arXiv:2302.06485v1 [cs.CC] 13 Feb 2023

差异最小化的稳定算法与在线算法的几何障碍

David Gamarnik* Eren C. Kızıldağ† Will Perkins‡ Changji Xu§

2023年2月14日

摘要

对于许多涉及随机性的计算问题，人们利用解空间复杂的几何特征，严格排除了一大类强大的算法。通常，这通过多重重叠间隙性质（m-重叠间隙性质）的视角来实现；它为具有输入稳定性的算法构成了严格障碍。本文聚焦于两个模型的算法可处理性：（i）差异最小化；（ii）对称二元感知机（SBP），这既是一个随机约束满足问题，也是单层神经网络的玩具模型。

我们首先关注在线算法的局限性。通过建立并利用一种新颖的几何障碍，我们为SBP和差异最小化这两个问题，得到了针对在线算法的尖锐困难性保证。我们的结果与目前已知的最佳算法保证相匹配，误差至多为常数因子。其次，我们关注如何在给定随机矩阵 $\mathcal{M} \in \mathbb{R}^{M\times n}$的情况下，高效地找到一个常数差异解。在光滑情形下，矩阵 $\mathcal{M}$ 的元素独立同分布，且服从标准正态分布；我们证明，对于 m，存在 $n = \Theta(M \log M)$-重叠间隙性质。因此，在这一取值下，我们排除了稳定算法这一类算法。我们的结果首次为Altschuler 和 Niles-Weed 的猜想提供了严格证据 [ANW21，猜想 1]。

我们利用解空间的复杂几何结构，为在线算法证明了紧致的困难性结果。我们所建立的屏障是 m-重叠间隙性质（OGP）的一种新变体。此外，它针对相关实例中的 m-元组解，且 m、 $m = \omega$ 的取值不断增大(1)。重要的是，我们的结果排除了在线算法即使以指数小的概率取得成功的可能性。

*麻省理工学院斯隆管理学院；电子邮件：gamarnik@mit.edu。

†哥伦比亚大学统计学系；电子邮件：eck2170@columbia.edu。

‡佐治亚理工学院计算机科学学院；电子邮件：math@willperkins.org。§哈佛大学数学科学与应用中心；电子邮件：cxu@cmsa.fas.harvard.edu。

目录

1 引言 **3**

1.1 差异最小化.31.2 对称二元感知机（SBP）.41.3 差异理论与 SBP 之间的联系.51.4 主要结果.61.5 背景与先前工作.71.6 未解决问题.9

2 在线算法的紧致困难性保证 **10**

3 寻找常数差异解的算法障碍 **12**

3.1 技术预备知识.123 Ensemble m-差异最小化中的重叠间隙性质.133.3 m-重叠间隙性质蕴含稳定算法的失效.13

4 证明 **14**

4.1 定理 2.3 .154.2 定理 2.4 .184.3 定理 2.5 .214.4 定理 3.2 .22

1 引言

本文研究差异最小化问题和感知机模型。组合差异理论 [Spe85，Mat99] 是组合数学、概率和算法交叉领域的核心课题。给定一个矩阵 $\mathcal{M} \in \mathbb{R}^{M\times n}$，差异理论的核心任务是计算或界定以下量

$$\mathcal{D}(\mathcal{M}) \triangleq \min_{\boldsymbol{\sigma}\in\Sigma_n} \left\|\mathcal{M}\boldsymbol{\sigma}\right\|_\infty,$$

称为 $\mathcal{M}$的差异度。

感知机是一种用于存储随机模式的玩具单层神经网络模型，同时也是一种非常自然的高维概率模型；关于该模型的早期研究，参见[JH60、Win61、Wen62、Cov65] 。给定随机模式 $X_i \in \mathbb{R}^n$、 $1 \le i \le M$，若能找到一个与所有 X_i都“相容”的 $\boldsymbol{\sigma} \in \mathbb{R}^n$，则可实现存储： $\langle \boldsymbol{\sigma}$、$X_i\rangle \ge 0$ 对于 $1 \le i \le M$成立。向量 $\boldsymbol{\sigma}$ 被解释为突触权重；它可以位于 $\mathbb{R}^n$中的球面上，$\|\boldsymbol{\sigma}\|_2 = \sqrt{n}$，也可以具有二元条目，$\boldsymbol{\sigma} \in \Sigma_n = \{-1, 1\}^n$。前者被称为球形感知机，相关研究参见[Gar88、ST03、Sto13、Tal11、AS20] 。本文仅关注后者，即二元感知机。感知机文献中研究的一个基本对象是存储容量：使用合适的 $\boldsymbol{\sigma}$时能够存储的（随机）模式的最大数量，参见[Gar87、Gar88、GD88]。Krauth 和 M´ezard [KM89] 给出了关于存储容量的详细但非严格刻画。近年来，人们研究了带有激活函数 $U: \mathbb{R} \to \{0, 1\}$ 的感知机模型，其中，如果模式 X_i 相对于 U 满足$U(\langle \boldsymbol{\sigma}$、 $X_i\rangle) = 1$，则称其被存储。我们尤其关注激活函数 $U(x) = \mathbf{1}_{|x|\le\kappa\sqrt{n}}$，它定义了由 Aubin、Perkins 和 Zdeborov´a [APZ19]提出的对称二元感知机（SBP）模型（关于更一般的感知机模型的结果，另见 [BNSX21、NS23] ）。正如下面将看到的，SBP 与差异最小化密切相关。

1.1 差异最小化

差异文献涵盖最坏情况和平均情况下的 $\mathcal{M}$。在最坏情况下，对 $\mathcal{M}$的结构不作任何假设；而在平均情况下， $\mathcal{M}$ 的各项是随机的，例如独立同分布的伯努利变量、拉ademacher 变量或标准正态变量。此外，差异理论既寻求存在性结果，也寻求算法方面的结果。

关于最坏情况分析，该领域的一个里程碑式结果由 Spencer 提出[Spe85]: $\mathcal{D}(\mathcal{M}) \le 6\sqrt{n}$ 如果 $\mathcal{M} \in \mathbb{R}^{n\times n}$ 满足 $|M_{ij}| \le 1$ 对 $1 \le i, j \le n$ (“六个标准差就足够了”)。这一结果的重要意义在于，它改进了基本概率方法所保证的差异度：随机赋予符号所产生的差~~异度为~~ $\Theta(\sqrt{n}$ $\log n$阶，显著大于 $O(\sqrt{n})$。值得注意的是，Spencer 的结果是最坏情况且非构造性的，但近期的研究 [Ban10、LM15、LRR17、Rot17] 已提出高效算法来寻找此类低差异解。

本文聚焦于平均情形差异度。假设 $\mathcal{M} \in \mathbb{R}^{M\times n}$ 的元素为独立同分布的 $\mathcal{N}(0, 1)$ 随机变量，且 $M = o(n)$。在这种情况下，始于 [KKLO86] （针对 $M = 1$）并随后由 [Cos09、TMR20] （针对 $M \ge 2$）继续推进的一系列研究证明了 $\mathcal{D}(\mathcal{M}) = \Theta(\sqrt{n}2^{-n/M})$ 以高概率成立。关于这一情形下的算法结果，见 [KK82、Yak96、TMR20]。在 $M = 1$ 的特殊情形下， [GK21a] 给出了有力证据，表明寻找一个满足 $\boldsymbol{\sigma}$ 且具有 $\|\mathcal{M}\boldsymbol{\sigma}\|_\infty = 2^{-\omega(\sqrt{n\log n})}$ 的算法

在算法上可能难以处理。另一方面，当 $M = \Theta(n)$ 时，结果表明$\mathcal{D}(\mathcal{M}) = \Theta(\sqrt{n})$，而这一情形与 SBP密切相关，详见第 1.3 节。

接下来，假设 $\mathcal{M}$ 的元素是独立同分布的二元变量，例如拉ademacher变量或伯努利(p)变量。在这种情况下，当 $M = \Theta(n)$ 时，尽管 w.h.p. 仍有 $\mathcal{D}(\mathcal{M}) = \Theta(\sqrt{n})$，但事实证明，当 n 远大于 M 时，可以达到常数差异度，实际上可以得到 $\mathcal{D}(\mathcal{M}) = 1$。这方面最精确的结果由 Altschuler 和 Niles-Weed 给出 [ANW21]，他们彻底解决了究竟何时 $\mathcal{D}(\mathcal{M}) \le 1$ 的问题：若 $\mathcal{M}$ 由伯努利(p)变量构成，且 p 任意、 $n \ge CM\log M$，其中 C 是任意大于 $(2\log 2)^{-1}$ 的常数，则 $\mathcal{D}(\mathcal{M}) \le 1$ 。他们的结果尤其涵盖稀疏区域，即 $p = o(1)$，并且是可能达到的最精确结果，因为要使 $\mathcal{D}(\mathcal{M})$ 达到 $O(1)$，必须有 $n = \Omega(M\log M)$：若对于 $C < (2\log 2)^{-1}$ 和 $p = 1/2$，有 $n = CM\log M$，那么 w.h.p. 不存在常数差异度解。关于这一方向的早期结果见 [HR19、FS20、Pot18]。有了 [ANW21]，给出的这一存在性保证，一个自然的算法问题是：能否在多项式时间内找到这样的常数差异度解？据推测，这一任务很难，参见 [ANW21，猜想 1]。本文主要关注该问题在算法上的可处理性。

1.2 对称二元感知机（SBP）

固定 $\kappa > 0$、 $\alpha > 0$，并设 $M = \lfloor n\alpha \rfloor \in \mathbb{N}$。令 $X_i \sim \mathcal{N}(0, I_n)$、 $1 \le i \le M$ 为独立同分布的随机向量，其中 $\mathcal{N}(0, I_n)$ 是 $\mathbb{R}^n$ 中具有单位协方差的中心多元正态分布。考虑以下（随机）集合

$$S_\alpha(\kappa) = \Big\{\boldsymbol{\sigma} \in \Sigma_n : |\langle \boldsymbol{\sigma}, X_i\rangle| \le \kappa\sqrt{n}, 1 \le i \le M\Big\} = \Big\{\boldsymbol{\sigma} \in \Sigma_n : \big\|\mathcal{M}\boldsymbol{\sigma}\big\|_\infty \le \kappa\sqrt{n}\Big\}, \tag{1}$$

其中 $\mathcal{M} \in \mathbb{R}^{M\times n}$ 的行是 X_1、……、 X_M。“对称”一词指的是： $\boldsymbol{\sigma} \in S_\alpha(\kappa)$ 当且仅当 $-\boldsymbol{\sigma} \in S_\alpha(\kappa)$。该 SBP 由 Aubin、Perkins 和 Zdeborová [APZ19] 提出，作为非对称二元感知机（ABP）的对称对应物；在后者中，约束的形式则是 $\langle \boldsymbol{\sigma}, X\rangle \ge \kappa\sqrt{n}$、 $1 \le i \le M$。事实证明， ABP 在数学上极具挑战性；相关工作和更多细节可参见 [KM89、KR98、Tal99、Xu19、DS19、PX21、ALS21a、GKPX22、Kız22] 。另一方面， SBP 保留了 ABP [BDVLZ20], 所猜想的相关结构性质，同时由于对称性，更易于进行严格分析。

该 SBP 经历了尖锐的相变，这一相变由 [APZ19]提出猜想，随后由 Perkins 和 Xu [PX21] 以及 Abbe、Li 和 Sly [ALS21b]证明。令

$$\alpha_c(\kappa) \triangleq -\frac{1}{\log_2 \mathbb{P}[|Z| \le \kappa]}, \quad \text{where} \quad Z \sim \mathcal{N}(0,1). \tag{2}$$

Then

$$\lim_{n\to\infty} \mathbb{P}\big[S_\alpha(\kappa) \ne \varnothing\big] = \begin{cases} 0, & \text{if } \alpha > \alpha_c(\kappa) \\ 1, & \text{if } \alpha < \alpha_c(\kappa) \end{cases}. \tag{3}$$

部分 $\alpha > \alpha_c(\kappa)$ 由 [APZ19] 得出，并通过应用一阶矩方法加以证明：若 $\alpha > \alpha_c(\kappa)$，则 $\mathbb{E}\big[|S_\alpha(\kappa)|\big] = o(1)$，因此根据马尔可夫不等式， $S_\alpha(\kappa) = \varnothing$ 以高概率成立。同一篇论文还研究了 $\alpha < \alpha_c(\kappa)$ 的情形，并通过二阶矩方法证明了 $\liminf_{n\to\infty} \mathbb{P}\big[S_\alpha(\kappa) \ne \varnothing\big] > 0$。要将其提升为高概率保证，则需要更强大的工具，详见 [PX21] 中精妙的鞅论证，以及 [ALS21b] 中的论证

基于小子图条件化方法的全连接类比。此外，最近有人证明，围绕 $\alpha_c(\kappa)$ 的临界窗口宽度为常数，参见 [Alt22、SS23]。这些事实表明，一阶矩对相变精确位置的“预测”是正确的，而且相变非常尖锐。

鉴于当 $\alpha < \alpha_c(\kappa)$ 时， $S_\alpha(\kappa)$ 以高概率非空，一个自然的后续问题便是算法方面的：能否高效地找到一个解 $\boldsymbol{\sigma} \in S_\alpha(\kappa)$? 高效算法的存在是否取决于 α? 对于 ABP 和 SBP，分别有研究在低密度下提出了高效算法： α 见于 [KR98] 和 [ALS21a] 。另一方面， [GKPX22] 研究了高效算法的极限（详见下文）。研究成果[BBBZ07、BIL+15、BDVLZ20]提出了对高效算法成功的可能解释：尽管几乎所有解都完全冻结（这一猜想见于[MMZ05、HK14]），高效算法却能找到位于大簇中的稀有解。包括 [PX21、ALS21b、ALS21a] 在内的近期研究探讨了这些结构性预测。

1.3 差异理论与 SBP之间的联系

为了阐明差异最小化与 SBP之间的联系，我们聚焦于比例区间，即 $M = \Theta(n)$。差异度视角是固定纵横比为 $\alpha = M/n$的 $\mathcal{M} \in \mathbb{R}^{M\times n}$，并寻求一个 $\boldsymbol{\sigma}$ ，使 $\|\mathcal{M}\boldsymbol{\sigma}\|_\infty$ 尽可能小。另一方面，感知机视角则相反：先固定一个 $\kappa > 0$ ，再寻找最大的 α ，使得存在一个满足 $\|\mathcal{M}\boldsymbol{\sigma}\|_\infty \le \kappa\sqrt{n}$ 的解 $\boldsymbol{\sigma}$ 。此外，比例区间内平均情形差异度的渐近值可直接由 SBP (3)的尖锐阈值结果得出： $\mathcal{D}(\mathcal{M}) = (1 + o(1))f(\alpha)\sqrt{n}$ 以高概率成立，其中 $f(\alpha)$是 $\alpha_c(\kappa)$（2）的“逆”。

算法上的联系 SBP 与差异理论之间的联系还进一步延伸到了算法领域：目前已知的针对 SBP 的最佳高效算法源自差异文献。假设 $\mathcal{M} \in \mathbb{R}^{M\times n}$ 具有独立同分布的 Rademacher 条目。Bansal 和 Spencer [BS20] 设计了一种高效的在线算法，该算法能够找到一个 $\boldsymbol{\sigma}_{\mathrm{ALG}} \in \Sigma_n$ ，使得$\|\mathcal{M}\boldsymbol{\sigma}_{\mathrm{ALG}}\|_\infty = O(\sqrt{M})$；当 $n \ge M = \omega(1)$时，该结论以高概率成立。直观地说，如果输出的 t^{th} 坐标 $\boldsymbol{\sigma}_{\mathrm{ALG}}$ 仅取决于 $\mathcal{M}$的前 t 列，那么该算法就是在线的；形式化定义见定义 2.1。作为直接推论，这给出了一种针对 SBP 的高效算法，该算法能够找到一个解 $\boldsymbol{\sigma} \in S_\alpha(\kappa)$；当 $\alpha = O(\kappa^2)$时，该结论以高概率成立；参见[GKPX22，推论 3.6]。事实上，这是目前已知的关于 SBP 以及随机比例区间中的差异度的最佳算法保证；参见 [GKPX22，第 3.3]节。

鉴于这些存在性和算法方面的结果， SBP 似乎可能存在显著的统计到计算的差距（SCG）：以高概率存在解的密度下限，即α_c（κ），远高于多项式时间搜索算法能够奏效的密度。此外，当 $\kappa \to 0$时，这种 SCG 最为显著。Bansal-Spencer 算法仅在 $\alpha = O$（κ^2）时有效，而在 α_c（κ）以下，以高概率确实存在解；根据（2），该值渐近地为^1 $\log_2$（$1/\kappa$）。关于这种 SCG 的起源，~~[GKPX22]~~, 进行了研究，证明当 $\alpha = \Omega(\kappa^2 \log_2 \frac{1}{\kappa})$时，SBP 具有一种称为多重重叠间隙性质（$m$-OGP）的复杂几何性质，因此，当$\alpha = \Omega(\kappa^2 \log_2 \frac{1}{\kappa})$时，稳定算法无法找到满足解。不过值得注意的是，稳定算法未必包含在线算法；对于 SBP，在线算法能够达到计算阈值。在线算法的能力边界仍是一个开放问题，也是本文要探讨的问题。

除了 SBP之外，差异最小化问题——尤其是当以高概率存在常数差异解时，高效寻找此类解的算法问题——也呈现出类似的 SCG。回顾一下，当 $\mathcal{M} \in \mathbb{R}^{M\times n}$时，只要 $n = \Omega(M \log M)$，就以高概率存在常数差异解。另一方面，目前已知的最佳多项式时间算法只能在小得多的 $M = o(\log n)$ [Ban22]，值上奏效，凸显了另一种显著的 SCG。这是本文的第二个研究重点。

1.4 主要结果

假设 $\mathcal{M}$ 每个（1）由独立同分布的 $\mathcal{N}(0, 1)$ 条目组成。我们的首个主要结果表明，在密度 SBP 达到 $\alpha = \Omega(\kappa^2)$ 时，在线算法无法找到一个满足解。

定理 1.1 （非正式表述，详见定理 2.3）。 对于密度 $\alpha = \Omega(\kappa^2)$，在线算法无法为 *SBP* 找到解，且该事件发生的概率大于 $e^{-\Theta(n)}$。

我们的下一个结果将定理 1.1 扩展到差异最小化问题，其中 $\mathcal{M} \in \mathbb{R}^{M\times n}$ 的条目为独立同分布的拉ademacher分布或独立同分布的伯努利分布（p）。

定理 1.2 （非正式表述，详见定理 2.4-2.5）。 存在 $c > 0$，使得在线算法无法返回差异度至多为 $c\sqrt{M}$ 的差异解，且该事件发生的概率大于 $e^{-\Theta(M)}$。

如果 $\mathcal{M}$ 的各项服从 Rademacher 分布，取 $c = 1/24$ 即可。对于伯努利情形，由此得到的常数取决于 p：取 c、 $c_p = \sqrt{p - p^2}/24$ 即可。综合定理 1.1 和定理 1.2 可知，在在线算法这一类算法中，Bansal-Spencer 算法 [BS20] 对于这两种模型在常数因子意义下都是最优的。我们的证明基于一种新型的 m-重叠间隙性质（OGP）：对于一组 m 个相关实例，给定适当的 $\Delta \in \in(0, 1)$，我们证明不存在在前 $1 - \Delta$ 比例坐标上相同的解元组，详情见下文。这一障碍比 m-重叠间隙性质更严格；后者断言不存在彼此处于预设距离的解元组。此外，对于定理 1.2，还必须考虑 m 元组，其中 m、 $m = \omega(1)$ 的值不断增长；这一思想最初由 Gamarnik 和 Kızılda˘g [GK21a] 提出，用于降低 m-重叠间隙性质的阈值。

据我们所知，定理 1.1-1.2 是通过几何障碍，对稳定算法之外的算法类别给出（在常数因子范围内）紧致困难性保证的首例，详情见第 1.5节。此外，与先前工作不同 [GJW20、Wei20、HS21、GK21a、GKPX22]，被排除的算法不必以高概率成功，甚至不必以常数概率成功：只要成功概率呈指数级小即可。之所以能够做到这一点，是因为巧妙地运用了詹森不等式，该方法最初由 Gamarnik 和 Sudan 提出 [GS17b]。

证明概要 我们概述定理 1.1 的证明。该证明基于一种新的 m- 重叠间隙性质，并结合反证法。假设存在这样一个在线算法 $\mathcal{A}$，其成功概率为 p_s。令 $\mathcal{M}_1 \in \mathbb{R}^{M\times n}$ 的各项为独立同分布的 $\mathcal{N}(0, 1)$ 随机变量。固定一个 $m \in \mathbb{N}$ 和一个 $\Delta \in (0, 1)$，通过对 $\mathcal{M}_1$ 的最后 Δn 列进行独立重新采样，生成随机矩阵 $\mathcal{M}_i \in \mathbb{R}^{M\times n}$、 $2 \le i \le m$。分别在每个 $\mathcal{M}_i$ 上运行 $\mathcal{A}$，得到解 $\boldsymbol{\sigma}_i$、 $\mathcal{A}(\mathcal{M}_i) \in \Sigma_n$、$1 \le i \le m$。应用詹森不等式可知，$\|\mathcal{M}_i\boldsymbol{\sigma}_i\|_\infty \le \kappa\sqrt{n}$、 $1 \le i \le m$ 以至少 p_s^m 的概率成立。此外，由于 $\mathcal{A}$ 是在线算法，任意 $\boldsymbol{\sigma}_i$ 和 $\boldsymbol{\sigma}_j$ 的前 $n - \Delta n$ 个坐标必然相同。也就是说，如果存在这样的 $\mathcal{A}$，那么以至少 p_s^m 的概率，存在一个由 m 个满足约束的解组成的元组（$\boldsymbol{\sigma}_1$，...，$\boldsymbol{\sigma}_m$），它们的前 $n - \Delta n$ 个坐标相同。

接下来，我们运用 一阶矩方法证明，对于适当选取的 m， Δ，存在这样的 m 元组的概率实际上严格小于 p_s^m。这就构成了矛盾。定理 1.2 的证明与此类似，但需要额外的技术步骤。特别是，需要通过 Berry-Esseen 定理，对二元变量的带符号和进行反集中论证。

我们接下来关注这样一个算法问题：给定随机的 $\mathcal{M} \in \mathbb{R}^{M\times n}$，如何高效地找到常数差异解。回顾一下，只要 $n = \Omega(M \log M)$ [ANW21]，，这类解就以高概率存在；而目前已知的最佳多项式时间算法仅在 $M = o(\log n)$ [Ban22] 时有效。此外， [ANW21] 猜想这一任务在算法上是困难的。为探究这一猜想，我们聚焦于光滑情形，其中 $\mathcal{M}$ 的元素是独立同分布的 $\mathcal{N}(0, 1)$。我们的下一个主要结果表明，当 $n = \Theta(M \log M)$ 时，存在具有 $m = O(1)$ 的 m-重叠间隙性质，从而为“边界” $n = \Theta(M \log M)$ 处的困难性提供了严格证据。

定理 1.3 (非正式表述，参见定理 3.2)。 对于 $n = \Theta(M \log M)$，常数差异解集合具有 m-重叠间隙性质（常数为 m），其中参数经过适当选取。

对数 $n \ll M \ll n/$ 对数 n 的情形，以及超越高斯无序的扩展——尤其是伯努利或拉德马赫情形——都是我们在第 1.6 节讨论的开放问题。

我们的最后一个主要结果利用 m-重叠间隙性质证明，稳定算法无法找到常数差异解，当 $n = \Theta(M$ 对数 $M)$。非正式地说，如果对输入进行微小扰动只会导致输出发生微小变化，那么该算法就是稳定的 $\boldsymbol{\sigma}$；形式化表述见定义 3.3。已有研究表明，稳定算法这一类涵盖了强大的算法类别，包括低次多项式 [GJW20、BH21], 近似消息传递（AMP） [GJ21], 以及低深度布尔电路 [GJW21]。

定理 1.4 （非正式表述，定理 3.4）。 对于 $n = \Theta(M$ 对数 $M)$，稳定算法以大于某个常数的概率无法找到常数差异解。

定理 1.4 的证明基于 [GK21a]中提出并且 [GKPX22] 也采用的 m-重叠间隙性质结果（定理 3.2），它排除了稳定算法以常数概率成功的可能性。

1.5 背景与先前工作

统计—计算间隙（SCGs） SBP 和差异最小化都存在 SCG：已知的高效算法表现严格弱于存在性保证。这种间隙是许多源自随机组合结构和高维统计推断的平均情形算法问题的普遍特征。具有 SCG 的一部分问题包括随机约束满足问题（[MMZ05, ART06, ACO08, GS17b, BH21], ）、随机图上的优化（ [GS14, COE15, Wei20], ）、自旋玻璃（ [GJ21, HS21], ）、植入团（ [DM15, BHK+19]）以及张量分解（ [Wei22], ）；另请参阅 Gamarnik [Gam21] 以及 Gamarnik、Moore 和 Zdeborov´a[GMZ22]撰写的综述。遗憾的是，由于这类问题具有平均情况性质，标准计算复杂性理论往往无济于事[1]。尽管如此，一条极有前景的研究路线提出了各种框架，为问题的困难性提供严格证据。这些框架包括平均情形归约——通常从植入团问题出发（[BR13, BBH18, BB19]）——以及针对受限算法类的无条件下界，

[1]除了少数例外情况，例如 [Ajt96、BABB21、GK21b]。

包括统计查询算法（ [DKS17, FGR+17、FPV18], ）、低次多项式（ [Hop18, KWB22, Wei22], ）、平方和层次（ [HSS15, HKP+17, RSS18, BHK+19]）、AMP（ [ZK16, BPW18], ）以及蒙特卡洛马尔可夫链（MCMC）方法（ [Jer92, DFJ02]）。另一种此类方法则基于解空间复杂而精细的几何结构，即重叠间隙性质。

复杂几何与重叠间隙性质（OGP）先前工作[MMZ05、ART06、ACO08] 发现，在随机约束满足问题的背景下，复杂几何与算法难度之间存在一种非常引人入胜的联系：算法难度的出现，大致与解空间中复杂几何结构的形成同时发生。OGP框架借鉴统计物理学的见解，严格地将复杂几何结构与形式化的难度联系起来。在随机优化的背景下，OGP的非正式表述是（关于随机性，以高概率）任意两个近最优解要么“接近”，要么“相距甚远”：存在 $0 < \nu_1 < \nu_2 < 1$ 使得$n^{-1}\langle \boldsymbol{\sigma},\ \boldsymbol{\sigma}'\rangle \in [0,\nu_1] \cup [\nu_2, 1]$ 对于任意一对近最优解 $\boldsymbol{\sigma},\ \boldsymbol{\sigma}' \in \Sigma_n$。也就是说，归一化重叠的区域在拓扑上不连通；不存在彼此处于中间距离的近最优解对。对于具有输入稳定性的大类算法，OGP构成一道严格的屏障——见下文。关于OGP的综述，请参见[Gam21]。

OGP 的算法含义 最早建立并利用 OGP 来排除算法的工作出自 Gamarnik 和 Sudan [GS14, GS17a]。他们关注的是在平均度为 d 的稀疏随机图中寻找大型独立集的问题；这类图有一个 SCG：最大独立集的大小为 $2\frac{\log d}{d}n$ [F L92, BGT10],，而目前已知的最佳高效算法只能找到大小为 $\log_d n$ 的独立集。他们证明，任意一对大小超过 $(1+1/\sqrt{2})\frac{\log d}{d}n$ 的独立集都会呈现 OGP。随后，他们利用 OGP 证明，局部算法无法找到大小超过 $(1+1/\sqrt{2})\frac{\log d}{d}n$ 的独立集。后续研究则在其他各种模型中（例如随机图[RV17, GJW20, Wei20], 、自旋玻璃模型 [C GPR19, GJ21, HS21], 、随机约束满足问题 [GS17b, BH21, GKPX22]），建立并利用 OGP 来排除其他类别的算法（例如 AMP [GJ21], 、低次多项式 [GJW20, Wei20, BH21], 、朗之万动力学 [GJW20, HS21], 、低深度电路 [GJW21]）。在 OGP 和其他复杂几何版本所排除的算法中，一个非常重要的共同特征是输入稳定性，与定义 3.3 类似（但植入模型中 MCMC 失效的情况除外，参见例如 [Jer92, AWZ20, GJS21]）。我们的工作首次表明，复杂几何能够对稳定算法之外的算法类别给出紧的算法难度界。

多重 OGP（m-OGP）前文讨论的工作 [GS14、GS17a] 证明了局部算法在取值（$1+1/\sqrt{2}$）$\frac{\log d}{d}n$时的失效。Rahman 和 Vir´ag [RV17] 随后通过考察涉及许多大型独立集的某种重叠模式，去除了额外的$1/\sqrt{2}$ 项；他们表明，OGP 的出现恰好与算法的对数$_d n$值相吻合。也就是说，通过更复杂的重叠模式，有可能降低 OGP 的出现阈值，并在更宽的参数范围内排除算法。类似地，Gamarnik 和 Sudan [GS17b] 研究了非全相等 k-SAT 问题，并证明对于近乎等距的满足赋值的 m-元组，存在 OGP。因此，他们针对序贯局部算法获得了近乎紧致的困难性保证。在 [GK21a、GKPX22], 中也采用了类似的 m-OGP，这同样是本文的研究重点。

最近，有人提出了针对更复杂模式的 m-OGP。这些禁忌模式涉及解的 m-元组，其中对于任意 $2 \le i \le m$， i^{th} 个解与前 $i-1$ 个解具有“中间”重叠。借此，Wein [Wei20] 针对在稀疏随机图中寻找独立集的问题，以及 Bresler 和 Huang [BH21]针对随机 k-SAT 问题，获得了针对低次多项式的紧致困难性保证。类似地，Huang 和 Sellke [HS21] 构造了一种非常复杂的禁忌结构，由解组成的超度量树，并将其称为分支重叠间隙性质。借助分支重叠间隙性质，他们针对 p-自旋模型获得了针对 Lipschitz 算法的紧致困难性保证。此外，这些论文还确立了系综m-OGP，该性质涉及相对于相关实例近最优的 m-元组。系综 OGP 首次出现在[CGPR19]中；此后，它一直被用于排除稳定算法。研究相对于相关实例的解的 m-元组是本文的核心内容。此外，我们考察了具有超常数 m、 $m = \omega(1)$ 的 m-元组，以便在定理 2.4-2.5 中排除在线算法。这一思想源自 [GK21a] ，旨在进一步降低 m-OGP 阈值。

1.6 未解决问题

其他类别算法的几何障碍 关于重叠间隙性质的先前研究表明，复杂的几何结构是算法难度的标志，并给出了针对稳定算法的下界。定理 2.3、2.4和2.5 将这一结论扩展到稳定算法之外；它们利用复杂的几何结构排除在线算法。若能通过类似的几何障碍排除其他类别的算法，将会非常有意思。

超越高斯无序的差异最小化 定理 3.2 表明，对于 $\mathcal{M} \in \mathbb{R}^{M\times n}$，其元素独立同分布于 $\mathcal{N}(0,1)$，且 $n = \Theta(M \log M)$，常数差异解的集合具有 m-重叠间隙性质。一个非常有意思的问题是，当$\mathcal{M}$ 的元素为二进制时， m-重叠间隙性质是否仍然成立。先前研究已在离散无序模型（例如随机k-SAT [GS17b、BH21], 随机图 [GS14、GS17a、GJW20、Wei20]）以及连续无序模型（例如自旋玻璃模型 [GJ21、HS21], 数划分 [GK21a],theSBP [GKPX22]）中证明了重叠间隙性质。这些结果表明，重叠间隙性质具有普适性：无序的分布细节并不重要[2]。鉴于此，我们提出以下猜想：

猜想 1.5. 对于具有独立同分布的拉ademacher或伯努利(p)条目的 $\mathcal{M} \in \mathbb{R}^{M\times n}$ ，且 $n = \Theta(M \log M)$，常数差异解的集合在适当参数下呈现 m-重叠间隙性质。

解决猜想 1.5 可能需要理解如下形式的概率项 $\mathbb{P}[M\boldsymbol{v} = \boldsymbol{x}]$，其中随机对象为 $\boldsymbol{v}$，其二元条目独立同分布，且确定性对象 $M \in \{-1,1\}^{m\times n}$ 的各行$\boldsymbol{\sigma}_1$、……、$\boldsymbol{\sigma}_m$ 满足 $n^{-1}\langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j\rangle = \beta$ 对于某个固定的 β 以及每个 $1 \le i < j \le m$。一种思路是采用局部极限定理论证；我们将此作为一个开放问题，留待未来研究。

超越 $n = \Theta(M \log M)$ 的差异最小化问题 回顾一下，只要 $n = \Theta(M \log M)$，即当 $M = O(n/\log n)$ 时，就存在常数差异解，而目前已知的最佳多项式时间算法仅在 $M = o(\log n)$ 时有效。鉴于此，定理 3.2 为困难性提供了严格证据，但仅限于“边界”处。$\log n \ll M \ll n/\log n$ 这一范围是未来值得研究的一个有趣方向。一种可能的途径是考虑

[2]事实上，这样的普适性结果已针对 SBP得到证明，见 [GKPX22，定理 5.2].

更复杂的重叠模式，例如 [Wei20、HS21]中的模式，或 分支重叠间隙性质 [HS21]。为此，我们发现了一个有趣的相变（证明略）：

定理 1.6。设 $\mathcal{M} \in \mathbb{R}^{M\times n}$ 的各项为独立同分布的 $\mathcal{N}(0,1)$。固定一个 $K>0$，并令 $S(m,\delta,K)$ 表示由满足 $d_H(\boldsymbol{\sigma}_i,\boldsymbol{\sigma}_j)=\delta$ 且 $\max_{i\le m}\|\mathcal{M}\boldsymbol{\sigma}_i\|_\infty \le K$ 的 $(\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_m)$ 组成的集合。

(a) 如果 $M=\omega(\log n)$，那么 $\mathbb{E}\big[\big|S\big(n\log^{-O(1)}n、\log^{O(1)}n,K\big)\big|\big]=o(1)$。

(b) 如果 $M=o(\log n)$，那么 $\mathbb{E}\big[\big|S\big(n\log^{-O(1)}n、\log^{O(1)}n,K\big)\big|\big]=\omega(1)$。

也就是说，$M=\log n$ 是一个阈值：在该值处，距离为 $\log^{O(1)}n$、具有 $m=\widetilde{\Theta}(n)$ 的常数差异解的 m 元组的（期望）数量会发生相变。值 $\log n$ 处的这一相变究竟只是巧合，还是算法难度的标志，仍是留待未来研究的开放问题。

论文结构与记号 本文其余部分安排如下。第 2 节将形式化定义在线算法类，并陈述困难性结果。第 3 节专门讨论寻找常数差异解的算法问题。关于预备知识以及与重叠间隙性质（OGP）相关的 m 元组集合的定义，参见第 3.1 节；关于主要的m-OGP 结果，参见第 3.3 节，其中给出了稳定算法的困难性结果。最后，完整证明见第 4 节。我们的记号基本遵循惯例，详情见第 4 节开头。

2 在线算法的紧致困难性保证

本节将探讨在我们的两种模型——SBP 和平均情形差异度——中在线算法的局限性。我们首先形式化这两种模型下在线算法的类别。

定义 2.1。

- （SBP）固定一个 $\kappa>0$和一个 $\alpha<\alpha_c(\kappa)$。令 $\mathcal{M}\in\mathbb{R}^{\alpha n\times n}$的元素为独立同分布的$\mathcal{N}(0,1)$。
- （差异度）固定 $n\ge M$，令 $\mathcal{M}\in\mathbb{R}^{M\times n}$的元素为独立同分布的拉ademacher或伯努利(p)变量。

固定一个 $p_s>0$ 和一个 $K>0$。若算法 $\mathcal{A}$ 满足以下条件，则称其对于 *SBP* 是(p_s,α)-在线的，或对于差异度是(p_s,K)在线的。令 $\mathcal{A}(\mathcal{M})$ $\boldsymbol{\sigma}_{\mathrm{ALG}}=(\boldsymbol{\sigma}_{\mathrm{ALG}}(i):\ 1\le i\le n)\in\Sigma_n$。

- (成功) 我们有

$$\mathbb{P}\big[\big\|\mathcal{M}\boldsymbol{\sigma}_{\mathrm{ALG}}\big\|_\infty\le\kappa\sqrt{n}\big]\ge p_s\quad \textit{(for the SBP)}$$
$$\mathbb{P}\big[\big\|\mathcal{M}\boldsymbol{\sigma}_{\mathrm{ALG}}\big\|_\infty\le K\big]\ge p_s\quad \textit{(for discrepancy)}.$$

- （在线）设 $\mathcal{C}_1$，……，$\mathcal{C}_n$是 $\mathcal{M}$的列。存在确定性函数 f_1，……，f_n，使得对于 $1\le t\le n$，$\boldsymbol{\sigma}_{\mathrm{ALG}}(t)=f_t(\mathcal{C}_1,\ \dots\dots,\ \mathcal{C}_t)\in\{-1,1\}$。

参数 p_s 是算法的成功保证，其中概率是相对于 $\mathcal{M}$中的随机性而言的。该算法的在线特性可作如下解释。列 $\mathcal{C}_i$ 依次到达。在第 $t-1$轮结束时，分配符号 $\boldsymbol{\sigma}(i) \in \{-1,1\}$，$1 \le i \le t-1$，随后一列 $\mathcal{C}_t$ 到达。符号 $\boldsymbol{\sigma}(t)$仅取决于先前的决策 $\boldsymbol{\sigma}(i)$、$1 \le i \le t-1$ 和 $\mathcal{C}_t$。也就是说，$\boldsymbol{\sigma}(t)$仅取决于 $\mathcal{C}_i$、$1 \le i \le t$。这种抽象尤其涵盖了 Bansal-Spencer 算法：

定理 2.2。*[BS20，定理 3。4]*设 $n \ge M$和 $\mathcal{M} \in \{-1,1\}^{M\times n}$ 具有独立同分布的拉ademacher条目。则存在绝对常数 $C>0$ 和 $\gamma<1$，以及一个在线算法 $\mathcal{A}$，该算法以 $\mathcal{M}$ 作为输入，并返回一个 $\boldsymbol{\sigma}$，$\mathcal{A}(\mathcal{M})$，使得

$$\mathbb{P}\left[\left\|\mathcal{M}\boldsymbol{\sigma}\right\|_\infty \le C\sqrt{M}\right] \le 1-e^{-\Theta(M^\gamma)}.$$

定理 2.2 立即给出了一个高效算法，用于 SBP 当 $\alpha \le \kappa^2/C^2$时，参见 [GKPX22，推论 4。6]。正如引言中所述，Bansal-Spencer 算法是目前已知的多项式时间算法中，在 SBP 以及随机成比例情形下的差异最小化方面表现最佳的算法。按照定义 2.1，它是 $1-e^{-\Theta(n^\gamma)}\kappa^2/C^2$)-在线算法，用于 SBP；同时也是 $1-e^{-\Theta(M^\gamma)}C\sqrt{M}$)-在线算法，用于差异度。

在线算法 SBP 我们的首个主要结果聚焦于 SBP 在 $\kappa \to 0$ 这一范围内的情形，并证明了在线算法类面临如下困难。

定理 2.3。任取足够小的 $\kappa>0$ 以及任意 $\alpha \ge 4\kappa^2$。则存在一个 $n_0 \in \mathbb{N}$和一个绝对常数 $c>0$，使得以下结论成立。对任意 $n \ge n_0$，不存在$(e^{-cn}$，$\alpha)$-在线算法，能够在定义 *SBP* 所述意义下处理 *2.1*。

我们将在第 2.3 节证明定理 4.1。下面给出几点说明。

定理 2.3 表明，在 $\kappa \to 0$ 这一范围内，当密度 SBP 为 $\alpha = \Omega(\kappa^2)$ 时，在线算法无法为 SBP 找到满足解。这显著改进了 [GKPX22 中定理 7 的早期结果。4]，该结果仅表明，当 α 足够接近可满足性阈值 $\alpha_c(\kappa)$ 时，在线算法才会失效。此外，根据定理 2.2，定理 2.3 已达到可能的最优界：Bansal-Spencer 算法在在线算法中（相差常数因子）是最优的；按照定义 2.1，任何在线算法都无法改进这一结果。

定理所排除的算法2.3 不必以高概率甚至以常数概率取得成功：指数微小的成功保证就足够了。这是定理2.3的一项特殊优势；据我们所知，对于以概率 $o(1)$成功的算法，我们还不了解任何类似的困难性保证。这基于对詹森不等式的一项巧妙应用，最初归功于 Gamarnik 和 Sudan [GS17b]。

差异最小化的在线算法我们的第二个主要结果将定理2.3 推广到了以下情形的差异最小化问题：的条目为 $\mathcal{M}$ 二值。

定理 2.4。设 $c<1/2$ 为任意的，$n \ge M=\omega(1)$，且 $\mathcal{M} \in \{-1,1\}^{M\times n}$ 具有独立同分布的 Rademacher 条目。则存在一个 $M_0 \in \mathbb{N}$ 使得以下结论成立。对于每个 $M \ge M_0$，不存在 $\left(e^{-cM}\right.$，$\left.\sqrt{M}/24\right)$-在线算法（关于差异度的定义见定义*2.1*）。

此外 e，定理 2.4 即使 $\mathcal{M}$ 的元素服从伯努利(p).

定理 2.5。 设 $c < 1/2$ 任意， $n \geq M = \omega(1)$，且 $\mathcal{M} \in \{0,1\}^{M\times n}$ 其元素为独立同分布的伯努利 (p)。那么存在一个 $M_0 \in \mathbb{N}$ 使得以下结论成立。对于每个 $(p-p^2)M \geq M_0$，不存在 $\left(e^{-cM}, \sqrt{M(p-p^2)/24}\right)$-在线算法能够按照定义 2.1 的意义解决差异度问题 2.1。

我们在第 2.4 节证明定理 4.2 ，并在第 2.5 节将结论推广至定理 4.3。

定理 2.4-2.5共同表明，在常数因子范围内，Bansal-Spencer 算法对于差异最小化问题，在在线算法这一类别中是最优的。再次强调，被排除的算法即使以指数级小的概率也可能成功。

从技术层面来说，定理 2.4-2.5 是通过证明前文所述的某些解的 m元组不存在而得出的，这些元组具有不断增大的 m, $m = \omega_M$（1）值。考虑具有 m个元组且满足 $m = \omega$（1）的“景观”这一想法，最初是在随机数划分问题 [GK21a]的背景下提出的。通过这样做，作者随后降低了 m- 重叠间隙性质的阈值，并排除了比常数 m所能涵盖范围更广的一系列参数下的稳定算法。我们的工作首次将这种障碍运用于稳定算法之外，并采用不断增大的 m 值；这进一步展现了，在随机计算问题中考虑超常数规模元组的潜在收益。我们证明中的另一个关键要素，是通过 Berry-Esseen 定理得到的关于伯努利/拉德马赫变量有符号和的反集中不等式。

3 寻找常数差异解的算法障碍

本节聚焦于寻找常数差异解这一算法问题。更具体地说，给定一个随机 $\mathcal{M} \in \mathbb{R}^{M\times n}$ ，我们提出以下问题：对于哪些M 和 n的取值，能够高效地找到一个差异度为常数的解 $\boldsymbol{\sigma} \in \Sigma_n$ ，即 $\|\mathcal{M}\boldsymbol{\sigma}\|_\infty = O(1)$?

首先，通过简单的一阶矩计算可知，要使这类解存在， $n = \Omega(M \log M)$ 是必要条件。事实证明，这一条件也是充分的； [ANW21]证明了，若 $\mathcal{M}$ 的元素服从独立同分布的伯努利(p) 分布，那么 $\mathcal{D}(\mathcal{M}) \leq 1$ 以高概率成立，只要 $n \geq CM \log M$ 其中 C 是任意一个大于 $(2 \log 2)$ 的常数[-1]。另一方面，目前已知能找到此类解的最佳多项式时间算法，仅在 $M = o(\log n)$ [Ban22]时有效。这凸显了显著的统计到计算的差距（SCG）。

在本节中，我们研究在光滑情形下 SCG 的性质，其中 $\mathcal{M}$ 的各项独立同分布于 $\mathcal{N}(0,1)$，且接近存在边界 $n = \Theta(M \log M)$。我们首先关注常数差异解集合的“景观”，并展示集合系综 m-重叠间隙性质这一复杂的几何性质。随后，我们利用 m-重叠间隙性质排除稳定算法这一类算法。

3.1 技术预备知识

我们形式化 所研究的常数差异解元组集合 .

定义 3.1。固定一个 $K > 0$、一个 $m \in \mathbb{N}$、 $0 < \eta < \beta < 1$ 和 $\mathcal{I} \subset [0, \pi/2]$。设 $\mathcal{M}_i \in \mathbb{R}^{M\times n}$ $0 \leq i \leq m$ 为独立同分布的随机矩阵，且每个矩阵的元素均为独立同分布的 $\mathcal{N}(0,1)$。用 $\mathcal{S}(K, m, \beta, \eta, \mathcal{I})$ 表示所有满足以下条件的 m 元组 $\boldsymbol{\sigma}_i \in \Sigma_n$, $1 \leq i \leq m$ 的集合：

- (成对 . e **Overlap Condition)** *For any* $1 \leq i < j \leq m$, $\beta - \eta \leq n^{-1} \langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j \rangle \leq \beta$

- (常数差异条件) 存在 $\tau_1, \ldots, \tau_m \in \mathcal{I}$，使得 $\max_{1\le i\le m} \left\|\mathcal{M}_i(\tau_i)\boldsymbol{\sigma}_i\right\|_\infty \le K$，其中 $\mathcal{M}_i(\tau_i) = \cos(\tau_i)\mathcal{M}_0 + \sin(\tau_i)\mathcal{M}_i \in \mathbb{R}^{M\times n}$。

定义 3.1 讨论差异度至多为 K的解元组。参数 m表示所考虑元组的大小，而 β 和 η 共同控制重叠的（禁止）区域。最后，集合 $\mathcal{I}$ 用于生成相关实例；这是建立集合 m-重叠间隙性质以排除稳定算法所必需的，详见下文。

3.2 差异最小化中的集合 m-重叠间隙性质

我们的下一个主要结果表明，常数差异解集合具有 m-重叠间隙性质。

定理 3.2. 任取常数 $C_1 > c_2 > 0$ 和一个 $K > 0$，假设 $C_1 M \log_2 M \ge n \ge c_2 M \log_2 M$。那么，存在一个 $m \in \mathbb{N}$、一个 $c > 0$ 和一个 $0 < \eta < \beta < 1$，使得以下结论成立。任取满足 $|\mathcal{I}| \le 2^{cn}$的 $\mathcal{I} \subset [0, \pi/2]$。那么，$\mathbb{P}\left[\mathcal{S}(K, m, \beta, \eta, \mathcal{I}) \ne \varnothing\right] \le 2^{-\Theta(n)}$.

我们在第4.4节证明定理3.2。这里有几点说明。定理3.2 表明，对于任意 $K > 0$ 以及整个范围 $n = \Theta(M \log M)$，差异度至多为 K 的解集合，对于适当的 m、 m， β 以及 η，都具有$n = \Theta(M \log M)$-OGP。结合前文讨论的先前工作，这为边界处的算法难度 [ANW21猜想 1] 提供了严格证据，并朝着该方向迈出了第一步。

我们的证明基于一阶矩方法：我们证明，对于适当选取的 m-元组，其期望数量呈指数级小 m， β， η 并应用马尔可夫不等式。此外，我们的证明揭示了 $\beta \gg \eta$：定理3.2 排除了彼此近乎等距的 m-元组常数差异解。此外，这些解不必都相对于同一个实例具有常数差异： $\mathcal{M}_i(\tau_i)$（出现在定义3.1 中）可能彼此相关。这称为集成 m-OGP，并且对于在定理3.4中排除稳定算法至关重要。

3.3 m-重叠间隙性质蕴含稳定算法的失效

在本节中，我们证明定理 m 中确立的-重叠间隙性质 3.2 蕴含 稳定算法 无法找到常数差异解。我们首先详细说明算法设定，并形式化定义我们所研究的稳定算法类。

算法设定 算法 $\mathcal{A}$ 是 $\mathbb{R}^{M\times n}$ 与 Σ_n 之间的映射，允许随机化：我们假设存在一个概率空间 $(\Omega\ \mathbb{P}_\omega)$，使得 $\mathcal{A}$：$\mathbb{R}^{M\times n} \times \Omega \to \Sigma_n$。对于任意 $\omega \in \Omega$ 和 $\mathcal{M} \in \mathbb{R}^{M\times n}$，我们希望 $\|\mathcal{M}\boldsymbol{\sigma}_{\mathrm{ALG}}\|_\infty = O(1)$，其中 $\boldsymbol{\sigma}_{\mathrm{ALG}} = \mathcal{A}(\mathcal{M}, \omega) \in \Sigma_n$。稳定算法类形式化定义如下。

定义 3.3。固定一个 $K > 0$。若算法 $\mathcal{A}$：$\mathbb{R}^{M\times n} \times \Omega \to \Sigma_n$ 满足以下条件（对所有充分大的 M 均成立），则称其为 $(K, \rho, p_f, p_{\mathrm{st}}, f, L)$-稳定的（用于差异最小化）。

- (成功) 对于具有独立同分布 $\mathcal{N}(0, 1)$ 条目的 $\mathcal{M} \in \mathbb{R}^{M\times n}$，

$$\mathbb{P}_{(\mathcal{M},\omega)}\left[\left\|\mathcal{M}\mathcal{A}(\mathcal{M}, \omega)\right\|_\infty \le K\right] \ge 1 - p_f.$$

• (稳定性) 设 $\mathcal{M}$、$\overline{\mathcal{M}} \in \mathbb{R}^{M\times n}$ 为随机矩阵，每个矩阵的条目均独立同分布且服从 $\mathcal{N}(0, 1)$，并且满足 $\mathbb{E}\left[\mathcal{M}_{ij}\overline{\mathcal{M}}_{ij}\right] = \rho$，其中 $1 \le i \le M$ 且 $1 \le j \le n$。那么，

$$\mathbb{P}_{(\mathcal{M},\overline{M},\omega)}\left[d_H\big(\mathcal{A}(\mathcal{M},\omega),\mathcal{A}(\overline{\mathcal{M}},\omega)\big) \le f + L\|\mathcal{M}-\overline{\mathcal{M}}\|_F\right] \ge 1 - p_{\text{st}}.$$

定义 3.3 与 [GKPX22 中的定义 3.1] 相同。以至少 $1-p_f$ 的概率，$\mathcal{A}$ 找到一个差异度低于 K 的解。$\mathcal{A}$ 能容忍值为 ρ 的输入相关性；参数 f 和 L 则量化了 $\mathcal{A}$ 的输出对输入变化的敏感性。该稳定性保证是概率性的——既针对 $\mathcal{M}$、$\overline{\mathcal{M}}$，也针对 $\mathcal{A}$ 的随机性 ω——且成立的概率至少为 $1-p_{\text{sto}}$。最后，f 一词只会使我们的否定性结果更强：即使 $\mathcal{M}$ 和 $\overline{\mathcal{M}}$ "过于接近"，$\mathcal{A}$ 也可以进行 f 次翻转。我们的最终主要结果如下。

定理 3.4。固定一个 $K > 0$，$C_1 > c_2 > 0$ 和一个 $\mathcal{L} > 0$。假设 $C_1 M$ 对数$_2$ $M \ge n \ge c_2 M$ 对数$_2$ M。令 $m \in \mathbb{N}$ 和 $0 < \eta < \beta < 1$ 为 m-OGP 参数，其取值由定理 3.2规定。设

$$C = \frac{\eta^2}{1600}, \quad Q = \frac{4800\mathcal{L}\pi}{\eta^2}, \quad \textit{and} \quad T = \exp_2\left(2^{4mQ\log_2 Q}\right). \tag{4}$$

那么，存在一个 $n_0 \in \mathbb{N}$，使得以下结论成立。对于每个 $n \ge n_0$，都不存在随机化算法 $\mathcal{A}: \mathbb{R}^{M\times n} \times \Omega \to \Sigma_n$ 按照定义 3.3 的意义，是

$$\left(K, \cos\left(\frac{\pi}{2Q}\right), \frac{1}{9(Q+1)T}, \frac{1}{9Q(T+1)}, Cn, \mathcal{L}\sqrt{\frac{n}{M}}\right) - \textit{stable}.$$

定理 3.4 的证明与 [GKPX22 的定理 3.2], 几乎完全相同，为简洁起见，此处略去。下面给出几点说明。

首先，$\mathcal{A}$的运行时间不受限制：定理 3.4 排除了任何按照定义 3.3 的意义、在适当参数下具有稳定性的 $\mathcal{A}$。其次，注意到 m、β、η、$\mathcal{L}$ 全都是 $O(1)$（因为 $n \to \infty$）。因此，根据式 (4)，C、Q、T也全都是 $O(1)$。这是我们结果的一个重要特点：定理 3.4 排除了具有常数成功/稳定性保证的算法。最后，由于 $C = O(1)$，即使 $\mathcal{M}$ 和$\overline{\mathcal{M}}$ 几乎相同，$\mathcal{A}$ 仍然可以进行 $\Theta(n)$次比特翻转。

4 证明

补充记号 本节首先介绍一组补充记号。对于任意集合 A，$|A|$ 表示其基数。给定任意事件 E，用 $\mathbb{I}\{E\}$表示其指示函数。对于任意 $v = (v(i): 1 \le i \le n) \in \mathbb{R}^n$ 和 $p > 0$，$\|v\|_p = \left(\sum_{1\le i\le n} |v(i)|^p\right)^{1/p}$ 和 $\|v\|_\infty = \max_{1\le i\le n} |v(i)|$。对于 v，$v' \in \mathbb{R}^n$，$\langle v, v'\rangle$，$\sum_{1\le i\le n} v(i)v'(i)$。对于任意 $\boldsymbol{\sigma}$，$\boldsymbol{\sigma}' \in \Sigma_n$，$\{-1,1\}^n$，$d_H(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$表示它们的汉明距离：$d_H(\boldsymbol{\sigma}, \boldsymbol{\sigma}') \sum_{1\le i\le n} \mathbb{I}\{\boldsymbol{\sigma}(i) \ne \boldsymbol{\sigma}'(i)\}$。对于任意 $r > 0$，$\log_r(\cdot)$ 和 $\exp_r(\cdot)$ 分别表示以 r 为底的对数函数和指数函数；当 $r = e$ 时，我们省略下标。对于 $p \in [0,1]$，$h_b(p)$ $-p\log_2 p - (1-p)\log_2(1-p)$。用 I_k 表示 $k \times k$单位矩阵，并用 1 表示全 1 向量，其维数将由上下文明确。给定 $\boldsymbol{\mu} \in \mathbb{R}^k$ 和 $\Sigma \in \mathbb{R}^{k\times k}$，用 $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ 表示 $\mathbb{R}^k$ 中的多元正态分布，其均值为 $\boldsymbol{\mu}$，协方差为 Σ。给定矩阵 $\mathcal{M}$，$\|\mathcal{M}\|_F$，$\|\mathcal{M}\|_2$和 $|\mathcal{M}|$ 分别表示 $\mathcal{M}$的弗罗贝尼乌斯范数、谱范数和行列式。

全文均采用标准的巴赫曼-朗道渐近记号，例如 $\Theta(\cdot)$、 $O(\cdot)$、 $o(\cdot)$、$\Omega(\cdot)$，其中底层的渐近关系通常可由上下文确定。在可能产生歧义的情况下，我们将底层的渐近关系以下标形式体现，例如 $\Theta_\kappa(\cdot)$。为简便起见，省略所有向下取整／向上取整算子。

4.1 定理 2.3的证明

令 $\kappa, \alpha > 0$, $M = n\alpha$, $m \in \mathbb{N}$ 和 $\Delta \in (0, 1/2)$。假设 $\mathcal{M}_1 \in \mathbb{R}^{M\times n}$ 的各项是独立同分布的 $\mathcal{N}(0,1)$，并令 $\mathcal{M}_2, \dots, \mathcal{M}_m \in \mathbb{R}^{M\times n}$ 为从 $\mathcal{M}_1$ 得到的随机矩阵，其中对 $\mathcal{M}_1$ 的最后 Δn 列分别进行独立重新采样。用 $\Xi(m, \Delta)$ 表示满足以下条件的所有 m-元组的集合：

- 最大值$_{1\le i\le m}\left\|\mathcal{M}_i\boldsymbol{\sigma}_i\right\|_\infty \le \kappa\sqrt{n}$。

$1 \le i < j \le n$ 和 $1 \le k \le n - \Delta n$ 时， $\boldsymbol{\sigma}_i(k) = \boldsymbol{\sigma}_j(k)$。

我们证明以下命题。

命题 4.1。取任意足够小的 $\kappa > 0$ ，并令 $\alpha \ge 4\kappa^2$。那么，存在一个 $m \in \mathbb{N}$ 和一个 $\Delta \in (0, 1/2)$, 使得

$$\mathbb{P}\left[\Xi(m, \Delta) = \varnothing\right] \ge 1 - e^{-\Theta(n)}.$$

我们首先假设命题 4.1成立，并说明如何推导出定理 2.3。固定一个 $c > 0$ ，并反设存在一个 (e^{-cn}, α)-在线 $\mathcal{A}$。对于 $\mathcal{M}_1, \dots, \mathcal{M}_m \in \mathbb{R}^{M\times n}$ ，如上所述，令

$$\boldsymbol{\sigma}_i \triangleq \mathcal{A}(\mathcal{M}_i) \in \Sigma_n, \quad 1 \le i \le m. \tag{5}$$

注意，对于任意 $1 \le i < j \le m$， $\mathcal{M}_i$ 和 $\mathcal{M}_j$ 的前 $n - \Delta n$ 列完全相同。因此，

$$\boldsymbol{\sigma}_i(k) = \boldsymbol{\sigma}_j(k) \quad \text{for} \quad 1 \le i < j \le m \quad \text{and} \quad 1 \le k \le n - \Delta n.$$

接下来，我们证明以下概率保证。

引理 4.2。

$$\mathbb{P}\left[\max_{1\le i\le m}\left\|\mathcal{M}_i\boldsymbol{\sigma}_i\right\|_\infty \le \kappa\sqrt{n}\right] \ge p_s^m.$$

引理 *4.2* 的证明。 我们的论证基于对詹森不等式的一种巧妙应用，该应用由 [GS17b 的引理 5 给出。3]。记 ζ 为 $1-\Delta$ 个n 的前几列 $\mathcal{M}_1$。也就是说， ζ 是 $\mathcal{M}_1, \dots, \mathcal{M}_m$之间共享的“共享随机性”。令

$$I_i = \mathbb{I}\left\{\|\mathcal{M}_i\boldsymbol{\sigma}_i\|_\infty \le \kappa\sqrt{n}\right\}.$$

于是，

$$\mathbb{P}\left[\max_{1\le i\le m}\left\|\mathcal{M}_i\sigma_i\right\|_\infty \le \kappa\sqrt{n}\right] = \mathbb{E}\left[I_1 \cdots I_m\right].$$

然后，我们只需注意到

$$\mathbb{E}\left[I_1 \cdots I_m\right] = \mathbb{E}_\zeta\left[\mathbb{E}\left[I_1 \cdots I_m|\zeta\right]\right] = \mathbb{E}_\zeta\left[\mathbb{E}[I_1|\zeta]^m\right] \ge \left(\mathbb{E}_\zeta\left[\mathbb{E}[I_1|\zeta]\right]\right)^m = \mathbb{E}[I_1]^m = p_s^m,$$

其中我们使用了 fac 注意到 $I_1, \dots, I_m$ 在给定 ζ 的条件下相互独立，以及詹森不等式 y. □

注意到显然 $(\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_m)\in\Xi(m,\Delta)$，特别是 $\Xi(m,\Delta)$ 以至少 e^{-cmn} 的概率非空。利用命题 4.1，我们得到

$$e^{-\Theta(n)}\geq\mathbb{P}\big[\Xi(m,\Delta)\neq\varnothing\big]\geq e^{-cmn}.$$

若 m 为常数且 $c>0$ 足够小，那么对所有足够大的 n都会导致矛盾。因此，只需证明命题4.1。

命题*4.1*的证明。我们的证明基于一阶矩方法。令

$$\mathcal{S}=\big\{(\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_m):\boldsymbol{\sigma}_i(k)=\boldsymbol{\sigma}_j(k),1\leq i<j\leq m,1\leq k\leq n-\Delta n\big\}.$$

注意到

$$\big|\Xi(m,\Delta)\big|=\sum_{(\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_m)\in\mathcal{S}}\mathbb{I}\left\{\max_{1\leq i\leq m}\|\mathcal{M}_i\boldsymbol{\sigma}_i\|_\infty\leq\kappa\sqrt{n}\right\}.\tag{6}$$

接下来，我们证明对于适当的 $m\in\mathbb{N}$ 和 $\Delta\in(0,1/2)$，

$$\mathbb{E}\big[|\Xi(m,\Delta)|\big]\leq e^{-\Theta(n)}.$$

计数估计 我们对 $|\mathcal{S}|$进行上界估计。 $\boldsymbol{\sigma}_1\in\Sigma_n$ 有 2^n 种选择。选定一个 $\boldsymbol{\sigma}_1$后，对于任意 $\boldsymbol{\sigma}_i$，$2\leq i\leq m$，都有 $2^{\Delta n}$ 种选择。因此，

$$|\mathcal{S}|\leq 2^n\big(2^{\Delta n}\big)^{m-1}\leq\exp_2\big(n+nm\Delta\big)\tag{7}$$

概率估计 固定任意 $(\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_m)\in\mathcal{S}$。分别将 $\mathcal{M}_1,\dots,\mathcal{M}_m$ 的首行记为 $R_1,\dots,R_m\in\mathbb{R}^n$；并设

$$Z_i=\frac{1}{\sqrt{n}}\langle R_i,\boldsymbol{\sigma}_i\rangle\overset{d}{=}\mathcal{N}(0,1),\quad 1\leq i\leq m.$$

注意，如果 $k\neq k'$ 或 $n-\Delta n+1\leq k=k'\leq n$, $\mathbb{E}\big[R_i(k)R_j(k')\big]=0$。利用这一事实，我们立即得出 $\mathbb{E}[Z_iZ_j]=1-\Delta$。特别地，$(Z_1,\dots,Z_m)\in\mathbb{R}^m$ 是一个协方差为 Σ 的中心化多元正态随机向量，其中

$$\Sigma=\Delta I_m+(1-\Delta)\mathbf{1}\mathbf{1}^T\in\mathbb{R}^{m\times m},$$

其中 $\mathbf{1}\in\mathbb{R}^m$ 是全 1 向量。特别地， Σ 的谱由特征值$\Delta+(1-\Delta)m$ 和特征值 Δ 构成，前者的重数为 1，后者的重数为 $m-1$。于是我们得到

$$\begin{aligned}\mathbb{P}\left[\max_{1\leq i\leq m}\|\mathcal{M}_i\boldsymbol{\sigma}_i\|_\infty\leq\kappa\sqrt{n}\right]&\leq\mathbb{P}\left[\max_{1\leq i\leq m}\big|\langle R_i,\boldsymbol{\sigma}_i\rangle\big|\leq\kappa\sqrt{n}\right]^{\alpha n}\\&\left((2\pi)^{-\frac{m}{2}}|\Sigma|^{-\frac{1}{2}}\int_{\boldsymbol{z}\in[-\kappa,\kappa]^m}\exp\left(-\frac{\boldsymbol{z}^T\Sigma^{-1}\boldsymbol{z}}{2}\right)\right)^{\alpha n}\\&\leq\left((2\pi)^{-\frac{m}{2}}\big(\Delta+(1-\Delta)m\big)^{-\frac{1}{2}}\Delta^{-\frac{m-1}{2}}(2\kappa)^m\right)^{\alpha n}.\end{aligned}\tag{8}$$

估计 $\mathbb{E}[|\Xi(\Delta,m)|]$ 现在我们结合（7）和（8）得到

$$\mathbb{E}\big[\big|\Xi(\Delta,m)\big|\big] \le \exp_2\Big(n\Psi(\Delta,m,\alpha)\Big), \tag{9}$$

其中

$$\Psi(\Delta,m,\alpha) = 1 + m\Delta - \frac{\alpha m}{2}\log_2(2\pi) + \alpha m\log_2(2\kappa) - \frac{\alpha(m-1)}{2}\log_2\Delta - \frac{\alpha}{2}\log_2\big(\Delta + (1-\Delta)m\big).$$

利用 $\log_2\frac{1}{\Delta} > 0$ if $\Delta < \frac{1}{2}$ 这一事实，我们进一步得到界

$$\Psi(\Delta,m,\alpha) \le m\left(\frac{1}{m} - \frac{\alpha}{2m}\log_2\big(\Delta + (1-\Delta)m\big) + \Upsilon(\Delta,\alpha)\right), \tag{10}$$

for

$$\Upsilon(\Delta,\alpha) = \Delta - \frac{\alpha}{2}\log_2(2\pi) + \alpha\log_2(2\kappa) - \frac{\alpha}{2}\log_2\Delta.$$

分析 $\Upsilon(\Delta,\alpha)$ 我们设定 $\Delta = (2\kappa)^2,$ ， 使得

$$\alpha\log_2(2\kappa) - \frac{\alpha}{2}\log_2\Delta = 0.$$

接下来，固定任意 $\alpha \ge 4\kappa^2$。于是，

$$\Upsilon(\Delta,\alpha) = \Delta - \frac{\alpha}{2}\log_2(2\pi) \le 4\kappa^2 - 2\kappa^2\log_2(2\pi) = -\kappa^2\big(2\log_2(2\pi) - 4\big) = -\Theta_\kappa(\kappa^2). \tag{11}$$

综合以上结果 对于固定的足够小的 $\kappa > 0$、 $\alpha \ge 4\kappa^2$ 和 $\Delta = (2\kappa)^2$，我们有 $\Upsilon(\Delta,\alpha) = -\Theta_\kappa(\kappa^2) < 0$。此外，

$$\frac{1}{m} - \frac{\alpha}{2m}\log_2\big(\Delta + (1-\Delta)m\big) = o_m(1)$$

即 $m \to \infty$。注意， $\Upsilon(\Delta、\ \alpha)$ 仅取决于 α， κ。因此，当 $m \in \mathbb{N}$ 足够大时，(11) 给出

$$\frac{1}{m} - \frac{\alpha}{2m}\log_2\big(\Delta + (1-\Delta)m\big) + \Upsilon(\Delta,\alpha) < 0.$$

因此，结合（9）和（10），我们得到

$$\mathbb{E}\big[\big|\Xi(m,\Delta)\big|\big] \le e^{-\Theta(n)}.$$

由此，根据马尔可夫不等式可得

$$\mathbb{P}\big[\big|\Xi(\Delta,m)\big| \ge 1\big] \le \mathbb{E}\big[\big|\Xi(\Delta,m)\big|\big] = \exp\big(-\Theta(n)\big).$$

□

4.2 定理2.4 的证明

定理 2.4 的证明与定理 2.3的证明类似。我们首先证明以下命题。

命题 4.3。令 $n \geq M = \omega(1)$、 $\mathcal{M}_1 \in \{-1,1\}^{M\times n}$ 具有独立同分布的 Rademacher 条目，并令 $m = \lceil \frac{2n}{M} \rceil$。通过独立地重新采样 $\mathcal{M}_1$ 的最后 M 列，生成 $\mathcal{M}_2$、……、 $\mathcal{M}_m \in \{-1,1\}^{M\times n}$。用 $\Xi_d(m, M)$ 表示所有满足以下条件的 m-元组 $\boldsymbol{\sigma}_1$、……、 $\boldsymbol{\sigma}_m \in \Sigma_n$ 的集合：

- 最大值$_{1\leq i\leq m} \|\mathcal{M}_i\boldsymbol{\sigma}_i\| \leq C_u\sqrt{M}$，其中 C_u， 124。
- 对于 $1 \leq i < j \leq m$ 和 $1 \leq k \leq n - M$， $\boldsymbol{\sigma}_i(k) = \boldsymbol{\sigma}_j(k)$。

于是，

$$\mathbb{P}\big[\Xi_d(m, M) = \varnothing\big] \geq 1 - e^{-n}.$$

在证明命题 4.3 之前，我们强调，如果 $n = \omega(M)$，那么 $m = \omega_M$（1），且重新采样的列所占比例趋于零。我们首先说明命题 4.3如何推出定理 2.4。为导出矛盾，假设存在一个 $\mathcal{A}$: $\{-1,1\}^{M\times n} \to \Sigma_n$其为（$e^{-cM}$， $\sqrt{M}/24$）-最优（其中 $c < 1/2$任意）。对于 $\mathcal{M}_i$, $1 \leq i \leq m$按照命题中的方式，定义

$$\boldsymbol{\sigma}_i \triangleq \mathcal{A}(\mathcal{M}_i) \in \Sigma_n, \quad 1 \leq i \leq n$$

并注意到

$$\boldsymbol{\sigma}_i(k) = \boldsymbol{\sigma}_j(k), \quad \text{for all} \quad 1 \leq i < j \leq m \quad \text{and} \quad 1 \leq k \leq n - M$$

由于 $\mathcal{A}$是在线的。接下来我们证明

引理 4.4。

$$\mathbb{P}\left[\max_{1\leq i\leq m} \|\mathcal{M}_i\boldsymbol{\sigma}_i\| \leq \frac{\sqrt{M}}{24}\right] \geq \left(e^{-cM}\right)^m \geq e^{-2cn}.$$

引理的证明 4.4 与引理 4.2相同。因此，在假设存在这样的 $\mathcal{A}$的情况下，我们得到 $\Xi_d(m,\ \Delta) \neq \varnothing$ 以至少 e^{-2cn}的概率。最后，利用命题 4.3，

$$e^{-n} \geq \mathbb{P}\big[\Xi_d(m, M) \neq \varnothing\big] \geq e^{-2cn}$$

这与 矛盾，因为 $c < 1/2$。因此，只需证明命题 4.3 .

命题的证明*4.3*。命题4.3 的证明与命题4.1类似；尤其基于一阶矩方法。设

$$\bar{\mathcal{S}} = \big\{(\boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_m) : \boldsymbol{\sigma}_i(k) = \boldsymbol{\sigma}_j(k), 1 \leq i < j \leq m, 1 \leq k \leq n - M\big\}.$$

注意，

$$\big|\Xi_d(m, M)\big| = \sum_{(\sigma_1,\ldots,\sigma_m)\in\bar{\mathcal{S}}} \mathbb{I}\left\{\max_{1\leq i\leq m} \|\mathcal{M}_i\boldsymbol{\sigma}_i\|_\infty \leq C_u\sqrt{M}\right\}, \quad \text{where} \quad C_u = \frac{1}{24}. \tag{12}$$

计数项 我们对 $|\bar{\mathcal{S}}|$ 进行界定。 $\boldsymbol{\sigma}_1$ 有 2^n 种选择；固定它后，对于任意 $\boldsymbol{\sigma}_i$、 $2 \le i \le m$，有 2^M 种选择。因此，

$$|\bar{\mathcal{S}}| \le 2^n (2^M)^{m-1} \le \exp_2\big(n + mM\big). \tag{13}$$

¯概率项。固定任意一个 $(\boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_m) \in \mathcal{S}$。令 $R_i \in \{\pm 1\}^n$, $1 \le i \le m$分别表示 $\mathcal{M}_i$, $1 \le i \le m$的首行。对每个固定的 i， $\mathcal{M}_i$ 的各行相互独立。因此，

$$\mathbb{P}\left[\max_{1\le i\le m} \|\mathcal{M}_i \boldsymbol{\sigma}_i\|_\infty \le C_u\sqrt{M}\right] = \mathbb{P}\left[\max_{1\le i\le m} |\langle R_i, \boldsymbol{\sigma}_i\rangle| \le C_u\sqrt{M}\right]^M. \tag{14}$$

接下来，令

$$R_i = \big(R_{ik} : 1 \le k \le n\big), \quad 1 \le i \le m.$$

任取 $1 \le i < j \le m$。注意，随机向量

$$\big(R_{ik} : 1 \le k \le n - M\big) \quad \text{and} \quad \big(R_{jk} : 1 \le k \le n - M\big)$$

是相同的。因此，我们省略第一个下标，改用 $\big(R_k\colon\ 1 \le k \le n - M\big)$。接下来，固定任意一个 $\boldsymbol{v} = (v_1, \ldots, v_{n-M}) \in \{-1, 1\}^{n-M}$，并定义

$$\Delta_i(\boldsymbol{v}) \triangleq \sum_{1\le k\le n-M} v_k \boldsymbol{\sigma}_i(k) \qquad \text{and} \qquad \Sigma_i \triangleq \sum_{n-M+1\le k\le n} R_{ik}\boldsymbol{\sigma}_i(k). \tag{15}$$

我们的目标是控制式（14）右侧的值。为此，我们的策略是对R_1、……、 R_{n-M}进行条件化，并对 Σ_i应用 Berry–Esseen 不等式。我们证明以下辅助结果。

引理 4.5。设 $Z_1, \ldots, Z_M$ 是独立同分布的拉德马赫随机变量， $\epsilon_i \in \{-1, 1\}$, $1 \le i \le M$ 是确定性符号，且 $I \subset \mathbb{R}$ 是长度为 $|I| = \omega_M(1)$ 的区间。那么

$$\mathbb{P}\big[Z_1\epsilon_1 + \cdots + Z_M\epsilon_M \in I\big] \le \frac{3|I|}{\sqrt{M}}$$

对所有足够大的 M，

引理 *4.5*的证明。 设 $\frac{1}{\sqrt{M}}I$ 表示集合 $\{c/\sqrt{M}\colon\ \ c \in I\}$。根据中心极限定理，

$$\frac{1}{\sqrt{M}} \sum_{1\le i\le M} Z_i\epsilon_i \Rightarrow \mathcal{N}(0, 1)$$

依分布收敛，其中收敛速度由 Berry–Esseen 不等式控制：

$$\left|\mathbb{P}\left[\sum_{1\le i\le M} Z_i\epsilon_i \in I\right] - \mathbb{P}\left[\mathcal{N}(0,1) \in \frac{1}{\sqrt{M}}I\right]\right| \le \frac{\mathcal{C}_{\text{be}}}{\sqrt{M}}. \tag{16}$$

这里， $\mathcal{C}_{\text{be}} > 0$ 是一个绝对常数。此外，我们有

$$\mathbb{P}\left[\mathcal{N}(0,1) \in \frac{1}{\sqrt{M}}I\right] = \frac{1}{\sqrt{2\pi}} \int_{u\in\frac{1}{\sqrt{M}}I} \exp(-u^2/2)\, du \le \frac{|I|}{\sqrt{2\pi M}}. \tag{17}$$

结合（ 16) 和 (17)，利用三角不等式，我们得到对于所有足够大的 M ，

$$\mathbb{P}\left[\sum_{1\le i\le M} Z_i\epsilon_i \in I\right] \le \frac{1}{\sqrt{M}}\left(\mathcal{C}_{\mathrm{be}} + \frac{|I|}{\sqrt{2\pi}}\right) \le \frac{3|I|}{\sqrt{M}}, \tag{18}$$

wh此处回顾了 $|I| = \omega_M$（1）以及 $\mathcal{C}_{\mathrm{be}} = O_M$（1）。由此得到引理 4.5 . □

接下来，任取 $\boldsymbol{v} \in \{-1,1\}^{n-M}$ 并令

$$I_i(\boldsymbol{v}) = \left[-C_u\sqrt{M} - \Delta_i(\boldsymbol{v}), C_u\sqrt{M} - \Delta_i(\boldsymbol{v})\right],$$

其中回顾 Δ_i（$\boldsymbol{v}$）来自（15）。特别地，

$$\left|I_i(\boldsymbol{v})\right| = 2C_u\sqrt{M}, \quad \text{for all} \quad 1 \le i \le m \quad \text{and} \quad v \in \{-1,1\}^{n-M}. \tag{19}$$

接下来取一个 $1 \le i \le m$ 并根据 Σ_i （15）回顾。应用引理 4.5，可得

$$\max_{1\le i\le m}\max_{\boldsymbol{v}\in\{-1,1\}^{n-M}} \mathbb{P}\Big[\Sigma_i \in I_i(\boldsymbol{v})\Big] \le 6C_u. \tag{20}$$

我们正 通过对 $R_1,\dots,R_{n-M}$进行条件化，从而界定概率项（14）。

$$\begin{aligned}
&\mathbb{P}\left[\max_{1\le i\le m} |\langle R_i, \boldsymbol{\sigma}_i\rangle| \le C_u\sqrt{M}\right] \\
&= \sum_{\boldsymbol{v}\in\{-1,1\}^{n-M}} \mathbb{P}\Big[\Sigma_i \in I_i(\boldsymbol{v}), 1\le i\le m \Big| (R_1,\dots,R_{n-M}) = \boldsymbol{v}\Big] \underbrace{\mathbb{P}\left[(R_1,\dots,R_{n-M}) = \boldsymbol{v}\right]}_{=2^{-(n-M)}} && (21)\\
&= 2^{-(n-M)} \sum_{\boldsymbol{v}\in\{-1,1\}^{n-M}} \mathbb{P}\Big[\Sigma_i \in I_i(\boldsymbol{v}), 1\le i\le m\Big] && (22)\\
&= 2^{-(n-M)} \sum_{v\in\{-1,1\}^{n-M}} \prod_{1\le i\le m} \mathbb{P}\big[\Sigma_i \in I_i(\boldsymbol{v})\big] && (23)\\
&\le (6C_u)^m. && (24)
\end{aligned}$$

现在我们来说明上述各行的依据。等式（21）是通过对“共同随机性” $R_1,\dots,R_{n-M}$进行条件化，并注意到它们在 $\{-1,1\}^{n-M}$上均匀分布而得出的。等式（22）利用了如下事实：对于任意固定的 $1 \le i \le m$， Σ_i与 $R_1,\dots,R_{n-M}$ 相互独立；而（ 23 ）利用了 $\Sigma_1,\dots,\Sigma_m$ 也是一组相互独立的随机变量这一事实。最后，（ 24 ）使用了（20）。

结合（14）和（24），因此我们得出

$$\max_{(\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_m)\in\bar{\mathcal{S}}} \mathbb{P}\left[\max_{1\le i\le m} \|\mathcal{M}_i\boldsymbol{\sigma}_i\|_\infty \le C_u\sqrt{M}\right] \le (6C_u)^{mM}. \tag{25}$$

界定 $\mathbb{E}\big[\big|\Xi_d(m,\ M)\big|\big]$ 现在我们可以估计 $\mathbb{E}\big[\big|\Xi_d(m,\ M)\big|\big]$。利用（12）、（13）和（25），

$$\mathbb{E}\big[\big|\Xi_d(m,M)\big|\big] \le \exp_2\left(n + mM - mM\log_2\frac{1}{6C_u}\right).$$

代入数值 $C_u = 1/24$ 和 $m \geq 2n/M$，可得

$$n + mM - mM \log_2 \frac{1}{6C_u} \leq -n,$$

使得 $\mathbb{E}\big[\big|\Xi_d(m, M)\big|\big] \leq e^{-n}$。最后，我们应用马尔可夫不等式得出结论：

$$\mathbb{P}\big[\Xi_d(m, M) \neq \varnothing\big] = \mathbb{P}\big[\big|\Xi_d(m, M)\big| \geq 1\big] \leq \mathbb{E}\big[\big|\Xi_d(m, M)\big|\big] \leq e^{-n}.$$

□

4.3 定理证明概要2.5

定理2.5 的证明与定理2.4非常相似；我们仅指出必要的改动。设 $\mathcal{M}_1$ 由独立同分布的伯努利(p) 条目组成。假设存在一个 $\mathcal{A}$： $\mathbb{R}^{M\times n} \to \Sigma_n$，按照定义2.1的含义，它是 $(e^{-cM}, C'_u\sqrt{M})$-在线的，其中 $c < 1/2$是任意的，且

$$C'_u = \frac{\sqrt{p - p^2}}{24}.$$

我们令 $m = 2n/M$，并说明如何将命题4.3 改用于此情形。完成这一步后，其余部分可直接照搬定理2.4的证明。首先，将定理2.4 证明中的所有 C_u 替换为$C'_u = C_u\sqrt{p - p^2}$。其次，式 (13) 中的计数估计保持不变。另一方面，引理4.5则替换为以下结果。

引理 4.6. 设 $Z_1, \ldots, Z_M$ 是独立同分布的伯努利(p)随机变量，$\epsilon_i \in \{-1, 1\}$, $1 \leq i \leq M$ 是确定性符号，且 $I \subset \mathbb{R}$ 是长度为 $|I| = \omega_{(p-p^2)M}(1)$ 的区间。那么

$$\mathbb{P}\big[Z_1\epsilon_1 + \cdots + Z_M\epsilon_M \in I\big] \leq \frac{3|I|}{\sqrt{M(p - p^2)}},$$

对于足够大的 M。

引理 *4.6* 的证明。 注意到 $\mathbb{E}[Z_i\epsilon_i] = p\epsilon_i$ 且

$$\mathrm{Var}(Z_i\epsilon_i) = \mathbb{E}[Z_i^2\epsilon_i^2] - p^2\epsilon_i^2 = p - p^2,$$

当 $\epsilon_i \in \{-1, 1\}$ 时。因此，根据中心极限定理，

$$\frac{1}{\sqrt{(p - p^2)M}}\left(\sum_{1\leq i\leq M} Z_i\epsilon_i - p\,\langle \mathbf{1}, \boldsymbol{\epsilon}\rangle\right) \Rightarrow \mathcal{N}(0, 1)$$

在分布意义下，其中 $\mathbf{1} \in \mathbb{R}^M$ 是全 1 向量，且 $\boldsymbol{\epsilon} = (\epsilon_i\colon\ 1 \leq i \leq M) \in \{-1, 1\}^M$。此外，根据 Berry–Esseen 不等式，我们有

$$\left|\mathbb{P}\left[\sum_{1\leq i\leq M} Z_i\epsilon_i \in I\right] - \mathbb{P}\left[\mathcal{N}(0, 1) \in \frac{I - p\,\langle \mathbf{1}, \boldsymbol{\epsilon}\rangle}{\sqrt{(p - p^2)M}}\right]\right| \leq \frac{C'_{\mathrm{be}}}{\sqrt{(p - p^2)M}}$$

对于某个绝对常数 $\mathcal{C}'_{\mathrm{be}} > 0$。这里，

$$\frac{I - p\langle \mathbf{1}, \boldsymbol{\epsilon}\rangle}{\sqrt{(p-p^2)M}} = \left\{ \frac{c - p\langle \mathbf{1}, \boldsymbol{\epsilon}\rangle}{\sqrt{(p-p^2)M}} : c \in I \right\},$$

使得

$$\left| \frac{I - p\langle \mathbf{1}, \boldsymbol{\epsilon}\rangle}{\sqrt{(p-p^2)M}} \right| = \frac{|I|}{\sqrt{(p-p^2)M}},$$

利用勒贝格测度的平移不变性。接下来，完全按照引理 4.5 的证明方式，我们证明引理 4.6。 □

结合引理 4.6 并沿用完全相同的记号，式（20）改写为式（26），其中

$$\max_{1\le i\le m} \max_{\boldsymbol{v}\in\{-1,1\}^{n-M}} \mathbb{P}\Big[\Sigma_i \in I_i(\boldsymbol{v})\Big] \le \frac{6C'_u}{\sqrt{p-p^2}} = \frac{1}{4}. \tag{26}$$

现在我们继续　类似地处理第（21）至（24）行。注意，对于任意的 $\boldsymbol{v} \in \{0,1\}^{n-M}$，

$$\mathbb{P}\Big[\Sigma_i \in I_i(\boldsymbol{v}), 1 \le i \le m \Big| (R_1, \dots, R_{n-M}) = \boldsymbol{v}\Big] \le 2^{-2m}, \tag{27}$$

利用式（26）。因此，

$$\begin{aligned}
&\mathbb{P}\left[\max_{1\le i\le m} |\langle R_i, \boldsymbol{\sigma}_i\rangle| \le C'_u\sqrt{M}\right] \\
&\le \sum_{\boldsymbol{v}\in\{0,1\}^{n-M}} \mathbb{P}\Big[\Sigma_i \in I_i(\boldsymbol{v}), 1 \le i \le m \Big| (R_1, \dots, R_{n-M}) = \boldsymbol{v}\Big] \mathbb{P}\Big[(R_1, \dots, R_{n-M}) = \boldsymbol{v}\Big] \\
&\le 2^{-2m} \sum_{\boldsymbol{v}\in\{0,1\}^{n-M}} \mathbb{P}\Big[(R_1, \dots, R_{n-M}) = \boldsymbol{v}\Big] \\
&= 2^{-2m},
\end{aligned}$$

其中最后第二行用到了（27）。这与（24）的界完全相同，因此证明的其余部分仍然成立。定理 2.5的证明至此完成。

4.4 定理 3.2 的证明

固定一个 $K > 0, C_1 > c_2 > 0$，并假设

$$C_1 M \log_2 M \ge n \ge c_2 M \log_2 M. \tag{28}$$

我们通过一阶矩方法证明结果。注意，根据马尔可夫不等式，

$$\mathbb{P}\big[\mathcal{S}(K, m, \beta, \eta, \mathcal{I}) \ne \varnothing\big] = \mathbb{P}\big[\big|\mathcal{S}(K, m, \beta, \eta, \mathcal{I})\big| \ge 1\big] \le \mathbb{E}\big[\big|\mathcal{S}(K, m, \beta, \eta, \mathcal{I})\big|\big].$$

因此，只需证明

$$\mathbb{E}\big[\big|\mathcal{S}(K, m, \beta, \eta, \mathcal{I})\big|\big] \le 2^{-\Theta(n)}.$$

现在我们估计 $\mathbb{E}\big[\big|\mathcal{S}(K, m, \beta, \eta, \mathcal{I})\big|\big]$。

计数项 固定 $m \in \mathbb{N}, 0 < \eta < \beta < 1$，并用 $M(m, \beta, \eta)$ 表示满足 $\beta - \eta \le n^{-1} \langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j \rangle \le \beta$（对 $1 \le i < j \le m$）的 m 元组 $(\boldsymbol{\sigma}_i \in \Sigma_n : 1 \le i \le m)$ 的数量。我们证明

引理 4.7。对于 $m = O(1)$，令 $n \to \infty$,

$$M(m, \beta, \eta) \le \exp_2 \left(n + n(m-1) h_b \left(\frac{1 - \beta + \eta}{2} \right) + O(\log_2 n) \right).$$

引理 4.7 照录自 [GKPX22，引理 6。7], 为完整起见，我们给出证明。

引理 *4.7*的证明。注意到 $\langle \boldsymbol{\sigma}, \boldsymbol{\sigma}' \rangle = n - 2 d_H(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ 对任意 $\boldsymbol{\sigma}, \boldsymbol{\sigma}' \in \Sigma_n$均成立。共有 2^n种 $\boldsymbol{\sigma}_1$选择。固定一个 $\boldsymbol{\sigma}_1$后，有

$$\sum_{\substack{\rho : \frac{1-\beta}{2} \le \rho \le \frac{1-\beta+\eta}{2} \\ \rho n \in \mathbb{N}}} \binom{n}{n\rho} \le \binom{n}{n\frac{1-\beta+\eta}{2}} n^{O(1)},$$

种选择，适用于任意 $\boldsymbol{\sigma}_i$, $2 \le i \le m$，且满足约束 $\beta - \eta \le n^{-1} \langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_i \rangle \le \beta$。接下来，对于任意$\rho \in (0, 1)$, $\binom{n}{n\rho} = \exp_2\big(n h(\rho) + O\big(\log_2 n\big)\big)$，由斯特林近似可得。结合以上结果并回忆 $m = O_n(1)$，得到引理 4.7。 □

概率估计：固定任意满足条件的 $(\boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_m)$

$$\frac{1}{n} \langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j \rangle = \beta - \eta_{ij}, \quad 1 \le i < j \le m.$$

显然 $0 \le \eta_{ij} \le \eta$。此外，令 $\boldsymbol{\eta} = (\eta_{ij} : 1 \le i < j \le m) \in \mathbb{R}^{m(m-1)/2}$。那么 $\|\boldsymbol{\eta}\|_\infty \le \eta$。我们最终对参数 β、η 和 m 的选择将确保

$$\eta = \frac{1 - \beta}{2m}. \tag{29}$$

现在我们来控制概率项。

引理 4.8。设 $\Sigma(\boldsymbol{\eta}) \in \mathbb{R}^{m \times m}$，其对角线元素均为 1，且对于 $1 \le i < j \le m$,

$$\big(\Sigma(\boldsymbol{\eta})\big)_{ij} = \big(\Sigma(\boldsymbol{\eta})\big)_{ji} = \beta - \eta_{ij}.$$

那么，下述结论成立。

(a) $\Sigma(\boldsymbol{\eta})$ 是正定的（PD），当且仅当 η 满足 (29)。

(b) 假设 η 满足(29)。那么,

$$\mathbb{P}\left[\exists \tau_1, \ldots, \tau_m \in \mathcal{I} : \max_{1 \le i \le m} \big\| \mathcal{M}_i(\tau_i) \boldsymbol{\sigma}_i \big\|_\infty \le K \right] \le |\mathcal{I}|^m (2\pi)^{-\frac{mM}{2}} \left(\frac{1 - \beta}{2} \right)^{-\frac{Mm}{2}} \left(\frac{2K}{\sqrt{n}} \right)^{Mm}.$$

引理 *4.8*的证明。

（a）部分　设 $E \in \mathbb{R}^{m\times m}$ 使得 $E_{ii}=0$ 且 $E_{ij}=E_{ji}=-\eta_{ij}$ 对于 $1\le i<j\le m$。那么　，

$$\Sigma(\boldsymbol{\eta}) = (1-\beta)I + \beta \mathbf{1}\mathbf{1}^T + E.$$

注意，$(1-\beta)I+\beta\mathbf{1}\mathbf{1}^T$ 的最小特征值为 $1-\beta$，且 $\|E\|_2 \le \|E\|_F < \eta m$。因此，若 $\eta < (1-\beta)/m$，则 $\Sigma(\boldsymbol{\eta})$ 可逆。回顾其为协方差矩阵这一事实，特别是它是半正定的，由此可得（a）部分。

（b）部分。首先，我们对 $\mathcal{I}$ 使用并集界，得到

$$\mathbb{P}\left[\exists \tau_1,\dots,\tau_m \in \mathcal{I} : \max_{1\le i\le m} \left\|\mathcal{M}_i(\tau_i)\boldsymbol{\sigma}_i\right\|_\infty \le K\right] \le |\mathcal{I}|^m \max_{\tau_i\in\mathcal{I}, 1\le i\le m} \mathbb{P}\left[\max_{1\le i\le m} \left\|\mathcal{M}_i(\tau_i)\boldsymbol{\sigma}_i\right\|_\infty \le K\right]. \tag{30}$$

接下来，记 $R_i \sim \mathcal{N}(0, I_n)$ 为 $\mathcal{M}_i(\tau_i) \in \mathbb{R}^{M\times n}$, $1\le i\le m$ 的第一行。注意，利用每个 $\mathcal{M}_i(\tau_i)$ 都具有独立行这一事实，

$$\mathbb{P}\left[\max_{1\le i\le m} \left\|\mathcal{M}_i(\tau_i)\boldsymbol{\sigma}_i\right\|_\infty \le K\right] \le \mathbb{P}\left[\max_{1\le i\le m} n^{-\frac{1}{2}} \left|\langle R_i, \boldsymbol{\sigma}_i\rangle\right| \le \frac{K}{\sqrt{n}}\right]^M. \tag{31}$$

接下来，我们考虑多元正态随机向量 $\left(n^{-1/2}\langle R_i, \boldsymbol{\sigma}_i\rangle : 1\le i\le m\right)$，其由标准正态坐标组成。令 $\overline{\Sigma}$ 表示其协方差矩阵，该矩阵取决于 $\tau_1,\dots,\tau_m$ 的选择。注意，对于 $1\le i<j\le m$,

$$\overline{\Sigma}_{ij} = \frac{1}{n}\mathbb{E}\left[\langle R_i,\boldsymbol{\sigma}_i\rangle\langle R_j,\boldsymbol{\sigma}_j\rangle\right] = \frac{1}{n}(\boldsymbol{\sigma}_i)^T \underbrace{\mathbb{E}[R_i R_j^T]}_{=\cos(\tau_i)\cos(\tau_j) I_m} \boldsymbol{\sigma}_j = \cos(\tau_i)\cos(\tau_j)(\beta-\eta_{ij}).$$

现在，我们利用 Sid τ_i ak´ Sid [Sid68的推论 1]，借助高斯比较不等式，消除对 [GKPX22定理 6 中给出的版本。5]。

定理 4.9。设 $(X_1,\dots,X_k)\in\mathbb{R}^k$ 是一个中心化多元正态随机向量。假设其协方差矩阵 $\Sigma\in\mathbb{R}^{k\times k}$ 的对角元素均为 1，且具有以下形式：存在$0\le\lambda_i\le 1$, $1\le i\le k$，使得对于任意 $1\le i\ne j\le k$, $\Sigma_{ij}=\lambda_i\lambda_j\rho_{ij}$，其中 $(\rho_{ij} : 1\le i\ne j\le k)$ 是一个固定的任意协方差矩阵。固定取值 $c_1,\dots,c_k>0$，并令

$$P(\lambda_1,\dots,\lambda_k) = \mathbb{P}\left[|X_1|<c_1, |X_2|<c_2,\dots,|X_k|<c_k\right].$$

那么，$P(\lambda_1,\dots,\lambda_k)$ 是每个 λ_i, $i=1,2,\dots,k$, $0\le\lambda_i\le 1$ 的非递减函数。也就是说，

$$P(\lambda_1,\lambda_2,\dots,\lambda_k) \le P(1,1,\dots,1).$$

现在令 $(Z_1,\dots,Z_m)$ 为一个协方差为$\Sigma(\boldsymbol{\eta})$ 的中心化多元正态随机向量。注意到

$$\max_{\tau_1,\dots,\tau_m\in\mathcal{I}} \mathbb{P}\left[\max_{1\le i\le m} n^{-\frac{1}{2}}\left|\langle R_i,\boldsymbol{\sigma}_i\rangle\right| \le \frac{K}{\sqrt{n}}\right] \le \mathbb{P}\left[\max_{1\le i\le m}|Z_i| \le \frac{K}{\sqrt{n}}\right] \tag{32}$$

$$= (2\pi)^{-\frac{m}{2}}|\Sigma(\boldsymbol{\eta})|^{-\frac{1}{2}} \int_{\boldsymbol{z}\in\left[-\frac{K}{\sqrt{n}},\frac{K}{\sqrt{n}}\right]^m} \exp\left(-\frac{\boldsymbol{z}^T\Sigma(\boldsymbol{\eta})^{-1}\boldsymbol{z}}{2}\right) d\boldsymbol{z}$$

$$\le (2\pi)^{-\frac{m}{2}}|\Sigma(\boldsymbol{\eta})|^{-\frac{1}{2}} \left(\frac{2K}{\sqrt{n}}\right)^m, \tag{33}$$

这里的（32）由定理 4.9推出，而（33）则由平凡事实 $\exp\left(-\frac{z^T\Sigma(\eta)^{-1}z}{2}\right)\le 1$推出。最后，我们对 $|\Sigma(\eta)|$进行界定。为此，我们将借助矩阵分析中的以下工具。

定理 4.10（霍夫曼–维兰特不等式）。设 $A\in\mathbb{R}^{m\times m}$ 和 $A+E\in\mathbb{R}^{m\times m}$ 是两个特征值为

$$\lambda_1(A)\ge\cdots\ge\lambda_m(A)\quad and\quad \lambda_1(A+E)\ge\cdots\ge\lambda_m(A+E).$$

则，

$$\sum_{1\le i\le m}\big(\lambda_i(A+E)-\lambda_i(A)\big)^2\le\|E\|_F.$$

参见 [HJ12，推论 6.3。8] 相关参考文献，以及 [HW53] 的原始论文。我们应用定理 4.10 于 $\Sigma(\eta)$。设 $A=(1-\beta)I+\beta\mathbf{1}\mathbf{1}^T$的特征值为 $\lambda_1=1-\beta+\beta m>\lambda_2=\cdots=\lambda_m=1-\beta$，且 E如上所述。假设 $A+E$ 的特征值为 $\mu_1\ge\cdots\ge\mu_m$。固定任意一个 $2\le i\le m$。定理 4.10 给出

$$\big|\mu_i-(1-\beta)\big|\le\|E\|_F\le\eta m=\frac{1-\beta}{2},$$

从而得到

$$\mu_i\ge\frac{1-\beta}{2},\quad 2\le i\le m.$$

此外，该界也适用于 μ_1，即

$$\mu_1\ge1-\beta+\beta m-\frac{1-\beta}{2}>\frac{1-\beta}{2}.$$

由于 $\eta\in\mathbb{R}^{m(m-1)/2}$ 在满足$\|\eta\|_\infty\le\eta\le\frac{1-\beta}{2m}$ 的条件下是任意的，我们得到

$$\inf_{\substack{\eta\in\mathbb{R}^{m(m-1)/2}\\ \|\eta\|_\infty\le\frac{1-\beta}{2m}}}\big|\Sigma(\eta)\big|=\prod_{1\le i\le m}\mu_i\ge\left(\frac{1-\beta}{2}\right)^m.\tag{34}$$

最后，结合（30）、（31）、（33）和（34），我们完成了（b）部分的证明。 □

估计期望 令 $\mathcal{F}(m,\beta,\eta)$ 为所有满足 $\beta-\eta\le n^{-1}\langle\boldsymbol{\sigma}_i,\boldsymbol{\sigma}_j\rangle\le\beta$, $1\le i<j\le m$ 的 m 元组 $(\boldsymbol{\sigma}_1,\ldots,\boldsymbol{\sigma}_m)$ 的集合。那么

$$\big|\mathcal{S}(K,m,\beta,\eta,\mathcal{I})\big|=\sum_{(\boldsymbol{\sigma}_1,\ldots,\boldsymbol{\sigma}_m)\in\mathcal{F}(m,\beta,\eta)}\mathbb{I}\left\{\exists\tau_1,\ldots,\tau_m\in\mathcal{I}:\max_{1\le i\le m}\big\|\mathcal{M}_i(\tau_i)\boldsymbol{\sigma}_i\big\|_\infty\le K\right\}.$$

利用期望的线性性、引理 4.7、引理 4.8以及 $\log_2|\mathcal{I}|\le cn$这一事实，我们得到

$$\mathbb{E}\big[\big|\mathcal{S}(K,m,\beta,\eta,\mathcal{I})\big|\big]\le\exp_2\Big(\Psi(m,\beta,\eta,c)+O(\log_2 n)\Big),\tag{35}$$

其中

$$\Psi(m,\beta,\eta,c)=n+mnh_b\left(\frac{1-\beta+\eta}{2}\right)+cmn+\frac{mM}{2}\log_2\frac{4K^2}{\pi(1-\beta)}-\frac{Mm}{2}\log_2 n.\tag{36}$$

我们将 η 和 c 设为

$$\eta = \frac{1-\beta}{2m} \quad \text{and} \quad c = \frac{1}{m}, \tag{37}$$

其中我们回顾了 η，见式（29）；参数 β 和 m 稍后将进行调节。现在回顾式（ n 中的28）的缩放。特别地,

$$\log_2 n \geq \log_2 c_2 + \log_2 M + \log_2 \log_2 M \geq \log_2 c_2 + \log_2 M.$$

由此，我们得到

$$\begin{aligned}\Psi\left(m, \beta, \frac{1-\beta}{2m}, \frac{1}{m}\right) \leq 2C_1 M \log_2 M + mC_1 M \log_2 M \cdot h_b\left(\frac{1-\beta}{2} + \frac{1-\beta}{4m}\right) \\ + \frac{mM}{2}\log_2 \frac{4K^2}{\pi(1-\beta)c_2} - \frac{Mm}{2}\log_2 M.\end{aligned} \tag{38}$$

注意，若 $\beta \in (1/2, 1)$ 且 $m \in \mathbb{N}$，显然有

$$h_b\left(\frac{1-\beta}{2} + \frac{1-\beta}{4m}\right) \leq h_b(1-\beta).$$

我们选择 $\beta^* > 1/2$ 使得

$$h_b(1-\beta^*) = \min\left\{\frac{1}{4C_1}, \frac{1}{2}\right\}.$$

So,

$$mC_1 M \log_2 M \cdot h_b\left(\frac{1-\beta^*}{2} + \frac{1-\beta^*}{4m}\right) \leq \frac{Mm}{4}\log_2 M. \tag{39}$$

结合式（38）和式（39），我们进一步给出的上界

$$\Psi\left(m, \beta^*, \frac{1-\beta^*}{2m}, \frac{1}{m}\right) \leq 2C_1 M \log_2 M - \frac{Mm}{4}\log_2 M + \Theta(mM). \tag{40}$$

最后，取 $m = m^* = \max\{2, 16C_1\}$，得到

$$\Psi\left(m^*, \beta^*, \frac{1-\beta^*}{2m^*}, \frac{1}{m^*}\right) = -\Theta(M \log_2 M). \tag{41}$$

结合（35）以及以下事实 $O(\log_2 n) = O(\log_2 M) = o(M \log_2 M)$，再结合 $M = \omega(1)$，我们得出

$$\mathbb{E}\left[\left|\mathcal{S}(K, m^*, \beta^*, \eta^*, \mathcal{I})\right|\right] \leq \exp_2\left(\Psi\left(m^*, \beta^*, \frac{1-\beta^*}{2m^*}, \frac{1}{m^*}\right) + o(M \log_2 M)\right) = 2^{-\Theta(M \log M)} = 2^{-\Theta(n)}.$$

至此，定理 3.2 的证明完成。

致谢

第一作者的部分经费由美国国家科学基金会项目 DMS-2015517 资助。第二作者获哥伦比亚大学统计学杰出博士后奖学金资助。第三作者的部分经费由美国国家科学基金会项目 DMS-1847451 资助。

参考文献

[ACO08] Dimitris Achlioptas 和 Amin Coja-Oghlan，来自相变的算法障碍，《2008年第49届IEEE计算机科学基础年会》，IEEE，2008年，第793–802页。[Ajt96] Mikl´os Ajtai，生成格问题的困难实例，《第二十八届ACM计算理论年度研讨会论文集》，1996年，第99–108页。[ALS21a] Emmanuel Abbe、Shuangping Li 和 Allan Sly，二元感知机：高效算法能够在一个罕见且连通性良好的簇中找到解，arXiv 预印本 arXiv:2111.03084（2021年）。[ALS21b] ———，对称感知机连续性猜想的证明及其对数正态极限，arXiv 预印本 arXiv:2102.13069（2021年）。[Alt22] Dylan J Altschuler，对称感知机的涨落，arXiv 预印本 arXiv:2205.02319（2022年）。[ANW21] Dylan J Altschuler 和 Jonathan Niles-Weed，随机矩形矩阵的差异度，《随机结构与算法》（2021年）。[APZ19] Benjamin Aubin、Will Perkins 和 Lenka Zdeborov´a，对称二元感知机的存储容量，《物理学杂志A：数学与理论》52（2019年），第29期，294003。[ART06] Dimitris Achlioptas 和 Federico Ricci-Tersenghi，随机约束满足问题的解空间几何，《第三十八届ACM计算理论年度研讨会论文集》，2006年，第130–139页。[AS20] Ahmed El Alaoui 和 Mark Sellke，负球形感知机的算法纯态，arXiv 预印本 arXiv:2010.15811（2020年）。[AWZ20] G´erard Ben Arous、Alexander S Wein 和 Ilias Zadik，稀疏主成分分析中的自由能阱与重叠间隙性质，《学习理论会议》，PMLR，2020年，第479–482页。[BABB21] Enric Boix-Adser`a、Matthew Brennan 和 Guy Bresler，计算 Erd ¨os–R ´enyi 超图中团计数的平均情况复杂度，《SIAM计算杂志》（2021年），第0期，FOCS19–39。[Ban10] Nikhil Bansal，差异最小化的构造性算法，《2010年IEEE第51届计算机科学基础年会》，IEEE，2010年，第3–10页。[Ban22] ———，个人通信，2022年。[BB19] Matthew Brennan 和 Guy Bresler，稀疏主成分分析的最优平均情形归约：从弱假设到强困难性，arXiv 预印本 arXiv:1902.07380（2019年）。

[BBBZ07] Carlo Baldassi、Alfredo Braunstein、Nicolas Brunel 和 Riccardo Zecchina，具有二值突触网络中的高效监督学习，《美国国家科学院院刊》104（2007），第26期，11079–11084。[BBH18] Matthew Brennan、Guy Bresler 和 Wasim Huleihel，具有植入稀疏结构问题的可约性与计算下界，arXiv预印本 arXiv:1806.07508（2018）。[BDVLZ20] Carlo Baldassi、Riccardo Della Vecchia、Carlo Lucibello、Luca Saglietti 和 Riccardo Zecchina，对称二元感知机中的解的聚类，《统计力学杂志：理论与实验》2020（2020），第7期，073303。[BGT10] Mohsen Bayati、David Gamarnik 和 Prasad Tetali，稀疏随机图中插值方法与标度极限的组合方法，2010年第四十二届ACM计算理论研讨会论文集，第105–114页。[BH21] Guy Bresler 和 Brice Huang，低阶多项式的随机 k-SAT问题的算法相变，arXiv预印本 arXiv:2106.02129（2021）。[BHK+19] Boaz Barak、Samuel Hopkins、Jonathan Kelner、Pravesh K Kothari、Ankur Moitra 和 Aaron Potechin，植入团问题近乎紧的平方和下界，《SIAM计算杂志》48（2019），第2期，687–735。[BIL+15] Carlo Baldassi、Alessandro Ingrosso、Carlo Lucibello、Luca Saglietti 和 Riccardo Zecchina，次主导稠密簇使离散突触神经网络实现简单学习与高计算性能，《物理评论快报》115（2015），第12期，128101。[BNSX21] Erwin Bolthausen、Shuta Nakajima、Nike Sun 和 Changji Xu，小密度下伊辛感知机模型的Gardner公式，arXiv预印本 arXiv:2111.02855（2021）。[BPW18] Afonso S Bandeira、Amelia Perry 和 Alexander S Wein，关于计算差距与统计差距的注记：利用统计物理进行预测，《葡萄牙数学》75（2018），第2期，159–186。[BR13] Quentin Berthet 和 Philippe Rigollet，稀疏主成分分析的计算下界，arXiv预印本 arXiv:1304.0828（2013）。[BS20] Nikhil Bansal 和 Joel H. Spencer，随机输入的在线平衡，《随机结构与算法》57（2020），第4期，879–891（美国英语）。[CGPR19] Wei-Kuo Chen、David Gamarnik、Dmitry Panchenko 和 Mustazee Rahman，一类最大割问题中局部算法的次优性，《概率年刊》47（2019），第3期，1587–1618。[COE15] Amin Coja-Oghlan 和 Charilaos Efthymiou，关于随机图中的独立集，《随机结构与算法》47（2015），第3期，436–486。[Cos09] Kevin P Costello，高斯向量的平衡，《以色列数学杂志》172（2009），第1期，145–156。

[Cov65] Thomas M Cover，具有线性不等式系统的几何与统计性质及其在模式识别中的应用，《IEEE电子计算机汇刊》（1965），第3期，326–334。[DFJ02] Martin Dyer、Alan Frieze 和 Mark Jerrum，关于稀疏图中独立集的计数，《SIAM计算杂志》31（2002），第5期，1527–1541。[DKS17] Ilias Diakonikolas、Daniel M Kane 和 Alistair Stewart，高维高斯分布与高斯混合模型鲁棒估计的统计查询下界，2017年IEEE第58届计算机科学基础年度研讨会（FOCS），IEEE，2017，第73–84页。[DM15] Yash Deshpande 和 Andrea Montanari，隐藏团与隐藏子矩阵问题的改进平方和下界，学习理论会议，2015，第523–562页。[DS19] Jian Ding 和 Nike Sun，伊辛感知器的容量下界，2019年第51届ACM SIGACT计算理论研讨会论文集，第816–827页。[FGR+17] Vitaly Feldman、Elena Grigorescu、Lev Reyzin、Santosh S Vempala 和 Ying Xiao，统计算法与检测植入团的下界，《ACM学报》64（2017），第2期，1–37。[F L92] Alan M Frieze 和 T Luczak，随机正则图的独立数与色数，《组合理论杂志B辑》54（1992），第1期，123–132。[FPV18] Vitaly Feldman、Will Perkins 和 Santosh Vempala，具有植入解的随机可满足性问题的复杂性，《SIAM计算杂志》47（2018），第4期，1294–1338。[FS20] Cole Franks 和 Michael Saks，具有许多列的随机矩阵的差异度，《随机结构与算法》57（2020），第1期，64–96。[Gam21] David Gamarnik，重叠间隙性质：优化随机结构的拓扑障碍，《美国国家科学院院刊》118（2021），第41期。[Gar87] Elizabeth Gardner，神经网络中的最大存储容量，EPL（《欧洲物理学快报》）4（1987），第4期，481。[Gar88] ———，神经网络模型中的相互作用空间，《物理学杂志A：数学与一般》21（1988），第1期，257。[GD88] Elizabeth Gardner 和 Bernard Derrida，神经网络模型的最优存储性质，《物理学杂志A：数学与一般》21（1988），第1期，271。

[GJ21] David Gamarnik 和 Aukosh Jagannath，《重叠间隙性质与 p 自旋模型的近似消息传递算法》，《概率年刊》49 (2021)，第 1 期，180–205。[GJS21] David Gamarnik、Aukosh Jagannath 和 Subhabrata Sen，《主子矩阵恢复中的重叠间隙性质》，《概率论及其相关领域》181 (2021)，757–814。[GJW20] David Gamarnik、Aukosh Jagannath 和 Alexander S Wein，《随机优化问题的低次数困难性》，2020 IEEE 第 61 届计算机科学基础年会（FOCS），IEEE，2020，第 131–140 页。[GJW21] ———，《p 自旋优化问题的电路下界》，arXiv 预印本 arXiv:2109.01342 (2021)。[GK21a] David Gamarnik 和 Eren C Kızıldağ，《随机数划分问题中的算法障碍》，arXiv 预印本 arXiv:2103.01369 (2021)。[GK21b] David Gamarnik 和 Eren C. Kızıldağ，《平均而言，计算 Sherrington–Kirkpatrick 模型的配分函数是困难的》，《应用概率年刊》31 (2021)，第 3 期，1474–1504。[GKPX22] David Gamarnik、Eren C Kızıldağ、Will Perkins 和 Changji Xu，《对称二元感知机模型中的算法与障碍》，arXiv 预印本 arXiv:2203.15667 (2022)。[GMZ22] David Gamarnik、Cristopher Moore 和 Lenka Zdeborová，《无序系统视角下的计算困难性》，《统计力学杂志：理论与实验》2022 (2022)，第 11 期，114015。[GS14] David Gamarnik 和 Madhu Sudan，《稀疏随机图上局部算法的极限》，第五届理论计算机科学创新会议论文集，2014，第 369–376 页。[GS17a] ———，《稀疏随机图上局部算法的极限》，《概率年刊》45 (2017)，第 4 期，2353–2376。[GS17b] ———，《随机 NAE-K-SAT 问题的序贯局部算法的性能》，《SIAM计算杂志》46 (2017)，第 2 期，590–619。[HJ12] Roger A Horn 和 Charles R Johnson，《矩阵分析》，剑桥大学出版社，2012。[HK14] Haiping Huang 和 Yoshiyuki Kabashima，《二值突触学习计算困难性的起源》，《物理评论 E》90 (2014)，第 5 期，052813。[HKP+17] Samuel 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力量》，2017 IEEE 第 58 届计算机科学基础年会（FOCS），IEEE，2017，第 720–731 页。

[Hop18] Samuel Brink Klevit Hopkins, 《统计推断与平方和方法》。[HR19] Rebecca Hoberg 和 Thomas Rothvoss, 《随机集合系统差异度的傅里叶分析方法》, 第三十届 ACM-SIAM 离散算法年会论文集, SIAM, 2019, 第 2547–2556 页。[HS21] Brice Huang 和 Mark Sellke, 《优化平均场自旋玻璃的紧 Lipschitz 困难性》, arXiv 预印本 arXiv:2110.07847 (2021)。[HSS15] Samuel B Hopkins、Jonathan Shi 和 David Steurer, 《通过平方和证明进行张量主成分分析》, 学习理论会议, 2015, 第 956–1006 页。[HW53] AJ Hoffman 和 HW Wielandt, 《正规矩阵谱的变化》, 《杜克数学杂志》20 (1953), 第 1 期, 37–39。[Jer92] Mark Jerrum, 《大团难以被 Metropolis 过程找到》, 《随机结构与算法》3 (1992), 第 4 期, 347–359。[JH60] Roger David Joseph 和 Louise Hay, 《n 维空间中与 s 维子空间相交的正交象限数》, 技术报告, 康奈尔航空实验室公司, 纽约州布法罗, 1960。[Kız22] Eren C Kızıldağ, 《高维统计与随机组合结构中的算法及算法障碍》, 博士论文, 麻省理工学院, 2022。[KK82] Narendra Karmarkar 和 Richard M Karp, 《集合划分的差分法》, 计算机科学系（EECS）, 加州大学伯克利分校, 1982。[KKLO86] Narendra Karmarkar、Richard M Karp、George S Lueker 和 Andrew M Odlyzko, 《最优划分的概率分析》, 《应用概率杂志》23 (1986), 第 3 期, 626–645。[KM89] Werner Krauth 和 Marc M´ezard, 《具有二元耦合的记忆网络的存储容量》, 《物理学杂志》50 (1989), 第 20 期, 3057–3066。[KR98] Jeong Han Kim 和 James R Roche, 《用随机半立方体覆盖立方体及其在二元神经网络中的应用》, 《计算机与系统科学杂志》56 (1998), 第 2 期, 223–252。[KWB22] Dmitriy Kunisky、Alexander S Wein 和 Afonso S Bandeira, 《关于假设检验计算困难性的札记：基于低次似然比的预测》, 《数学分析、应用与计算：ISAAC 2019》, 葡萄牙阿威罗, 2019 年 7 月 29 日至 8 月 2 日, 施普林格, 2022, 第 1–50 页。[LM15] Shachar Lovett 和 Raghu Meka, 《沿边行走实现构造性差异最小化》, 《SIAM计算杂志》44 (2015), 第 5 期, 1573–1582。

[LRR17] Avi Levy、Harishchandra Ramadas 和 Thomas Rothvoss，《通过乘法权重更新方法实现确定性差异最小化》，国际整数规划与组合优化会议，施普林格，2017年，第380–391页。[Mat99] Jiri Matousek，《几何差异：图解指南》，第18卷，Springer Science & Business Media，1999年。[MMZ05] Marc M´ézard、Thierry Mora 和 Riccardo Zecchina，《随机可满足性问题中解的聚类》，《物理评论快报》94（2005），第19期，197205。[NS23] Shuta Nakajima 和 Nike Sun，《伊辛感知机模型的尖锐阈值序列与普适性》，2023年ACM-SIAM离散算法年会论文集（SODA），SIAM，2023年，第638–674页。[Pot18] Aditya Potukuchi，《随机超图模型中的差异度》，arXiv预印本 arXiv:1811.01491（2018）。[PX21] Will Perkins 和 Changji Xu，《对称伊辛感知机的冻结1-RSB结构》，第53届ACM SIGACT计算理论研讨会年会论文集，2021年，第1579–1588页。[Rot17] Thomas Rothvoss，《凸集的构造性差异度最小化》，《SIAM计算杂志》46（2017），第1期，224–234。[RSS18] Prasad Raghavendra、Tselil Schramm 和 David Steurer，《通过平方和证明进行高维估计》，arXiv预印本 arXiv:1807.11419 6（2018）。[RV17] Mustazee Rahman 和 Balint Virag，《用于独立集的局部算法仅达到半最优》，《概率年刊》45（2017），第3期，1543–1577。[Sid68] Zbynek Sid´ak，《矩形的多元正态概率：其对相关系数的依赖》，《数理统计年刊》39（1968），第5期，1425–1434。[Spe85] Joel Spencer，《六个标准差就足够》，《美国数学会汇刊》289（1985），第2期，679–706。[SS23] Ashwin Sah 和 Mehtaab Sawhney，《对称感知机阈值的分布》，arXiv预印本 arXiv:2301.10701（2023）。[ST03] Mariya Shcherbina 和 Brunello Tirozzi，《Gardner问题的严格解》，《数学物理通讯》234（2003），第3期，383–422。[Sto13] Mihailo Stojnic，《再探Gardner问题》，arXiv预印本 arXiv:1306.3979（2013）。[Tal99] Michel Talagrand，《随机半立方体的交集》，《随机结构与算法》15（1999），第3–4期，436–449。[Tal11] ———，《自旋玻璃的平均场模型：高级副本对称性与低温》，施普林格，2011年。

[TMR20] Paxton Turner、Raghu Meka 和 Philippe Rigollet，《高维高斯向量平衡》，学习理论会议，PMLR，2020年，第3455–3486页。[Wei20] Alexander S Wein，《最大独立集的最优低次困难性》，arXiv预印本 arXiv:2010.06563（2020）。[Wei22] ———，《低次多项式张量分解的平均情形复杂度》，arXiv预印本 arXiv:2211.05274（2022）。[Wen62] James G Wendel，《几何概率中的一个问题》，Mathematica Scandinavica 11（1962年），第1期，109–111。[Win61] Robert O Winder，《单级阈值逻辑》，第二届开关电路理论与逻辑设计年会（SWCT 1961），IEEE，1961年，第321–332页。[Xu19] Changji Xu，《伊辛感知机模型的尖锐阈值》，arXiv预印本 arXiv:1905.05978（2019）。[Yak96] Benjamin Yakir，《用于划分的差分算法ldm：证明Karmarkar和Karp的一个猜想》，Mathematics of Operations Research 21（1996年），第1期，85–99。[ZK16] Lenka Zdeborov´a 和 Florent Krzakala，《推断的统计物理：阈值与算法》，Advances in Physics 65（2016年），第5期，453–552。