则当 c_0 足够大时（取决于引理 *3.3*中的常数），有

$$\mathcal{E}_{\text{noise}} \;\subseteq\; \left\{ \left\| \frac{1}{T} X^\top \varepsilon \right\|_\infty \;\leq\; \frac{\lambda}{2} \right\}. \tag{26}$$

特别地，存在常数 c_1、$c_2 > 0$ 使得

$$\mathbb{P}\left(\left\| \frac{1}{T} X^\top \varepsilon \right\|_\infty > \frac{\lambda}{2} \right) \;\leq\; c_1 m^{-c_2}.$$

Proof 根据引理 3.3，在 $\mathcal{E}_{\text{noise}}$ 上有

$$\left\| \frac{1}{T} X^\top \varepsilon \right\|_\infty \;\leq\; C_1 \sigma \sqrt{\frac{\log m}{T}}.$$

若 $c_0 \geq 2C_1$ 在（25）中成立，则 $C_1\sigma\sqrt{(\log m)/T} \leq \lambda/2$，这意味着（26）。概率界通过结合引理 3.3 与此包含关系并调整常数即可得到。□

4.3 基本不等式与锥约束

现在我们推导 Lasso 的基本不等式，并证明在噪声事件（26）发生时，估计误差 Δ^{ℓ_1} 位于真实支撑集 $S^\star$周围的一个 ℓ_1-锥内。

引理 4.4（基本不等式）。 设 $\hat{\theta}^{\ell_1}$ 是 *Lasso* 问题 （21）的任意一个解，并令 $\Delta^{\ell_1} = \hat{\theta}^{\ell_1} - \theta^\star$。那么

$$\frac{1}{2T}\|X\Delta^{\ell_1}\|_2^2 + \lambda\|\hat{\theta}^{\ell_1}\|_1 \;\le\; \frac{1}{T}\varepsilon^\top X\Delta^{\ell_1} + \lambda\|\theta^\star\|_1. \tag{27}$$

Proof 根据（θ^{ℓ_1}）中21的最优性,

$$\frac{1}{2T}\|Y - X\hat{\theta}^{\ell_1}\|_2^2 + \lambda\|\hat{\theta}^{\ell_1}\|_1 \;\le\; \frac{1}{2T}\|Y - X\theta^\star\|_2^2 + \lambda\|\theta^\star\|_1.$$

代入 $Y = X\theta^\star + \varepsilon$，得到

$$\frac{1}{2T}\|X\Delta^{\ell_1} - \varepsilon\|_2^2 + \lambda\|\hat{\theta}^{\ell_1}\|_1 \;\le\; \frac{1}{2T}\|\varepsilon\|_2^2 + \lambda\|\theta^\star\|_1.$$

展开左侧的平方范数，并从两边消去 $\|\varepsilon\|_2^2/(2T)$，得到

$$\frac{1}{2T}\|X\Delta^{\ell_1}\|_2^2 - \frac{1}{T}\varepsilon^\top X\Delta^{\ell_1} + \lambda\|\hat{\theta}^{\ell_1}\|_1 \;\le\; \lambda\|\theta^\star\|_1,$$

整理后得到（27）。 □

引理 4.5（误差的锥约束）。 假设条件 *4.2*（*B2*）成立，且事件 *Let*

$$E_{\ell_1} := \left\{\frac{1}{T}\|X^\top\varepsilon\|_\infty \le \frac{\lambda}{2}\right\}. \tag{28}$$

发生。则误差 Δ^{ℓ_1} 满足

$$\|\Delta^{\ell_1}_{(S^\star)^c}\|_1 \;\le\; 3\|\Delta^{\ell_1}_{S^\star}\|_1. \tag{29}$$

Proof 将 ℓ_1-范数分解为支撑集及其补集:

$$\|\hat{\theta}^{\ell_1}\|_1 = \|\theta^\star_{S^\star} + \Delta^{\ell_1}_{S^\star}\|_1 + \|\Delta^{\ell_1}_{(S^\star)^c}\|_1, \qquad \|\theta^\star\|_1 = \|\theta^\star_{S^\star}\|_1,$$

因为 $\theta^\star_{(S^\star)_c} = 0$。利用三角不等式,

$$\|\theta^\star_{S^\star} + \Delta^{\ell_1}_{S^\star}\|_1 \;\ge\; \|\theta^\star_{S^\star}\|_1 - \|\Delta^{\ell_1}_{S^\star}\|_1.$$

代入（27）并整理可得

$$\frac{1}{2T}\|X\Delta^{\ell_1}\|_2^2 + \lambda\|\Delta^{\ell_1}_{(S^\star)^c}\|_1 \;\le\; \frac{1}{T}\varepsilon^\top X\Delta^{\ell_1} + \lambda\|\Delta^{\ell_1}_{S^\star}\|_1.$$

在事件 $\mathcal{E}_{\ell_1}$ 上,

$$\frac{1}{T}\varepsilon^\top X\Delta^{\ell_1} \;\le\; \left\|\frac{1}{T}X^\top\varepsilon\right\|_\infty \|\Delta^{\ell_1}\|_1 \;\le\; \frac{\lambda}{2}\left(\|\Delta^{\ell_1}_{S^\star}\|_1 + \|\Delta^{\ell_1}_{(S^\star)^c}\|_1\right).$$

结合这些不等式可得

$$\frac{1}{2T}\|X\Delta^{\ell_1}\|_2^2 + \lambda\|\Delta^{\ell_1}_{(S^\star)^c}\|_1 \;\le\; \frac{\lambda}{2}\left(\|\Delta^{\ell_1}_{S^\star}\|_1 + \|\Delta^{\ell_1}_{(S^\star)^c}\|_1\right) + \lambda\|\Delta^{\ell_1}_{S^\star}\|_1.$$

从左侧去掉非负项 $\|X\Delta^{\ell_1}\|_2^2/(2T)$，并化简右侧，可得

$$\lambda\|\Delta^{\ell_1}_{(S^\star)^c}\|_1 \;\le\; \frac{\lambda}{2}\,\|\Delta^{\ell_1}_{(S^\star)^c}\|_1 + \frac{3\lambda}{2}\,\|\Delta^{\ell_1}_{S^\star}\|_1,$$

因此

$$\frac{\lambda}{2}\,\|\Delta^{\ell_1}_{(S^\star)^c}\|_1 \;\le\; \frac{3\lambda}{2}\,\|\Delta^{\ell_1}_{S^\star}\|_1.$$

两边除以 $\lambda/2$，得到（29）。 □

4.4 误差界与支撑集控制

我们现在将锥约束与受限特征值条件相结合，得到关于 ℓ_2误差和 ℓ_1误差的非渐近界，以及 Lasso 所选伪索引数量的界。

定理 4.6 (受限特征值条件下的Lasso误差界)。 假设 *4.2* 成立，且令$\hat{}\theta^{\ell_1}$ 为 *Lasso* 问题 (21)的任意一个解。那么，存在常数 $C_1, C_2, C_3 >0$，它们仅取决于引理 *3.3* 和假设 *4.2*中的常数，使得以至少 $1 - C_1 m^{-C_2}$ 的概率，以下界成立：

$$\|\hat{\theta}^{\ell_1} - \theta^\star\|_2 \;\le\; \frac{C_3\sigma}{\kappa_\star}\sqrt{\frac{K\log m}{T}}, \tag{30}$$

$$\|\hat{\theta}^{\ell_1} - \theta^\star\|_1 \;\le\; \frac{C_3\sigma K}{\kappa_\star^2}\sqrt{\frac{\log m}{T}}, \tag{31}$$

$$\frac{1}{T}\,\|X(\hat{\theta}^{\ell_1} - \theta^\star)\|_2^2 \;\le\; \frac{C_3^2\sigma^2 K}{\kappa_\star^2}\,\frac{\log m}{T}. \tag{32}$$

此外，假阳性的数量受到如下控制：

$$\left|\operatorname{supp}(\hat{\theta}^{\ell_1}) \setminus S^\star\right| \;\le\; \frac{C_3^2}{\kappa_\star^2}\,K, \tag{33}$$

在同一个高概率事件上。

Proof 根据引理 4.3，事件 $\mathcal{E}_{\ell_1}$ （在28）中定义的事件发生的概率至少为 $1 - C_1 m^{-C_2}$，其中常数 C_1、$C_2 > 0$ 取值适当。我们在 $\mathcal{E}_{\ell_1}$ 上进行分析，并应用引理 4.5，该引理推出锥约束 $\|\Delta^{\ell_1}{}_{(S^\star)_c}\|_1 \le 3\|\Delta^{\ell_1}_{S^\star}\|_1$。根据受限特征值常数的定义（23）以及假设 4.2(B1),

$$\frac{1}{\sqrt{T}}\,\|X\Delta^{\ell_1}\|_2 \;\ge\; \kappa_\star\|\Delta^{\ell_1}_{S^\star}\|_2. \tag{34}$$

回到基本不等式（27），并使用 $\mathcal{E}_{\ell_1}$，如引理4.5的证明中所示，我们得到

$$\frac{1}{2T}\,\|X\Delta^{\ell_1}\|_2^2 \;\le\; \frac{\lambda}{2}\left(\|\Delta^{\ell_1}_{S^\star}\|_1 + \|\Delta^{\ell_1}_{(S^\star)^c}\|_1\right) + \lambda\|\Delta^{\ell_1}_{S^\star}\|_1 - \lambda\|\hat{\theta}^{\ell_1}\|_1 + \lambda\|\theta^\star\|_1.$$

与引理4.5中一样，我们可以用 $\lambda\|\Delta^{\ell_1}_{S^\star}\|_1$来界定最后两项，从而得到

$$\frac{1}{2T}\|X\Delta^{\ell_1}\|_2^2 \;\le\; \frac{\lambda}{2}\left(\|\Delta^{\ell_1}_{S^\star}\|_1+\|\Delta^{\ell_1}_{(S^\star)^c}\|_1\right)+\lambda\|\Delta^{\ell_1}_{S^\star}\|_1.$$

利用锥约束（29），

$$\|\Delta^{\ell_1}_{(S^\star)^c}\|_1 \;\le\; 3\|\Delta^{\ell_1}_{S^\star}\|_1,$$

我们得到

$$\frac{1}{2T}\|X\Delta^{\ell_1}\|_2^2 \;\le\; \frac{\lambda}{2}(1+3)\|\Delta^{\ell_1}_{S^\star}\|_1+\lambda\|\Delta^{\ell_1}_{S^\star}\|_1 \;=\; 3\lambda\|\Delta^{\ell_1}_{S^\star}\|_1.$$

因此

$$\frac{1}{T}\|X\Delta^{\ell_1}\|_2^2 \;\le\; 6\lambda\|\Delta^{\ell_1}_{S^\star}\|_1. \tag{35}$$

现在使用（34）和不等式 $\|\Delta^{\ell_1}_{S^\star}\|_1\le\sqrt{K}\,\|\Delta^{\ell_1}_{S^\star}\|_2$来界定（35）的右侧:

$$\|\Delta^{\ell_1}_{S^\star}\|_1 \;\le\; \sqrt{K}\,\|\Delta^{\ell_1}_{S^\star}\|_2 \;\le\; \frac{\sqrt{K}}{\kappa_\star\sqrt{T}}\|X\Delta^{\ell_1}\|_2.$$

代入（35）可得

$$\frac{1}{T}\|X\Delta^{\ell_1}\|_2^2 \;\le\; 6\lambda\frac{\sqrt{K}}{\kappa_\star\sqrt{T}}\|X\Delta^{\ell_1}\|_2.$$

若 $\|X\Delta^{\ell_1}\|_2=0$，则根据（ $\Delta^{\ell_1}_{S^\star}=0$ （34），界限（30）–（32）显然成立。否则，我们可以将两边同时除以 $\|X\Delta^{\ell_1}\|_2/T$，得到

$$\frac{1}{\sqrt{T}}\|X\Delta^{\ell_1}\|_2 \;\le\; 6\lambda\frac{\sqrt{K}}{\kappa_\star}.$$

因此,

$$\frac{1}{T}\|X\Delta^{\ell_1}\|_2^2 \;\le\; \frac{36\lambda^2K}{\kappa_\star^2}, \tag{36}$$

并且，根据（34），

$$\|\Delta^{\ell_1}_{S^\star}\|_2 \;\le\; \frac{6\lambda}{\kappa_\star^2}\sqrt{K}.$$

将 λ 在（25）中的取值代入，得到

$$\|\Delta^{\ell_1}_{S^\star}\|_2 \;\le\; \frac{6c_0\sigma}{\kappa_\star^2}\sqrt{\frac{K\log m}{T}},$$

由此得到 ℓ_2 误差界（30），将常数吸收到 C_3 中即可。类似地,

$$\|\Delta^{\ell_1}\|_1 \;=\; \|\Delta^{\ell_1}_{S^\star}\|_1+\|\Delta^{\ell_1}_{(S^\star)^c}\|_1 \;\le\; 4\|\Delta^{\ell_1}_{S^\star}\|_1 \;\le\; 4\sqrt{K}\,\|\Delta^{\ell_1}_{S^\star}\|_2,$$

因此

$$\|\Delta^{\ell_1}\|_1 \;\le\; \frac{24c_0\sigma K}{\kappa_\star^2}\sqrt{\frac{\log m}{T}},$$

调整常数后得到（31）。最后，将（25）代入（36），并吸收常数，即可得到预测误差界（32）。

为得到（33），注意到

$$\left|\mathrm{supp}(\hat\theta^{\ell_1})\setminus S^\star\right| \;\le\; \frac{\|\hat\theta^{\ell_1}_{(S^\star)^c}\|_1^2}{\|\hat\theta^{\ell_1}_{(S^\star)^c}\|_2^2} \;\le\; \frac{\|\Delta^{\ell_1}_{(S^\star)^c}\|_1^2}{\|\Delta^{\ell_1}_{(S^\star)^c}\|_2^2},$$

并使用锥约束 $\|\Delta^{\ell_1}{}_{(S^\star)_c}\|_1\le 3\|\Delta^{\ell_1}_{S^\star}\|_1$ ，结合关系 $\|\Delta^{\ell_1}_{S^\star}\|_1\le$ $\sqrt{K}\|\Delta^{\ell_1}_{S^\star}\|_2$ ℓ_2 30 | θ^{ℓ_1} \以及 -界（）。简单计 $S^\star|\le C_3^2/\kappa_\star^2\ K$ $C_3>0$ □ 算可知，这将得到 supp(ˆ) ( )，其中 为适当的常数。

备注 4.7 (精确支撑集恢复)。定理 4.6保证 Lasso 至多选出 $O(K)$ 个假阳性，并在受限特征值条件下实现较小的 ℓ_2-误差和预测误差。若要实现精确支撑集恢复 $\mathrm{supp}(\hat{\theta}^{\ell_1}) = S^\star$，通常需要更强的条件，例如 [30] 的不可表示条件，或 [31]，的互不相干条件，并且需要满足类似于（10）的最小信号强度假设。这里不详细列出这些附加条件，因为后续章节中对概率联盟结构生成的分析只要求控制误差和伪联盟的数量。

4.5 对概率联盟结构生成的启示

如第 3.6节所述，Lasso 估计量会诱导出一个约简的候选有利可图联盟集合

$$\widehat{\mathcal{C}}^{\ell_1} := \{C_j \in \mathcal{C} : j \in \mathrm{supp}(\hat{\theta}^{\ell_1})\}.$$

定理 4.6 表明，在假设 4.2下，真正有利可图联盟的系数估计误差较小，且伪联盟的数量 $\widehat{\mathcal{C}}^{\ell_1} \setminus \{C_j : j \in S^\star\}$ 至多为 K的常数倍。因此，对于采样回合和噪声，以高概率而言，将任一经典 CSG 算法应用于约简联盟集 $\widehat{\mathcal{C}}^{\ell_1}$，所得到的联盟结构的期望福利接近由真正有利可图联盟构成的预言机结构的期望福利。

与 BGCP（第 3节）相比，ℓ_1方法提供了不同的权衡：BGCP 在相干性条件下能够保证精确恢复支撑集，但它是贪婪且非凸的；而 Lasso 求解一个凸优化问题，并享有基于受限特征值的误差界，但代价是允许一定数量的可控假阳性。不过，这两种方法都属于第 2节所述的同一概率联盟结构生成框架，在该框架中，联盟结构生成被分解为统计阶段（对有利可图联盟进行稀疏学习）和组合阶段（在稀疏化联盟集上运行经典 CSG）。

5 概率设计与福利保证

第3 节和4 节针对 BGCP 和 ℓ_1-松弛，基于设计矩阵 X 的几何性质（相干性和受限特征值条件），给出了高概率支撑集和误差保证。本节将完成这一论证，并证明在针对 X 各行的一种自然随机回合模型下，只要回合数 T 达到 $K \log m$ 的量级，这些几何条件就会以高概率成立。随后，我们将由此得到的估计保证转化为概率联盟结构生成的端到端福利保证。

5.1 随机情节设计模型

我们沿用线性模型

$$Y = X\theta^\star + \varepsilon, \qquad Y \in \mathbb{R}^T,\ X \in \mathbb{R}^{T\times m},\ \theta^\star \in \mathbb{R}^m, \tag{37}$$

真实支撑集为 $S^\star$，稀疏性为 $|S^\star| = K$，如式（2）所示。噪声 ε 满足次高斯假设（9）。现在，我们对 X的行施加概率模型，以刻画这样一种思想：每个回合都对应于联盟的一种随机可行配置。

用 $X_t \in \mathbb{R}^m$ 表示 X的第 t行。

假设 5.1（随机情节设计）。设计矩阵 X 按如下方式生成。

（*C1*）（独立同分布情景）各行 $\{X_t\}_{t=1}^T$ 是随机向量 $X^{(1)} \in \mathbb{R}^m$的独立同分布副本。

（*C2*）（有界稀疏行）存在常数 $L > 0$ 和整数 $q \geq 1$，使得对所有 t，

$$\|X_t\|_0 \leq q, \qquad \|X_t\|_\infty \leq L \quad \textit{almost surely.}$$

具体而言，每个回合至多激活 q 个联盟，且联盟激活均有界。

(C3)（总体协方差条件）令 t

$$\Sigma \coloneqq \mathbb{E}\left[\frac{1}{T}X^\top X\right] = \mathbb{E}\left[X^{(1)\top}X^{(1)}\right] \in \mathbb{R}^{m\times m} \tag{38}$$

表示设计矩阵的总体格拉姆矩阵。假设：

(i)（对角元远离零）存在 $\nu_{\min} > 0$，使得对所有 $j \in S^\star$，$\Sigma_{jj} \geq \nu_{\min}$。

(ii)（总体相干性）总体相干性

$$\mu(\Sigma) \coloneqq \max_{\substack{i,j\in\{1,\dots,m\}\\ i\neq j}} \frac{|\Sigma_{ij}|}{\sqrt{\Sigma_{ii}\,\Sigma_{jj}}}$$

满足

$$\mu(\Sigma) < \frac{1}{4K}. \tag{39}$$

(iii) (总体受限特征值) 存在 $\kappa_\Sigma > 0$ 使得

$$\inf_{\substack{u\in\mathbb{R}^m\setminus\{0\}\\ \|u_{(S^\star)^c}\|_1 \leq 3\|u_{S^\star}\|_1}} \frac{u^\top \Sigma u}{\|u_{S^\star}\|_2^2} \geq \kappa_\Sigma^2. \tag{40}$$

假设 5.1 刻画了这样一种情形：每个回合激活少量联盟（有界行稀疏性），而总体协方差在真实支撑集上同时具有低相干性和受限特征值性质 $S^\star$。已知这类条件在高维回归中的一大类随机设计中均成立（例如 [20, 21, 29]）。

我们记

$$\hat{\Sigma} \;:=\; \frac{1}{T} X^\top X \tag{41}$$

为经验格拉姆矩阵，使得 $\mathbb{E}[\hat{\Sigma}] = \Sigma$。

5.2 经验格拉姆矩阵的集中

$\hat{}$ 我们首先证明，在假设 5.1成立的条件下， Σ 的各元素会一致地集中于 Σ的对应元素周围。

引理 5.2 ($\hat{}$ 的逐元素集中Σ)。 假设*5.1(C1)–(C2)*成立。那么存在常数 c_1、 $c_2 > 0$，它们仅取决于 L 和 q，使得对于任意 $T \geq 1$ 以及任意 $\delta \in (0,1)$，

$$\mathbb{P}\left(\left\| \hat{\Sigma} - \Sigma \right\|_{\max} \;>\; c_1 \sqrt{\frac{\log(m^2/\delta)}{T}} \right) \;\leq\; \delta, \tag{42}$$

其中 $\|A\|_{\max} = \max_{i,\,j} |A_{ij}|$ 表示逐元素 ℓ_∞-范数。

*Proof*固定索引 i， $j \in \{1, \dots, m\}$。根据定义，

$$\hat{\Sigma}_{ij} - \Sigma_{ij} \;=\; \frac{1}{T} \sum_{t=1}^{T} \Big(X_{ti} X_{tj} - \mathbb{E}[X_{ti} X_{tj}] \Big).$$

这些被加项是相互独立、均值为零的随机变量，其绝对值由 L^2 限制（因为 $|X_{ti}|$， $|X_{tj}| \leq L$ 几乎处处成立）。此外，根据行稀疏性假设 $\|X_t\|_0 \leq q$，任意给定行中至多有 q^2 对 $(i,\ j)$ 可以为非零，这保证了每个被加项的方差具有统一上界。对每个固定的索引对 $(i,\ j)$ 应用 Hoeffding不等式（或 Bernstein不等式），可得

$$\mathbb{P}\Big(|\hat{\Sigma}_{ij} - \Sigma_{ij}| > t \Big) \;\leq\; 2 \exp\left(-c \frac{T t^2}{L^4} \right)$$

对于某个数值常数 $c > 0$。总共有 m^2 个条目。根据联合界，

$$\mathbb{P}\Big(\|\hat{\Sigma} - \Sigma\|_{\max} > t \Big) \;\leq\; 2m^2 \exp\left(-c \frac{T t^2}{L^4} \right).$$

取 $t = c_1 \sqrt{(\log(m^2/\delta))/T}$，其中 c_1 足够大，并化简右端，可得（42），其中常数仅取决于 L和 q。 □

$\hat{}$ 作为一个直接推论，当 T 的阶为 $\log m$ 时， Σ 的对角元素和非对角元素继承 Σ 的总体性质。

推论 5.3 (对角稳定性和经验相干性)。 假设 5.1成立。存在常数 c_3, $c_4 > 0$ 使得如果

$$T \;\geq\; c_3 \log(m), \tag{43}$$

那么，以至少 $1 - m^{-c_4}$ 的概率，以下结论同时成立：

(i)（对角元远离零）对所有 $j \in S^\star$,

$$\hat{\Sigma}_{jj} \;\geq\; \frac{\nu_{\min}}{2}. \tag{44}$$

(ii)（经验相干性界）经验相干性

$$\mu(\hat{\Sigma}) \;:=\; \max_{\substack{i,j\in\{1,\dots,m\}\\ i\neq j}} \frac{|\hat{\Sigma}_{ij}|}{\sqrt{\hat{\Sigma}_{ii}\,\hat{\Sigma}_{jj}}}$$

满足

$$\mu(\hat{\Sigma}) \;\leq\; \mu(\Sigma) + \frac{1}{8K} \;<\; \frac{1}{2K}, \tag{45}$$

其中 $\mu(\Sigma)$定义见假设5.1*(C3)(ii)*。

*Proof*应用引理 5.2 并令 $\delta = m^{-c}$，选择 T满足 (43)，使得 $\|\hat{}\Sigma - \Sigma\|_{\max} \leq \eta$ 以高概率成立，其中 $\eta > 0$ 的取值将相对于 $\nu_{\min}$ 和 $\mu(\Sigma)$ 选得较小。

对于 (i)，注意到对于 $j \in S^\star$,

$$\hat{\Sigma}_{jj} \;\geq\; \Sigma_{jj} - |\hat{\Sigma}_{jj} - \Sigma_{jj}| \;\geq\; \nu_{\min} - \eta.$$

如果我们选择 $\eta \leq \nu_{\min}/2$，则得到 $\Sigma_{jj} \hat{\geq} \nu_{\min}/2$，从而得到 (44)。

对于 (ii,) 对于任意 $i \neq j$，我们有

$$|\hat{\Sigma}_{ij}| \;\leq\; |\Sigma_{ij}| + \eta, \qquad \hat{\Sigma}_{ii} \;\geq\; \Sigma_{ii} - \eta, \qquad \hat{\Sigma}_{jj} \;\geq\; \Sigma_{jj} - \eta.$$

因此，

$$\frac{|\hat{\Sigma}_{ij}|}{\sqrt{\hat{\Sigma}_{ii}\,\hat{\Sigma}_{jj}}} \;\leq\; \frac{|\Sigma_{ij}| + \eta}{\sqrt{(\Sigma_{ii} - \eta)(\Sigma_{jj} - \eta)}}.$$

利用 Σ_{ii}, $\Sigma_{jj} \geq \nu_{\min}$ 并选择 $\eta \leq \nu_{\min}/4$，我们得到

$$\sqrt{(\Sigma_{ii} - \eta)(\Sigma_{jj} - \eta)} \;\geq\; \frac{\nu_{\min}}{2},$$

从而

$$\frac{|\hat{\Sigma}_{ij}|}{\sqrt{\hat{\Sigma}_{ii}\,\hat{\Sigma}_{jj}}} \;\leq\; \frac{2}{\nu_{\min}}\big(|\Sigma_{ij}| + \eta\big) \;\leq\; \frac{2}{\nu_{\min}}\,|\Sigma_{ij}| + \frac{2\eta}{\nu_{\min}}.$$

对 $i \neq j$ 取最大值，并回顾 $\mu(\Sigma)$ 的定义，可得

$$\mu(\hat{\Sigma}) \;\leq\; \mu(\Sigma) + \frac{2\eta}{\nu_{\min}}.$$

通过选择 η 足够小，使得 $2\eta/\nu_{\min} \leq 1/(8K)$，我们得到（45）。不等式（45）中的最后一个不等式由假设 $\mu(\Sigma) < 1/(4K)$ 得出。通过在引理 5.2 中令 η 足够大，即可满足所需的 T 取值要求。 □

5.3 经验设计矩阵的受限特征值

ˆ接下来，我们证明当 T 足够大时，经验格拉姆矩阵 Σ 继承了 Σ的受限特征值性质。证明采用高维回归中的标准论证 [28, 29], ，并针对定义 4.1中使用的锥作了调整。

引理 5.4（由总体受限特征值得到经验受限特征值）。 假设 *5.1* 成立。存在常数 c_5, $c_6 > 0$ 使得若

$$T \;\geq\; c_5 K \log(m), \tag{46}$$

则以至少 $1 - m^{-c_6}$ 的概率，

$$\inf_{\substack{u\in\mathbb{R}^m\setminus\{0\}\\ \|u_{(S^\star)^c}\|_1\le 3\|u_{S^\star}\|_1}} \frac{u^\top\hat{\Sigma}u}{\|u_{S^\star}\|_2^2} \;\geq\; \frac{\kappa_\Sigma^2}{2}. \tag{47}$$

等价地，在此事件上，

$$\kappa(S^\star, 3) \;\geq\; \frac{\kappa_\Sigma}{\sqrt{2}},$$

其中 κ（$S^\star$，3）是式 （23）中定义的受限特征值常数。

Proof 为简洁起见，记锥为

$$\mathcal{C} \;\coloneqq\; \left\{u \in \mathbb{R}^m \setminus \{0\} : \|u_{(S^\star)^c}\|_1 \le 3\|u_{S^\star}\|_1\right\}.$$

对于任意满足 $\|u_{S^\star}\|_2 = 1$ 的 $u \in \mathcal{C}$，我们可以写作

$$u^\top\hat{\Sigma}u \;=\; u^\top\Sigma u + u^\top(\hat{\Sigma} - \Sigma)u.$$

根据假设 5.1(C3)(iii),

$$u^\top\Sigma u \;\geq\; \kappa_\Sigma^2.$$

因此，我们需要控制

$$\sup_{\substack{u\in\mathcal{C}\\ \|u_{S^\star}\|_2=1}} \left|u^\top(\hat{\Sigma} - \Sigma)u\right|.$$

注意到

$$u^\top(\hat{\Sigma} - \Sigma)u \;=\; \sum_{i,j=1}^{m} u_i u_j(\hat{\Sigma}_{ij} - \Sigma_{ij}),$$

从而

$$\left|u^\top(\hat{\Sigma} - \Sigma)u\right| \;\le\; \|\hat{\Sigma} - \Sigma\|_{\max}\left(\sum_{i=1}^{m}|u_i|\right)^2 \;=\; \|\hat{\Sigma} - \Sigma\|_{\max}\|u\|_1^2.$$

对于 $u \in \mathcal{C}$,

$$\|u\|_1 \;=\; \|u_{S^\star}\|_1 + \|u_{(S^\star)^c}\|_1 \;\le\; 4\|u_{S^\star}\|_1 \;\le\; 4\sqrt{K}\,\|u_{S^\star}\|_2 \;=\; 4\sqrt{K},$$

其中我们使用了 $\|u_{S^\star}\|_1 \le \sqrt{K}\|u_{S^\star}\|_2$。因此，

$$\left|u^\top(\hat{\Sigma}-\Sigma)u\right| \;\le\; 16K\,\|\hat{\Sigma}-\Sigma\|_{\max}$$

对所有 $u \in \mathcal{C}$ 且满足 $\|u_{S^\star}\|_2 = 1$。

根据引理 5.2，如果 $T \ge c_5 K\log(m)$ 对某个足够大的常数 c_5成立，那么以至少 $1-m^{-c_6}$ 的概率，我们有

$$\|\hat{\Sigma}-\Sigma\|_{\max} \;\le\; \frac{\kappa_\Sigma^2}{32K},$$

从而

$$\sup_{\substack{u\in\mathcal{C}\\ \|u_{S^\star}\|_2=1}} \left|u^\top(\hat{\Sigma}-\Sigma)u\right| \;\le\; \frac{\kappa_\Sigma^2}{2}.$$

将此结果与总体下界 $u^\top\Sigma u \ge \kappa_\Sigma^2$ 结合，可得

$$u^\top\hat{\Sigma}u \;\ge\; \kappa_\Sigma^2 - \frac{\kappa_\Sigma^2}{2} \;=\; \frac{\kappa_\Sigma^2}{2},$$

对所有 $u \in \mathcal{C}$ 且满足 $\|u_{S^\star}\|_2 = 1$，这正是式（47）。受限特征值常数的等价性可通过比较 $\kappa(S^\star, 3)$ 与锥 $\mathcal{C}$的定义得出。□

5.4 BGCP 和 Lasso 的设计条件

现在结合推论 5.3和引理5.4可以证明，在随机回合设计模型下，一旦 T达到 $K\log m$ 的数量级，BGCP 所需的相干性条件和 Lasso 所需的受限特征值条件均以高概率成立。

定理 5.5（随机回合下的设计矩阵条件）。假设 *5.1*成立。则存在常数 C_0、c_7、$c_8 > 0$，使得当

$$T \;\ge\; C_0 K\log(m), \tag{48}$$

以至少 $1-c_7 m^{-c_8}$ 的概率，以下条件同时成立：

（*i*）（BGCP 的相干性条件）在（4）中定义的互相干性 $\mu(X)$ 满足

$$\mu(X) \;<\; \frac{1}{2K-1}, \tag{49}$$

因此假设*3.2*（*A1*）成立。

（*ii*）（Lasso 的受限特征值条件）在（23）中定义的受限特征值常数 $\kappa(S^\star, 3)$ 满足

$$\kappa(S^\star, 3) \;\ge\; \frac{\kappa_\Sigma}{\sqrt{2}}, \tag{50}$$

从而假设 *4.2**(B1)* 在 $\kappa_\star = \kappa_\Sigma/\sqrt{2}$ 下成立。

Proof 根据推论 5.3，对于 $T \geq c_3 \log(m)$，我们以至少 $1 - m^{-c_4}$ 的概率有：

$$\mu(\hat{\Sigma}) \leq \mu(\Sigma) + \frac{1}{8K} < \frac{1}{2K},$$

利用假设 $\mu(\Sigma) < 1/(4K)$。在同一事件下，根据式（44），对角元素 $\hat{\Sigma}_{jj}$ 在 $S^\star$ 上均与零保持一定距离。利用列归一化（3）以及 $\hat{}$ 与 Σ 和 X之间的关系，可得

$$\mu(X) = \max_{i \neq j} \frac{|X_i^\top X_j|}{\|X_i\|_2 \|X_j\|_2} = \max_{i \neq j} \frac{|\hat{\Sigma}_{ij}|}{\sqrt{\hat{\Sigma}_{ii}\,\hat{\Sigma}_{jj}}} \leq \mu(\hat{\Sigma}) < \frac{1}{2K},$$

并且由于 $1/(2K) < 1/(2K-1)$ 对于 $K \geq 1$成立，因此可得式（49）。因此，当 $T \geq c_3 \log(m)$ 时，相干性条件以高概率成立。

类似地，根据引理 5.4，当 $T \geq c_5 K \log(m)$ 时，以至少 $1 - m^{-c_6}$ 的概率，有

$$\kappa(S^\star, 3) \geq \frac{\kappa_\Sigma}{\sqrt{2}},$$

这恰好就是式（50）。取 $C_0 = \max\{c_3, c_5\}$，并通过联合界合并两个高概率事件，即可在适当的常数 c_7、$c_8 > 0$ 下得到所述结果。□

定理 5.5 表明，在随机情节设计模型下，随着回合数增至 $T \gtrsim K \log m$，定理 3.7 和 4.6 所需的几何假设便成为典型情形，这与压缩感知和高维回归中的经典结果一致 [20, 21, 29]。

5.5 BGCP 的端到端福利最优性

现在，我们将 BGCP 的支集恢复保证转化为概率联盟结构生成情形下的福利最优性结果。

回顾第 2 节，对于每个联盟结构 $\mathcal{P}$（即智能体集合的一种划分），福利可写成如下参数形式：

$$W(\mathcal{P}; \theta) = z(\mathcal{P})^\top \theta, \tag{51}$$

其中，$z(\mathcal{P}) \in \{0,1\}^m$ 是所用联盟的指示向量：若联盟 C_j 出现在 $\mathcal{P}$ 中，则 $z_j(\mathcal{P}) = 1$；否则，$z_j(\mathcal{P}) = 0$。根据可行性，每个联盟结构至多使用 M 个联盟，其中 M 为仅取决于参与者数量的某个有限值，因此

$$\|z(\mathcal{P})\|_0 \leq M, \qquad \|z(\mathcal{P})\|_\infty \leq 1. \tag{52}$$

Let

$$\mathcal{P}^\star \in \arg\max_{\mathcal{P}} W(\mathcal{P}; \theta^\star) \tag{53}$$

将真实参数 $\theta^\star$下的最优联盟结构记为.....。不失一般性地，我们假设在 $\theta^\star$ 下所有贡献非零的联盟都属于 $S^\star$，因此，在 $\mathcal{P}^\star$ 中任何有利可图地使用的联盟，其索引都属于 $S^\star$。

基于 BGCP 的联盟结构生成流程如下：

1. 运行 BGCP（定义 3.1），使用 $K_{\max} = K$ 和 $\eta = 0$ 以获得支撑集 $S^{\mathrm{BGCP}} \subseteq \{1, \dots, m\}$。2. 构建约简联盟集 $\widehat{\mathcal{C}}^{\mathrm{BGCP}} = \{C_j \in \mathcal{C} : \ j \in S^{\mathrm{BGCP}}\}$。3. 使用精确算法（例如动态规划或分支定界 [1, 2]）求解限制在 $\widehat{\mathcal{C}}^{\mathrm{BGCP}}$ 上的确定性 CSG 问题，从而得到联盟结构 $\widehat{\mathcal{P}}^{\mathrm{BGCP}}$。

定理 5.6 (BGCP 的福利最优性具有高概率)。 假设 3.2和 5.1成立，且回合数 T 满足 (48)。进一步假设，$\theta_{\min}$ 在 (10) 中的最小信号足够大，以满足推论 3.6的分离条件。那么存在常数 c_9，$c_{10} > 0$ 使得

$$\mathbb{P}\Big(\widehat{\mathcal{P}}^{\mathrm{BGCP}} = \mathcal{P}^\star\Big) \ \geq \ 1 - c_9 m^{-c_{10}}. \tag{54}$$

特别地，以相同的概率，

$$W\big(\widehat{\mathcal{P}}^{\mathrm{BGCP}}; \theta^\star\big) \ = \ W\big(\mathcal{P}^\star; \theta^\star\big). \tag{55}$$

Proof 5 5, T 48 1 − 根据定理，对于满足（）的情 $c_7 m^{-c_8}$ 况，我们以至少为 的概率同时得到相干性条件（）和受限特征值条件（）。特别地，假设 3.2（A1）在此事件下成立。 49 50

另一方面，引理 3.3 和假设 3.2（A2）意味着，式（$\mathcal{E}_{\mathrm{noise}}$ 12）中的噪声事件以至少 $1 - c_{11} m^{-c_{12}}$ 的概率发生，其中常数 c_{11}，$c_{12} > 0$取值适当。在这些事件与推论 3.6的最小信号条件同时成立时，定理 3.7 适用，并可得 $S^{\mathrm{BGCP}} = S^\star$。因此，在这些事件的交集上，

$$\widehat{\mathcal{C}}^{\mathrm{BGCP}} \ = \ \{C_j \in \mathcal{C} : j \in S^\star\}.$$

由于 $\theta^\star$ 下所有有利可图的联盟都属于 $S^\star$，因此式（53）中的福利最大化联盟结构 $\mathcal{P}^\star$ 仅使用 $\widehat{\mathcal{C}}^{\mathrm{BGCP}}$ 中的联盟，因此对受限 CSG 问题是可行的。反过来，任何仅使用 $\widehat{\mathcal{C}}^{\mathrm{BGCP}}$ 中联盟的联盟结构，其福利由式（51）给出，其中 $\theta^\star_j = 0$ 对于 $j \notin S^\star$成立，因此受限 CSG 问题与原始 CSG 问题完全相同。由于我们假设流程第 3 步使用精确 CSG 求解器，受限问题会得到一个联盟结构 $\widehat{\mathcal{P}}^{\mathrm{BGCP}}$，它与 $\mathcal{P}^\star$相同。

通过联合界合并设计矩阵事件和噪声事件的概率，并将常数吸收到 c_9中，c_{10}可得高概率结论（54），由此立即得到（55）。□

注 5.7 （近似 CSG 求解器）。如果第 3 步中的确定性 CSG 求解器仅为 α-近似（即其输出的联盟结构的福利与最优受限福利相差不超过因子 $\alpha \geq 1$），那么在事件 $S^{\mathrm{BGCP}} = S^\star$ 发生时，我们仍可得到

$$W\big(\widehat{\mathcal{P}}^{\mathrm{BGCP}}; \theta^\star\big) \ \geq \ \frac{1}{\alpha} W\big(\mathcal{P}^\star; \theta^\star\big).$$

因此，定理 5.6 可以扩展为，在 BGCP 与近似 CSG 求解器结合使用时，为福利提供高概率 α-近似保证。

5.6 通过 ℓ_1 松弛进行福利近似

最后，我们利用定理 ℓ_1 中的误差界和定理 4.6 中的设计矩阵保证，为该流程推导出定量的福利差距界。5.5。

对于基于 ℓ_1 的流程，我们考虑以下步骤:

1. 通过（21）计算 Lasso 估计量 $\hat{}\theta^{\ell_1}$ ，其中 λ 根据假设 4.2(B2) 选取。2. 在完整的联盟集合 $\mathcal{C}$上，针对替代参数 $\hat{}\theta^{\ell_1}$ 求解确定性 CSG 问题，得到

$$\widehat{\mathcal{P}}^{\ell_1} \in \arg\max_{\mathcal{P}} W(\mathcal{P}; \hat{\theta}^{\ell_1}).$$

也就是说，与 BGCP 不同，我们不预先限制联盟集合；$\hat{}\theta^{\ell_1}$ 的稀疏性会隐式地抑制无利可图的联盟。

我们首先注意到，福利泛函相对于 θ 具有一个简单的利普希茨性质。

引理 5.8（福利的利普希茨连续性）。令 $\mathcal{P}$ 为任意联盟结构，并令θ、 $\theta' \in \mathbb{R}^m$。则

$$\left|W(\mathcal{P};\theta) - W(\mathcal{P};\theta')\right| \le \|\theta - \theta'\|_1. \tag{56}$$

Proof 根据参数形式（51），

$$W(\mathcal{P};\theta) - W(\mathcal{P};\theta') = z(\mathcal{P})^\top(\theta - \theta').$$

Thus

$$\left|W(\mathcal{P};\theta) - W(\mathcal{P};\theta')\right| \le \|z(\mathcal{P})\|_\infty \|\theta - \theta'\|_1 \le \|\theta - \theta'\|_1,$$

由于 $\|z(\mathcal{P})\|_\infty \le 1$ 由（52）可得。 □

现在可以陈述福利近似定理。

定理 5.9（通过 Lasso 进行福利近似）。假设 *5.1* 和 *4.2*成立，且回合数 T 满足（48）。设 $\mathcal{P}^\star$ 为根据（53）定义的 $\theta^\star$ 下的最优联盟结构， $\widehat{\mathcal{P}}^{\ell_1}$ 为通过上述基于 ℓ_1 的流程得到的联盟结构。那么，存在常数 C_2、 $C_3 > 0$ ，使得其发生概率至少为 $1 - C_2 m^{-C_3}$,

$$W\big(\mathcal{P}^\star;\theta^\star\big) - W\big(\widehat{\mathcal{P}}^{\ell_1};\theta^\star\big) \le C_3 \frac{\sigma K}{\kappa_\Sigma^2}\sqrt{\frac{\log m}{T}}. \tag{57}$$

Proof 根据定理 5.5，对于满足条件 T（48）的情况，以高概率，受限特征值条件（50）以 $\kappa_\star = \kappa_\Sigma/\sqrt{2}$ 成立。在此事件下，定理 4.6 适用，并得出

$$\|\hat{\theta}^{\ell_1} - \theta^\star\|_1 \le \frac{C_4 \sigma K}{\kappa_\star^2}\sqrt{\frac{\log m}{T}} = \frac{2C_4 \sigma K}{\kappa_\Sigma^2}\sqrt{\frac{\log m}{T}},$$

其中某个常数为 $C_4 > 0$。

现在，我们比较 $\theta^\star$ 下 $\mathcal{P}^\star$ 和 $\widehat{\mathcal{P}}^{\ell_1}$ 的福利。我们有

$$W(\mathcal{P}^\star;\theta^\star) - W(\widehat{\mathcal{P}}^{\ell_1};\theta^\star)$$
$$= \Big[W(\mathcal{P}^\star;\theta^\star) - W(\mathcal{P}^\star;\hat{\theta}^{\ell_1})\Big] + \Big[W(\mathcal{P}^\star;\hat{\theta}^{\ell_1}) - W(\widehat{\mathcal{P}}^{\ell_1};\hat{\theta}^{\ell_1})\Big] + \Big[W(\widehat{\mathcal{P}}^{\ell_1};\hat{\theta}^{\ell_1}) - W(\widehat{\mathcal{P}}^{\ell_1};\theta^\star)\Big].$$

根据 $\widehat{\mathcal{P}}^{\ell_1}$ 是 $W(\cdot;\hat{\theta}^{\ell_1})$ 的最大化解这一点，中间项非正：

$$W(\mathcal{P}^\star;\hat{\theta}^{\ell_1}) - W(\widehat{\mathcal{P}}^{\ell_1};\hat{\theta}^{\ell_1}) \le 0.$$

将引理 5.8 应用于第一项和第三项，可得

$$W(\mathcal{P}^\star;\theta^\star) - W(\widehat{\mathcal{P}}^{\ell_1};\theta^\star) \le \|\hat{\theta}^{\ell_1} - \theta^\star\|_1 + \|\hat{\theta}^{\ell_1} - \theta^\star\|_1 = 2\|\hat{\theta}^{\ell_1} - \theta^\star\|_1.$$

将此不等式与上文的 ℓ_1-误差界结合，可得

$$W(\mathcal{P}^\star;\theta^\star) - W(\widehat{\mathcal{P}}^{\ell_1};\theta^\star) \le \frac{4C_4\sigma K}{\kappa_\Sigma^2}\sqrt{\frac{\log m}{T}},$$

在定理 5.5 和4.6中的设计事件与噪声事件的交集上。调整常数并通过联合界合并概率，即完成了对（57）的证明。□

定理 5.6和 5.9 为概率联盟结构生成提供了端到端保证：当设计矩阵具有相干性且信号足够强时，BGCP 以高概率实现福利最优性；而基于 ℓ_1 的方法则在受限特征值条件下，将福利差距控制在 $\sigma K\sqrt{(\log\, m)/T}/\kappa_\Sigma^2$ 的量级。两种情况下，所需回合数均为 $K\log m$ 的量级，这与稀疏恢复的经典样本复杂度界一致[20, 21, 29]。

6 与其他概率 CSG 方案的比较

第 3–5 节分析了本文提出的概率 CSG 流程：稀疏回合价值模型，可通过 BGCP 或 ℓ_1 正则化估计器进行估计，随后由确定性 CSG 求解器求解。本节将这些流程与自然的概率 CSG 基线方法进行比较，并指出我们的方法在哪些参数范围内具有可证明的优势，以及经典方法在哪些范围内可能具有竞争力，甚至更可取。

全文沿用以下数据模型

$$Y = X\theta^\star + \varepsilon, \qquad Y \in \mathbb{R}^T,\ X \in \mathbb{R}^{T\times m},\ \theta^\star \in \mathbb{R}^m, \tag{58}$$

其真实支撑集为 $S^\star$，稀疏性为 $|S^\star| = K$，噪声为次高斯噪声 ε，如假设 3.2（A2）所述；并采用假设 5.1中的随机情节设计模型。福利定义为 $W(\mathcal{P};\theta) = z(\mathcal{P})^\top\theta$，如式（51）所示，其中 $z(\mathcal{P})$ 是 $\mathcal{P}$ 中所用联盟的指示向量。

6.1 概率 CSG 基线方案

我们首先形式化两种简单但自然的概率 CSG 基线方法。

基线 ***A***：情节式插件 ***CSG***（***EPC***）。
对于每个联盟索引 $j \in \{1, \ldots, m\}$，令

$$\mathcal{T}_j \coloneqq \{t \in \{1, \ldots, T\} : X_{tj} \neq 0\}$$

表示联盟 C_j 处于激活状态的回合集合，并令 $n_j = |\mathcal{T}_j|$。针对 $\theta_j^\star$ 的情节式插件估计量为

$$\hat{\theta}_j^{\mathrm{EPC}} \coloneqq \begin{cases} \dfrac{1}{n_j} \displaystyle\sum_{t \in \mathcal{T}_j} \dfrac{Y_t}{X_{tj}}, & \text{if } n_j > 0, \\ 0, & \text{if } n_j = 0. \end{cases} \tag{59}$$

换言之，$\hat{}\,\theta_j^{\mathrm{EPC}}$是联盟 C_j 在其处于激活状态的回合中的单位激活平均收益，不考虑与其他联盟的交互。

EPC 流程如下：
1. 通过 (59) 计算 $\hat{}\theta^{\mathrm{EPC}}$ 。2. 在价值函数 $C_j \mapsto \hat{}\,\theta_j^{\mathrm{EPC}}$上运行精确确定性 CSG 求解器 [1, 2]，并得到联盟结构 $\widehat{\mathcal{P}}^{\mathrm{EPC}}$。

EPC 是一种自然的“多臂老虎机式”方法：它使用情节数据，但不利用 X的稀疏性或联合结构。

基线 ***B***：稠密最小二乘 ***CSG***（***DLS***）。
第二种基线方法是将 $\theta^\star$ 视为稠密，并通过普通最小二乘法（OLS）进行估计，

$$\hat{\theta}^{\mathrm{DLS}} \in \arg\min_{\theta \in \mathbb{R}^m} \frac{1}{2T} \|Y - X\theta\|_2^2, \qquad (T \geq m,\ X^\top X \text{ invertible}), \tag{60}$$

或者更一般地，通过岭回归或带有关于 θ [32, 33] 的高斯先验的稠密贝叶斯线性回归进行估计。DLS 流程如下：
1. 通过（60）计算 $\hat{}\theta^{\mathrm{DLS}}$ 。 2. 对 $C_j \mapsto \hat{}\,\theta_j^{\mathrm{DLS}}$运行精确
CSG 求解器，得到 $\widehat{\mathcal{P}}^{\mathrm{DLS}}$。
该基线方法使用完整的设计矩阵，但同样没有利用稀疏性；通常需要 $T \gtrsim m$ 个回合才能保持稳定。

相比之下，我们的 BGCP 和 ℓ_1-流程明确假设 $K \ll m$ ，并在高维情形 $T \ll m$下运行，只需 $T \asymp K \log m$ 回合，并满足假设 5.1。

6.2 福利差距作为比较指标

对于任何从数据中生成联盟结构 $\widehat{\mathcal{P}}^{\mathsf{Alg}}$ 的算法 Alg ，我们将样本量为 T 时的福利差距 定义为

$$\Delta_{\mathsf{Alg}}(T) \;:=\; W(\mathcal{P}^\star;\theta^\star) - W(\widehat{\mathcal{P}}^{\mathsf{Alg}};\theta^\star), \tag{61}$$

其中 $\mathcal{P}^\star$ 是在 $\theta^\star$ 下的最优联盟结构，如式（53）所示。我们关注 $\Delta_{\mathsf{Alg}}(T)$ 随 T 增长时的高概率上界，并希望找出这样一些情形：我们的方法所得上界严格优于 EPC 或 DLS 的上界。

对于 BGCP 和基于 ℓ_1 的流程，定理 5.6 和 5.9 意味着：

$$\Delta_{\mathrm{BGCP}}(T) = 0 \quad \text{with high probability, for } T \gtrsim K\log m, \tag{62}$$

$$\Delta_{\ell_1}(T) \;\le\; C\,\frac{\sigma K}{\kappa_\Sigma^2}\sqrt{\frac{\log m}{T}} \quad \text{with high probability,} \tag{63}$$

其中常数 $C>0$ 取决于受限特征值常数 κ_Σ 和噪声水平σ。现在，我们将在简化假设下为 EPC 和 DLS 推导类似的界，并比较由此得到的情形。

6.3 我们的方法占优的稀疏高维情形

我们首先确定一种情形，在该情形下，可以证明稀疏概率 CSG 流程优于 EPC 和 DLS。为清楚起见，我们采用一种简化的独立激活设计，该设计与假设5.1一致。

假设 6.1（独立激活设计）。 除假设 5.1外，假设 $X_{tj}\in\{0,1\}$ 对所有 t， j，并且对每个联盟 j，

$$\mathbb{P}(X_{tj}=1) = p_j \quad \textit{for all } t,$$

其中 $p_j \in (0,1]$，在 t 和 j 上相互独立，并且几乎必然满足行稀疏约束$\|X_t\|_0 \le q$ 。令

$$p_{\min} \;:=\; \min_{j\in\{1,\dots,m\}} p_j \quad \textit{and} \quad p_{\max} \;:=\; \max_{j\in\{1,\dots,m\}} p_j.$$

假设 $p_{\min}>0$ ，且 ε_t 独立同分布，满足 $\mathbb{E}[\varepsilon_t]=0$ 和 $\mathbb{E}[\varepsilon_t^2]=\sigma^2$。

为分离稀疏性和样本量的影响，我们考虑如下情形：

$$K \ll m, \qquad C_1 K\log m \;\le\; T \;\le\; C_2 m \tag{64}$$

对于常数 C_1， $C_2>0$。因此，该问题是高维的（$T\ll m$），但回合数足够大，能够满足定理5.5的条件。

我们先给出 EPC 估计量的一个简单界。

引理 6.2 (EPC 估计误差)。 假设 *6.1* 成立。那么对于任意 j满足 $p_j > 0$ 以及任意 $T \geq 1$,

$$\mathbb{E}\Big[\big(\hat{\theta}_j^{\mathrm{EPC}} - \theta_j^\star\big)^2\Big] \;\geq\; \frac{\sigma^2}{2Tp_j}, \tag{65}$$

已提供 $Tp_j \geq 2$。此外，存在一个常数 $c > 0$ （仅取决于 $p_{\min}$、$p_{\max}$），使得以至少 $1 - m^{-2}$ 的概率，

$$\|\hat{\theta}^{\mathrm{EPC}} - \theta^\star\|_2^2 \;\geq\; c\,\frac{\sigma^2 K}{T}. \tag{66}$$

Proof sketch 在给定 n_j的条件下，索引 j 的 EPC 估计量是 n_j 个含噪观测的平均值，其方差至少为 σ^2 $\mathbb{E}[\,\theta_j^{\mathrm{EPC}} - \theta_j^\star\,^2 \mid n_j] \geq$，忽略干 σ^2/n_j $\mathbb{E}[1/n_j\,\mathbf{1}\{n_j > 0\}] \geq\; 2Tp_j\,^{-1}$ $Tp_j \geq 2$ 扰。因此（ˆ）。取期望，并在满足 时使用（），得到（65）。对 $j \in S^\star$ 求和并使用 $p_j \leq p_{\max}$，得到 $\mathbb{E}\|\hat{}\theta^{\mathrm{EPC}} - \theta^\star\|_2^2 \geq c(\sigma^2 K/T)$，其中 $c > 0$。对独立次指数变量之和应用标准集中论证，即可得到（66）以高概率成立。□

结合引理 6.2 与引理5.8的福利利普希茨性质，我们得到 EPC 福利差距的一个下界。

命题 6.3 （稀疏情形下的 EPC 福利差距）。 假设 *5.1*和*6.1* 成立，且满足（64）。那么存在常数 c_1、$c_2 > 0$，使得对于所有充分大的 T，以至少 $1 - c_1 m^{-2}$ 的概率，

$$\Delta_{\mathrm{EPC}}(T) \;\geq\; c_2\,\sigma\sqrt{\frac{K}{T}}. \tag{67}$$

*Proof*根据引理 6.2， $\|\hat{}\theta^{\mathrm{EPC}} - \theta^\star\|_2^2 \geq c$（$\sigma^2 K/T$）以高概率成立。由于 $\|\cdot\|_1 \geq \|\cdot\|_2$，我们有 $\|\hat{}\theta^{\mathrm{EPC}} - \theta^\star\|_1 \geq \sqrt{c}\,\sigma\sqrt{K/T}$在同一事件上成立。应用引理 5.8 并令 $\mathcal{P} = \mathcal{P}^\star$ 和$\theta = \theta^\star$， $\theta' = \hat{}\theta^{\mathrm{EPC}}$，可得

$$\big|W(\mathcal{P}^\star;\theta^\star) - W(\mathcal{P}^\star;\hat{\theta}^{\mathrm{EPC}})\big| \;\geq\; \sqrt{c}\,\sigma\sqrt{\frac{K}{T}}.$$

EPC 联盟结构 $\widehat{\mathcal{P}}^{\mathrm{EPC}}$最大化 W W $\geq$ $(\cdot;\hat{}\theta^{\mathrm{EPC}})$，因此（$\widehat{\mathcal{P}}^{\mathrm{EPC}};\hat{}\theta^{\mathrm{EPC}}$）$W$ $\mathcal{P}^\star$ θ^{EPC}（$;\,$）。由简单的三角不等式可得

$$W(\mathcal{P}^\star;\theta^\star) - W(\widehat{\mathcal{P}}^{\mathrm{EPC}};\theta^\star) \;\geq\; \frac{1}{2}\,\big|W(\mathcal{P}^\star;\theta^\star) - W(\mathcal{P}^\star;\hat{\theta}^{\mathrm{EPC}})\big|,$$

在.....事件发生时， $W(\,\widehat{\mathcal{P}}^{\mathrm{EPC}};\hat{}\theta^{\mathrm{EPC}}) \geq W(\mathcal{P}^\star;\hat{}\theta^{\mathrm{EPC}})$。结合这些不等式可得 (67)，其中 $c_2 = \sqrt{c}/2$。□

相比之下，定理 5.6和 5.9 表明，在相同的情形（64）下，BGCP 以 $\Delta_{\mathrm{BGCP}}(T) = 0$ 高概率实现 ℓ_1-驱动的流程满足（63）。由于当 $\sqrt{K/T} \gg K\sqrt{(\log m)/T}$时， $K \ll m$ 且 $T \asymp K \log m$，我们得到：

推论 6.4 (我们方法的稀疏高维优势)。 假设条件 *3.2*、*4.2*、*5.1* 和 *6.1* 成立，并且当常数 C_1、C_2 足够大时，(64) 成立。那么对于范围 $C_1 K \log m \le T \le C_2 m$内的 T，

(i) BGCP 以高概率达到零福利差距： $\Delta_{\mathrm{BGCP}}(T) = 0$。 *(ii)* ℓ_1流程满足 $\Delta_{\ell_1}(T) \le C\,(\sigma K/\kappa_\Sigma^2)\sqrt{}$（对数 m）$/T$。 *(iii) EPC* 基线以高概率满足 $\Delta_{\mathrm{EPC}}(T) \ge c_2\sigma\sqrt{K/T}$。

特别地，对于该情形下固定的 (K, m) 和 T，*BGCP* 和 ℓ_1 流程的福利差距都严格小于 *EPC*，差距仅相差常数因子。此外，当 $T < m$ 时，*DLS* 基线不适用，因为 $X^\top X$ 是秩亏的，且（60）定义不适定。

推论 6.4 形式化了这样一个直观观点：当价值函数确实具有稀疏性（$K \ll m$），且回合数处于高维情形（$T \ll m$）时，利用稀疏性至关重要：EPC 无法达到 BGCP 或流程的福利保证，而 DLS 等稠密方法甚至无法定义。

6.4 经典方法具有竞争力的稠密情形

现在我们讨论一种互补情形，在该情形下，经典稠密概率 CSG 方案可能具有竞争力或更为合适。

考虑一系列满足 $m \to \infty$ and $K = K(m)$ 且 $K/m \to \rho \in (0, 1]$ 的问题。假设 T 的增长速度快于 m，即

$$T \;\ge\; C_3 m \log m \tag{68}$$

对于某个常数 $C_3 > 0$，并且假设 5.1 在具有良好条件的协方差 Σ下成立。在此情形下，价值函数实际上是稠密的，并且已知经典 OLS（或稠密贝叶斯）估计量能够达到参数速率

$$\|\hat{\theta}^{\mathrm{DLS}} - \theta^\star\|_2 \;=\; O_{\mathbb{P}}\!\left(\sigma\sqrt{\frac{m}{T}}\right) \tag{69}$$

在 X [32, 33] 满足温和条件时。

通过与定理 5.9 的证明类似的论证（使用引理 5.8 以及不等式 $\|u\|_1 \le \sqrt{m}\|u\|_2$），我们得到以下结果。

命题 6.5 （稠密情形下的 DLS 福利差距）。 假设条件 *5.1*在 $K/m \to \rho \in (0, 1]$ 下成立，且满足（68）。 进一步假设 $X^\top X$ 以高概率可逆。那么存在常数 C_4、C_5、$C_6 > 0$，使得概率至少为 $1 - C_5 m^{-C_6}$，

$$\Delta_{\mathrm{DLS}}(T) \;\le\; C_4\,\sigma\sqrt{\frac{m}{T}}. \tag{70}$$

在稠密情形（68）下，BGCP 和ℓ_1 流程所依赖的稀疏性假设变得不切实际：BGCP 所要求的互相干条件

在 $\theta^\star$ 的许多坐标均非零时，难以满足；而 ℓ_1 估计量会产生普通最小二乘法（OLS）所没有的稀疏性诱导偏差。特别是，当63中的速率按 $K\sqrt{}$（log m）$/T \approx m\sqrt{}$（log m）$/T$缩放时， $K \approx m$与（ $m^{1/2}T^{-1/2}$ 70）中的参数速率相比是次优的。因此，在数据充足的稠密情形下，基于最小二乘法或稠密贝叶斯线性模型的经典稠密概率 CSG 方案可能更可取：它们达到极小极大参数率，且无需付出稀疏性惩罚。

6.5 概率型 CSG 方案的计算复杂度

现在，我们将所提出的算法（BGCP 和 ℓ_1 流程）与 EPC 和 DLS 基线方法进行计算复杂度比较。下文中， T 表示回合数， m 表示候选联盟数， K 表示稀疏度 $|S^\star|$, q 表示每回合活跃联盟数的上限（$\|X_t\|_0 \le q$）。

BGCP。
一种直接的 BGCP 实现会在迭代 k时维护当前支撑集 S_k 和残差 $r_k = Y - X_{S_k}\theta_{S_k}$。每一步中，BGCP 都会使用形如 $X_j^\top r_k$的内积，在候选集合中对所有 j 计算添加/移除评分（例如，添加时考虑所有 $j \notin S_k$，移除时考虑所有 $j \in S_k$）。从头计算所有此类评分需要 $O(Tm)$ 次运算。经过 K 次迭代后，总成本为

$$\mathrm{Time}_{\mathrm{BGCP}} \;=\; O(TmK) \tag{71}$$

忽略低阶项，假设 $K \le K_{\max}$ 预先固定。如果我们维护 $X_{S_k}^\top X_{S_k}$的 Cholesky 分解因子，那么每一步更新后验分布或模型选择准则所需的额外计算量至多为 $O(K^2)$, 当 $Tm \gg K^2$时，这一开销低于 $O(Tm)$的成本。

ℓ_1 流程。
对于 Lasso 估计量 (21)，坐标下降法或近端梯度法等标准求解器的单次迭代开销为 $O(Tm)$，主要由与 X的矩阵-向量乘积构成。令 I_{ℓ_1} 表示达到指定优化精度所需的迭代次数。那么， ℓ_1 阶段的计算开销为

$$\mathrm{Time}_{\ell_1} \;=\; O(TmI_{\ell_1}). \tag{72}$$

在高维情形下， I_{ℓ_1} 通常至多随 $1/\varepsilon$ 呈多对数级增长，以获得 ε-精度的解。计算出 $\hat\theta^{\ell_1}$后，可以将低于阈值的坐标 $|\hat\theta_j^{\ell_1}|$ 剪除，从而在运行联盟结构生成（CSG）之前缩小有效联盟集合。

EPC.
对于 EPC，只需遍历一次数据，即可为每个 j 累积充分统计量 $\sum_{t\in\mathcal{T}_j} Y_t/X_{tj}$ 和 n_j。由于每个回合 t 中最多有 q 个非零条目

X_t，因此形成所有插件估计量（59）所需的操作总数为

$$\mathrm{Time}_{\mathrm{EPC}} \;=\; O(Tq) \;\le\; O(Tm). \tag{73}$$

因此，EPC 在估计方面的计算成本较低，但如命题 6.3所示，它在稀疏高维情形下的福利保证较差。

DLS.
对于 DLS，使用直接方法（例如 Cholesky 分解）求解正规方程 $X^\top X\hat{}\theta^{\mathrm{DLS}} = X^\top Y$，需要构造 $X^\top X$，成本为 $O(Tm^2)$，并以 $O(m^3)$ 的成本对其进行分解，从而得到

$$\mathrm{Time}_{\mathrm{DLS}} \;=\; O(Tm^2 + m^3). \tag{74}$$

共轭梯度法等迭代方法可以将因子从 m^3 降低到 $O(TmI_{\mathrm{DLS}})$，其中 I_{DLS} 为迭代次数，但在最坏情况下，对 m 的依赖仍为二次。

确定性 ***CSG*** 阶段。
所有考虑的方案（BGCP、ℓ_1、EPC、DLS）最终都会在某个联盟集合上调用确定性 CSG 求解器：

- BGCP：在 $\widehat{\mathcal{C}}_{\mathrm{BGCP}} = \{C_j$ 上：$j \in S_{\mathrm{BGCP}}\}$，大小为 $|S_{\mathrm{BGCP}}| \approx K$。
- ℓ_1：可能在一个经过剪枝的集合 $\widehat{\mathcal{C}}^{\ell_1}$ 上，该集合大小为 $\tilde{K}$，通过对 $\hat{}\theta^{\ell_1}$进行阈值化得到。
- EPC、DLS：在完整的联盟集合 $\mathcal{C}$ 上，大小为 m（除非进行额外剪枝）。

基于动态规划或分支定界的精确联盟结构生成算法 [1, 2]，其最坏情况复杂度随参与者数量呈指数增长 $|N|$；我们的框架不会改变这种最坏情况下的依赖关系。不过，通过将˜相关 联盟的数量从 m 减少到约 K（或 K），BGCP 和 ℓ_1 流程在实践中可以显著缩小联盟结构生成求解器的状态空间并减少剪枝工作量。

复杂度权衡概述。
忽略联盟结构生成阶段时，EPC 的估计成本最低（O（Tq）），其次是 BGCP 和 ℓ_1 流程（二者均为 O（Tm），另有迭代因子），然后是 DLS（O（$Tm^2 + m^3$））。将组合型联盟结构生成成本纳入考虑后，情况就有所不同：积极缩减联盟集合的方法（BGCP，以及在较小程度上的 ℓ_1 流程）可以通过缩小联盟结构生成求解器的指数级部分，抵消较高的估计成本。EPC 和 DLS 的拟合成本较低，但会将规模较大的联盟全集传递给联盟结构生成阶段，可能导致总运行时间增加。

6.6 总结：何时使用哪种概率 CSG 方案?

本节的理论与计算比较可概括如下。

• 在稀疏高维情形 $K \ll m$ 和 $T \asymp K \log m$下，我们的概率型联盟结构生成流程在理论上更优。BGCP 以高概率实现精确的福利最优性，而 ℓ_1 流程则具有非平凡的福利差距界，其阶为 $\sigma K\sqrt{(\log\ m)/T}/\kappa_{\Sigma}^2$。在相同情形下，EPC 的福利差距阶为$\sigma\sqrt{K/T}$，且无法利用稀疏性；而 DLS 则不适用，因为 $T < m$。在计算方面，BGCP 和 ℓ_1 流程的估计成本为 $O(Tm)$，但在调用 CSG 求解器之前，可将联盟集合从 m 缩减至约 K 。

• 在稠密数据丰富情形下，当 K 与 m 及 $T \gg m$成比例增长时，DLS 等稠密概率型 CSG 方案（或与 CSG 结合的稠密贝叶斯线性回归）可达到参数速率 $\sigma\sqrt{m/T}$，且无需承担稀疏性惩罚。在这种情形下，BGCP 和 ℓ_1 流程的结构优势基本不复存在，而后者可能会遭受不必要的稀疏性偏差，并需要采取保守的调参方式。由于 T 较大，DLS 在 m 上的二次成本可以摊销；无论如何，组合型 CSG 阶段面对的联盟全集都会被充分利用。

• BGCP 与 ℓ_1 流程之间的选择，本身就体现了组合优化与凸优化之间的权衡。在相干性强且满足最小信号条件时，BGCP 可以实现精确支持集恢复和福利恢复，但它依赖于贪心非凸搜索，成本为 $O(TmK)$。ℓ_1 流程采用凸优化，能够稳健应对稀疏性假设的小幅偏离，但只能近似达到福利最优，其估计成本为 $O(TmI_{\ell_1})$，福利差距则以 $K\sqrt{(\log\ m)/T}$的速率递减。

• EPC 和 DLS 则处于相反的一端：它们要么易于拟合且速度快（EPC），要么在稠密情形下具有统计最优性（DLS），但要么忽略稀疏性，要么需要$T \gtrsim m$。它们会将大型联盟集合交给 CSG 求解器，且无法在稀疏高维情形下提供任何保证。

总体而言，本文所建立的理论为“为什么要通过稀疏回合式模型实现概率联盟结构生成？”这一问题提供了有理有据的答案：当真正有利可图的联盟只有少数几个且数据有限时，利用稀疏性能够带来可证明更优的福利保证，以及比自然的回合式插件方法或稠密最小二乘替代方案更有利的端到端复杂性；而当价值确实是稠密的且回合充足时，经典稠密方案仍然适用。

参考文献

[1] T. Sandholm、K. Larson、M. Andersson、O. Shehory 和 F. Tohm´e，《具有最坏情况保证的联盟结构生成》，人工智能，第 111 卷，第 1–2 期，第 209–238 页，1999 年。[2]T. Rahwan、T. P. Michalak、M. Wooldridge 和 N. R. Jennings，《联盟结构生成：综述》，人工智能，第 229 卷，第 139–174 页，2015 年。[3] T. Rahwan、S. D. Ramchurn、V. D. Dang 和 N. R. Jennings，《近似最优的随时联盟结构生成》，载于第 *20* 届国际人工智能联合会议论文集（*IJCAI-07*），2007 年，第 2365–2371 页。

[4] T. Rahwan、S. D. Ramchurn、N. R. Jennings 和 A. Giovannucci，《用于最优联盟结构生成的随时算法》，人工智能研究杂志，第 34 卷，第 521–567 页，2009 年。[5] T. P. Michalak、J. Sroka、T. Rahwan、M. Wooldridge、P. McBurney 和 N. R. Jennings，《用于随时联盟结构生成的分布式算法》，载于第 *9* 届自主智能体与多智能体系统国际会议论文集（*AAMAS 2010*），2010 年，第 1007–1014 页。[6] B. Banerjee 和 L. Kraemer，《具有混合外部性的多智能体系统中的联盟结构生成》，载于第 *9* 届自主智能体与多智能体系统国际会议论文集（*AAMAS 2010*）。加拿大多伦多：IFAAMAS，2010 年。[7] T. Rahwan、T. Michalak、M. Wooldridge 和 N. R. Jennings，《具有正外部性或负外部性的多智能体系统中的随时联盟结构生成》，人工智能，第 186 卷，第 95–122 页，2012 年。[8] F. Brandt 和 M. Bullinger，《寻找并识别受欢迎的联盟结构》，人工智能研究杂志，第 74 卷，第 571–628 页，2022 年。[9] H. Aziz 和 B. de Keijzer，《联盟结构生成的复杂性》，载于第 *10* 届自主智能体与多智能体系统国际会议论文集（*AAMAS 2011*），台湾台北，2011 年，第 191–198 页。[10] G. Chalkiadakis 和 C. Boutilier，《不确定条件下联盟形成的贝叶斯强化学习》，载于第 *3* 届自主智能体与多智能体系统国际联合会议论文集（*AAMAS 2004*）。IEEE计算机学会，2004 年，第 1090–1097 页。[11]——，《不确定条件下序贯最优的重复联盟形成》，自主智能体与多智能体系统，第 24 卷，第 3 期，第 441–484 页，2012 年。[12] S. Sarkar、M. C. Malta 和 A. Dutta，《多智能体系统中联盟形成应用综述》，并发与计算：实践与经验，第 34 卷，第 11 期，2022 年。[13] R. R˘adulescu、P. Mannion、D. M. Roijers 和 A. Now´e，《多目标多智能体决策：基于效用的分析与综述》，自主智能体与多智能体系统，第 34 卷，第 25 期，2020 年。[14] N. Schwind、T. Okimoto、K. Inoue、K. Hirayama、J. Lagniez 和 P. Marquis，《概率联盟结构生成》，载于第 *16* 届知识表示与推理原理国际会议论文集（*KR 2018*）。IEEE / AAAI出版社，2018 年。[15] K. Matsumura、B. Kodric、T. Okimoto 等，《具有质量界的概率联盟结构生成的两种近似算法》，自主智能体与多智能体系统，第 34 卷，第 25 期，2020 年。[16] N. Schwind、T. Okimoto、K. Inoue、K. Hirayama、J. Lagniez 和 P. Marquis，《概率联盟结构的计算》，自主智能体与多智能体系统，第 35 卷，第 1 期，第 14 页，2021 年。[17] T. Rahwan、T. P. Michalak 和 N. R. Jennings，《用于联盟结构生成的混合算法》，载于第二十六届 *AAAI*人工智能会议论文集（*AAAI-12*），2012 年，第 1443–1449 页。

[18] M. H. Rothkopf、A. Peke˘c 和 R. M. Harstad，“计算上易于处理的组合拍卖”，Management Science，第44卷，第8期，第1131–1147页，1998年。[19] T. Sandholm，“组合拍卖中的获胜者确定方法”，Decision Support Systems，第28卷，第1–2期，第165–176页，2000年。[20] D. L. Donoho，“压缩感知”，IEEE Transactions on Information Theory，第52卷，第4期，第1289–1306页，2006年。[21] E. J. Cand`es 和 T. Tao，“通过线性规划进行解码”，IEEE Transactions on Information Theory，第51卷，第12期，第4203–4215页，2005年。[22] C. Herzet 和 A. Dr´emeau，“贝叶斯追踪算法”，IEEE Transactions on Signal Processing，第62卷，第23期，第6178–6192页，2014年。[23] Y. C. Pati、R. Rezaiifar 和 P. S. Krishnaprasad，“正交匹配追踪：递归函数逼近及其在小波分解中的应用”，载于《第27届阿西洛马会议：信号、系统与计算机》论文集。加利福尼亚州太平洋丛林市：IEEE，1993年，第40–44页。[24] J. A. Tropp 和 A. C. Gilbert，“通过正交匹配追踪从随机测量中恢复信号”，IEEE Transactions on Information Theory，第53卷，第12期，第4655–4666页，2007年。[25]M. A. Davenport 和 M. B. Wakin，“使用受限等距性质分析正交匹配追踪”，IEEE Transactions on Information Theory，第56卷，第9期，第4395–4401页，2010年。[26] T. T. Cai 和 L. Wang，“带噪声的稀疏信号恢复正交匹配追踪”，IEEE Transactions on Information Theory，第57卷，第7期，第4680–4688页，2011年。[27] R. Tibshirani，“通过 Lasso 进行回归收缩与选择”，英国皇家统计学会会刊：B辑（方法论），第58卷，第1期，第267–288页，1996年。 [在线]。网址：https://www.jstor.org/stable/2346178[28] P. J. Bickel、Y. Ritov 和 A. B. Tsybakov，“Lasso 与 Dantzig选择器的同步分析”，Annals of Statistics，第37卷，第4期，第1705–1732页，2009年。[29] P. B¨uhlmann 和 S. van de Geer，《高维数据统计：方法、理论与应用》。柏林、海德堡：施普林格，2011年。[30] P. Zhao 和 B. Yu，“关于 Lasso 的模型选择一致性”，Journal of Machine Learning Research，第7卷，第2541–2563页，2006年。[31] M. J. Wainwright，“使用 ℓ_1-约束二次规划（Lasso）实现高维含噪稀疏性恢复的精确阈值”，IEEE Transactions on Information Theory，第55卷，第5期，第2183–2202页，2009年。[32] T. Hastie、R. Tibshirani 和 J. Friedman，《统计学习要素：数据挖掘、推理与预测》，第2版。纽约：施普林格，2009年。[33] C. M. Bishop，《模式识别与机器学习》。纽约：施普林格，2006年。